Springer

Handbook of Systematic Musicology

Bader
Editor

Springer

Springer Handbook of Systematic Musicology

Springer Handbooks provide a concise compilation of approved key information on methods of research, general principles, and functional relationships in physical and applied sciences. The world's leading experts in the fields of physics and engineering will be assigned by one or several renowned editors to write the chapters comprising each volume. The content is selected by these experts from Springer sources (books, journals, online content) and other systematic and approved recent publications of scientific and technical information.

The volumes are designed to be useful as readable desk reference book to give a fast and comprehensive overview and easy retrieval of essential reliable key information, including tables, graphs, and bibliographies. References to extensive sources are provided.

Springer Handbook

# of Systematic Musicology

Rolf Bader (Ed.)

With 952 Figures and 61 Tables

Springer

Editor
Rolf Bader
University of Hamburg
Institute of Systematic Musicology
Neue Rabenstr. 13
20354 Hamburg
r_bader@t-online.de

ISBN: 978-3-662-55002-1 e-ISBN: 978-3-662-55004-5
https://doi.org/10.1007/978-3-662-55004-5
Library of Congress Control Number: 2018930527

Production and typesetting: le-tex publishing services GmbH, Leipzig
Typography and layout: schreiberVIS, Seeheim
Illustrations: Hippmann GbR, Schwarzenbruck
Cover design: eStudio Calamar Steinen, Barcelona
Cover production: WMXDesign GmbH, Heidelberg

Printed on acid free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Systematic Musicology is a field that grew tremendously over the last decades to such an extent and into so many topics, that for quite some time researchers as well as students in the field have been demanding a comprehensive overview of the different parts of the discipline. As the field of Systematic Musicology is so wide, there is a need for introductions, reviews, teaching materials, and the display of recent trends for researchers in related fields who want to connect and interact within this highly interdisciplinary area. Therefore, the *International Working Group of Systematic and Comparative Musicology* decided to compile such a volume, covering the major areas of research in Systematic Musicology, and making them easily accessible to the community. This volume tries to present the state-of-the-art in the field while also giving an overview of basic and fundamental methodologies and terminologies. It also discusses recent trends and topics and therefore hopefully is also inspiring in terms of an interchange of ideas and subjects.

Systematic Musicology is a highly interdisciplinary field, which it has been ever since its ancient origins with philosophers like Pythagoras, Archytas of Tarrent, or Aristoxenos. Here the connection between music, mathematics, geometry, astronomy, well-being, politics, and other fields were discussed. Much attention was given to music theory, especially tonal systems or rhythm theory. Ancient writings of music theory, like the Indian *Natyashastra* on music and the arts, or the music theory embedded in the ancient Tamil epic *Cilappatikāram* of today's Sri Lanka show complex tonal system theories. The Chinese *Yueji* from 1st and 2nd century BC is not only a music theory, but also discusses aspects of nature and the social role of music. In all these traditions musicology, concerned with all aspects of music, continued to the present day in many forms.

In modern times the roots of Systematic Musicology lie in Comparative Musicology, which tries to determine laws and universals by comparing the musical traditions of the world. This was only possible from around 1900 after the invention of the Edison phonograph, which made it possible for the first time to record and playback music on wax cylinders. Following the systematic approach, right from the start the Hornbostel/Sachs classification of musical instruments used a taxonomy based on the acoustical and mechanical driving mechanisms of musical instruments, i. e., plucked, bowed, blown, etc. The Berlin phonogram archive, recording music from around the world on wax cylinders, started with a recording of a Thai *phi pha* orchestra in the Berlin *Tiergarten* in 1900. Jaap Kunst was among the first to collect recordings in today's Indonesia in the 1920s and 1930s developing Ethnomusicology on his many fieldtrips. In the US, Charles Seeger was co-founder of the Society of Comparative Musicology and Frances Desmond was collecting the music mainly of native Dakota Indians.

Carl Stumpf was among the first to introduce music psychology in modern terms, by discussing how a pitch sensation can appear from a multisensory input of many frequencies entering the ear. A distinct influence on these early findings came from Gestalt psychologists, who indeed first used musical melodies to derive their Gestalt laws. Early experiments on musical timbre were performed again by Stumpf and others, especially focusing on music and speech as well as on musical transients and tone color.

Many problems of music psychology, like the problem of tonal fusion addressed by Stumpf, were based on the experimental evidence that sounds consists of overtones, which was published by Hermann von Helmholtz in 1863. The growing field of electronic music enhanced research in these fields in the Bell labs in the US or in the Heinrich-Hertz Institute for vibration of Karl Willy Wagner in Berlin, where also Barkhausen was working on tubes and on the light bow for building a loudspeaker and microphone in one device. Research on musical instruments was prominently performed by Felix Savart around 1800, especially on violins, building a trapezoid-shaped violin body or an *octobass* ranging over two floors. Also around this time Friedrich Chladni was discovering longitudinal waves and building musical instruments with this invention, the *clavicylinder* and the *euphonium*, and wrote the first comprehensive work on vibrating musical systems.

All these and many more early works building up the field of Systematic Musicology were followed by a tremendous increase in the number of works after WWII. Today, the topics discussed in Systematic Musicology range from Musical Acoustics and Musical Signal Processing, Music Psychology and Neuromusicology, Music Ethnology and Comparative Musicology, Musical Syntax and Philosophy to Musical Applications in modern music production and distribution,

alongside many related topics. The aim of the discipline is still to understand music, its production and perception, its cultural, historical and philosophical background in a systematic way. This cannot be done by discussing one aspect alone. So, for example, the development of musical instruments is shaped not only by tradition, but by acoustical, physiological and psychological constraints, the needs of different compositional styles, to be technically state-of-the-art, or by political and economic demands. Therefore, Systematic Musicology needs to consider all aspects of music so as to aim for a system of music, the *Kunst* (art) in the Renaissance or Baroque sense of a system of rules, as J.S. Bach used the term in his *Die Kunst der Fuge* (The art of the fugue). This late work by Bach concerned displaying the rules and possibilities that lie within the polyphonic system of fugue composition.

So Systematic Musicology believes that understanding music can only truly be achieved by considering all aspects of music. The Handbook tries to reflect this interdisciplinary nature of Systematic Musicology in seven sections. These sections follow the main topics in the field, Musical Acoustics, Music Psychology, and Music Ethnology while also taking recent research trends into consideration, such as Embodied Music Cognition and Media Applications. Other topics, like Music Theory or Philosophy of Music are also discussed. Of course, they could also have been presented as sections in their own right, taking their salience into account. These topics are incorporated within one of the main sections under recent developments in the field. This may, however, be subject to respective developments over the next decades.

Part A, edited by Rolf Bader, presents an overview of the state-of-the-art of research in *Musical Acoustics*. After an introduction to basic models, mathematical frameworks and formulas, the section discusses materials for musical instruments as well as modern measurement techniques used in the field. Then, the main instrument families are discussed, string and wind instruments, organs and percussion instruments, in terms of their basic properties as well as special features. A discussion on nonlinearities and synchronization in musical instruments tries to give a global framework for instrument families. Finally, modern Room Acoustics is discussed, both in terms of general findings as well as modeling techniques.

In Part B, edited by Jonas Braasch, modern *Musical Signal Processing* is discussed. Again the section starts with the fundamentals of modern music recording studio equipment and methods, digital audio processing using delay lines and waveguides, and an introduction to Fourier and convolution methods. It then turns to sound source separation and state-of-the-art automatic musical score extraction and adaptive music control. Modern synthesis methods are discussed in terms of multichannel audio reproduction systems using wavefield synthesis, physical modeling using Finite-Difference methods (FDM), as well as hardware implementation of FDM with highly parallel processing Field-Programmable Gate Array (FPGA) hardware.

Part C, edited by Stefan Koelsch, turns to *Music Psychology*, addressing many aspects of music perception and associated musical parameters. It starts with a review of music time perception, which is a basic issue of an art form developing in time. This leads to the topic of the dependency of sensual information stored in memory, both short- and long-term memory processing and auditory working memory. Here many findings from brain research using EEG, fMRI, MEG, or related techniques are reported and discussed. On a higher level musical syntax needs to be understood, both in terms of music theory often based on syntax rules found in speech and its relations to music, as well as in terms of experimental findings again using brain research techniques. Following on from brain research findings, rhythm and beat perception are discussed, as well as the relation between music perception and active music production. Finally, music and emotion are discussed in terms of their relation to sensory data.

Part D, edited by Albrecht Schneider, is devoted to *Psychoacoustics*, its fundamentals, history, methods, and findings. The philosophical background of sensation as found in ancient as well as Renaissance and Baroque times is discussed and basic models of perception and psychophysics are presented. Thereafter, pitch and pitch perception is addressed, with basic parameters like virtual or residue pitch, tonal fusion, scales, or intonation. The auditory pathway is presented from the cochlea up to the auditory cortex. Then timbre and sound color are discussed in terms of their notion as well as their use in musical instrument organology, both of acoustic as well as electronic music, together with the relatedness of pitch and timbre and its use for audio segregation. Finally, loudness is treated as a psychological sensation related to physical sound pressure level, and different loudness models are presented and discussed in modern disco and club performance scenarios.

Part E, edited by Marc Leman, presents the emerging field of *Embodied Music Cognition*. After an introduction presenting ontological and epistemological foundations, the method is applied to musical parameters like timing, expressiveness, or expectation. The role of musical gestures are discussed and new trends and questions are addressed. The ideas are then transferred to compositional ideas using sonic objects as basic compositional units of timbre in terms of analysis

and synthesis. Music therapy is also subject to embodiment applications, where the relation between rhythm perception and body coordination is shown to improve well-being and the health of patients. Finally, the interaction of musicians and listeners using sensors and other techniques is treated, both in terms of basic frameworks as well as technical issues concerning sensors and their applications.

Part F, edited by Isabel Barbancho, concentrates on the application side within the *Modern Media Application* domain of music production and reproduction in terms of Music Information Retrieval (MIR). One aim is to extract musical features from audio files and use them for content analysis, song suggestions, or further analysis. Problems of data mining, automatic analysis, or visualizations are discussed on a technical level. Advanced hearing aids are also a subject of music applications, especially as normal hearing aids only have poor abilities to reproduce music. Modern tools and ideas help to enhance music perception through hearing aids. Applications of MRI also improve music education and education as a whole, and this is discussed in general as well as in a case study. Embedding music in a media structure is another recent challenge, as users expect highly automatic guidance and help tools, as well as many additional features besides the auditory stream. Changes in music production and perception due to modern media are a recent issue discussed along with their development over the last decades. Applications to music ethnology are treated in the domain of Computational Ethnomusicology, detecting high-level features of non-Western music styles through MIR. Also composition is due to automatization, especially voice but also for other instruments. Modern instrument building uses sensors or additional components to enhance musical instrument features, also addressing a world music approach.

The final Part G, edited by Rolf Bader, discusses problems in modern *Ethnomusicology* along with some examples of musical styles from around the world. It begins with two papers introducing the history of the discipline focusing on the relation between Systematic Musicology and Ethnomusicology, as well as discussing modern Analytical Ethnomusicology, focusing on musical features, instruments, and compositions rather than extramusical features like social or economic issues. Displaying relations between a wide range of musical styles, a systematic overview of the music of Africa is presented. As an example of the relationship between musical diversity and relatedness, the music of Southeast Asian minorities are addressed. Many musical features, like the origin of music or its historical development are often discussed in analogy with the music of the stone age, therefore an overview of music archaeology is given. As improvisation is a common feature of many musical styles around the world, a framework, as well as problems and discussions on free improvisation are also present. Finally, music and its relation to politics is presented with the history of protest songs in the postwar US.

We hope that the Handbook is not only of use to students in the discipline but also to advanced scholars and people working outside the field that are interested in an overview of Systematic Musicology, its recent developments, findings, and problems. Of course not all issues, problems, and topics of modern Systematic Musicology may be addressed in a single Handbook, especially not for a field that is in its own right so interdisciplinary. Still, it is our hope that researchers from other fields may be inspired by the topics and problems found in this Handbook and may enlarge their views in a truly interdisciplinary way.

Hamburg, January 2018
Rolf Bader

About the Editor

Rolf Bader is Professor for Systematic Musicology at the Institute of Systematic Musicology, University of Hamburg, Germany. He studied Systematic Musicology, Physics, Ethnology, and Historic Musicology at the University of Hamburg where he obtained his PhD and Habilitation on topics of Musical Acoustics, Music Psychology, and Musical Signal Processing. He was a visiting scholar at the Center for Computer Music and Research (CCRMA) at Stanford University (2005-2006). His major fields of research are musical acoustics and musical signal processing, musical hardware and software development, music psychology and neurocognition, music ethnology, and philosophy of music.

He is co-editor of the Springer series *Current Research in Systematic Musicology* and published several books here as an author, like *Nonlinearities and Synchronization in Musical Acoustics and Music Psychology* or *Computational Mechanics of the Classical Guitar*, or as an editor like *Sound – Perception – Performance* or *Concepts, Experiments, and Fieldwork: Studies in Systematic Musicology*, alongside many contributing book chapters or peer-reviewed papers.

He has conducted fieldwork as an Ethnomusicologist in Bali, Nepal, Thailand, Cambodia, Myanmar, Sri Lanka, China, and India since 1999. He has also worked as a professional musician, composer, artist, running recording studios, as a music journalist, leading exhibitions, and running a cinema.

List of Authors

Jakob Abeßer
Fraunhofer IDMT
Ehrenbergstr. 31
98693 Ilmenau, Germany
jakob.abesser@idmt.fraunhofer.de

Judit Angster
Fraunhofer Institute of Building Physics (IBP)
Acoustics
Nobelstr. 12
70569 Stuttgart, Germany
judit.angster@ibp.fraunhofer.de

Simha Arom
12 rue Ernest Psichari
75007 Paris, France
simha.arom@gmail.com

Rolf Bader
University of Hamburg
Institute of Systematic Musicology
Neue Rabenstr. 13
20354 Hamburg, Germany
r_bader@t-online.de

Ana M. Barbancho
Universidad de Málaga
ATIC Research Group, Dep. Ingeniería de Comunicaciones, ETSI Telecomunicación
Campus de Teatinos s/n
29071 Malaga, Spain
abp@ic.uma.es

Isabel Barbancho
Universidad de Málaga
ATIC Research Group, Dep. Ingeniería de Comunicaciones, ETSI Telecomunicación
Campus de Teatinos s/n
29071 Malaga, Spain
ibp@ic.uma.es

Juan Pablo Bello
New York University
35 W 4th St.
New York, NY 10012, USA
jpbello@nyu.edu

Stefan Bilbao
University of Edinburgh
Acoustics and Audio Group
Mayfield Rd.
Edinburgh, EH9 3JZ, UK
s.bilbao@ed.ac.uk

David Borgo
University of California San Diego
9500 Gilman Dr.
La Jolla, CA 92093-0099, USA
dborgo@ucsd.edu

Jonas Braasch
Rensselaer Polytechnic Institute
110 8th St.
Troy, NY 12180, USA
braasj@rpi.edu

Elvira Brattico
Aarhus University
Department of Clinical Medicine
Nørrebrogade 44
8000 C, Aarhus, Denmark
elvira.brattico@clin.au.dk

Estefanía Cano
Fraunhofer IDMT
Ehrenbergstr. 31
98693 Ilmenau, Germany
cano@idmt.fraunhofer.de

Amin Chaachoo
Tetouan-Asmir Center
Place 9 Avril
Tetuán, Morocco
chaachooamin@gmail.com

Marshall Chasin
Musicians' Clinics of Canada
340 College St.
Toronto, M5T 3A9, Canada
marshall.chasin@rogers.com

Jesús Corral García
Universidad de Málaga
Departamento de Ingeniería de Comunicaciones, ETSI Telecomunicación
Campus de Teatinos s/n
29071 Malaga, Spain
jcorral@ic.uma.es

Lola L. Cuddy
Queen's University
Dept. of Psychology
Kingston, K7L3N6, Canada
lola.cuddy@queensu.ca

Phillippe Depalle
McGill University
Schulich School of Music
555 Sherbrooke St. West
Montreal, QC H3A1E3, Canada
philippe.depalle@mcgill.ca

José-Miguel Díaz-Báñez
Escuela Superior de Ingenieros, Universidad de Sevilla
Departamento de Matemática Aplicada II
Camino de los Descrubrimientos, s/n
41092 Sevilla, Spain
dbanez@us.es

Christian Dittmar
International Audio Laboratories Erlangen
Am Wolfsmantel 33
91058 Erlangen, Germany
christian.dittmar@audiolabs-erlangen.de

Peter Driessen
University of Victoria
Dept. of Electrical and Computer Engineering
3800 Finnerty Rd.
Victoria, V8P 5C2, Canada
peter@ece.uvic.ca

Zhiyao Duan
University of Rochester
Dept. of Electrical and Computer Engineering
308 Hopeman
Rochester, NY 14627, USA
zhiyao.duan@rochester.edu

Tuomas Eerola
Durham University
Dept. of Music
Palace Green
Durham, DH13RL, UK
tuomas.eerola@durham.ac.uk

Ricardo Eichmann
Deutsches Archäologisches Institut
Orient-Abteilung
Podbielskiallee 69 – 71
14195 Berlin, Germany
ricardo.eichmann@dainst.de

Benoit Fabre
Sorbonne Universités, UPMC Univ Paris 06
LAM – Institut d'Alembert
4, place Jussieu
75252 Cedex 05, Paris, France
benoit.fabre@upmc.fr

Ichiro Fujinaga
McGill University
Schulich School of Music
555 Sherbrooke St. West
Montreal, QC H3A1E3, Canada
ichiro.fujinaga@mcgill.ca

Aaron Gibbings
The University of Western Ontario, Natural Sciences Centre
The Brain and Mind Institute, Department of Psychology
London, N6A 5B7, Canada
jgrahn@uwo.ca

Joël Gilbert
Université du Maine – CNRS
Laboratoire d'Acoustique
Avenue Olivier Messiaen
72085 Cedex 9, Le Mans, France
joel.gilbert@univ-lemans.fr

Nicholas Giordano
Auburn University
Dept. of Physics
Auburn, AL 36849, USA
njg0003@auburn.edu

Rolf Inge Godøy
University of Olso
Dept. of Musicology
0315 Oslo, Norway
r.i.godoy@imv.uio.no

Emilia Gómez
Universitat Pempeu Fabra
Music Technology Group
Roc Boronat 138
8018 Barcelona, Spain
emilia.gomez@upf.edu

Francisco Gómez
Technical University of Madrid
Dept. of Applied Mathematics
Escuela Técnica Superior de Ingeniería de Sistemas Informáticos Carret. Valencia Km 7
28031 Madrid, Spain
fmartin@etsisi.upm.es

Jessica Grahn
The University of Western Ontario
The Brain and Mind Institute, Dept. of Psychology
London, N6A 5B7, Canada

Sascha Grollmisch
Fraunhofer IDMT
Ehrenbergstr. 31
98693 Ilmenau, Germany
goh@idmt.fraunhofer.de

Simon Grondin
Université Laval
École de psychologie
2325 rue des Bibliothèques
Québec, G1V0A6, Canada
simon.grondin@psy.ulaval.ca

Peter Grosche
Huawei Technologies Duesseldorf GmbH
Riesstr. 25
80992 München, Germany
peter.grosche@huawei.com

Brian Hamilton
University of Edinburgh
Acoustics and Audio Group
Mayfield Rd.
Edinburgh, EH9 3JZ, UK
b.hamilton-2@sms.ed.ac.uk

Andrew Hankinson
University of Oxford
Bodleian Libraries
Osney One Building, Osney Mead
Oxford, OX2 0EW, UK
andrew.hankinson@bodleian.ox.ac.uk

Reginald Harrison
University of Edinburgh
Acoustics and Audio Group
Mayfield Rd.
Edinburgh, EH9 3JZ, UK
s0916351@sms.ed.ac.uk

Emi Hasuo
Tokyo Denki University
School of Information Environment
2-1200 Muzai-Gakuendai
270-1382 Chiba, Japan

Avraham Hirschberg
Zwartenberg 1
5508 AK, Veldhoven, The Netherlands

Neil S. Hockley
Bernafon AG
Morgenstr. 131
3018 Bern, Switzerland
neho@bernafon.com

Alexander Refsum Jensenius
University of Oslo
Dept. of Musicology
PB 1017 Blindern
0315 Oslo, Norway
a.r.jensenius@imv.uio.no

Wilfried Kausel
University of Music and Performing Arts, Vienna
Dept. of Musical Acoustics
Anton-von-Webern-Platz 1
1030 Vienna, Austria
kausel@mdw.ac.at

Christian Kehling
Neways Technologies
Fichtenweg 8
99098 Erfurt, Germany
christian@k-ling.de

Peter E. Keller
Western Sydney University
MARCS Institute for Brain, Behaviour and Development
Locked Bad 1797
Penrith, NSW 2751, Australia
p.keller@uws.edu.au

Stefan Koelsch
University of Bergen
Jonas Liesvei 91, BB-Bygget
5009 Bergen, Norway
koelsch@cbs.mpg.de

Nadine Kroher
Escuela Superior de Ingenieros, Universidad de Sevilla
Departamento de Matemática Aplicada II
Camino de los Descrubrimientos, s/n
41092 Sevilla, Spain
nkroher@us.es

Panos Kudumakis
Queen Mary University of London
School of Electronic Engineering and Computer Science
Mile End Rd.
London, E1 4NS, UK
panos.kudumakis@eecs.qmul.ac.uk

Tsuyoshi Kuroda
Shizuoka University
Faculty of Informatics
3-5-3 Johuku, Naka-ku
432-8011 Hamamatsu, Japan
tkuroda@inf.shizuoka.ac.jp

Marc Leman
Ghent University
IPEM – Musicology, Department of Art, Music and Theatre Sciences
Sint-Pietersnieuwstraat 41
9000 Ghent, Belgium
marc.leman@ugent.be

Alexander Lerch
Georgia Institute of Technology
Center for Music Technology
840 McMillan St.
Atlanta, GA 30332-0456, USA
alexander.lerch@gatech.edu

Micheline Lesaffre
Ghent University
IPEM – Musicology, Department of Art, Music and Theatre Sciences
Sint-Pietersnieuwstraat 41
9000 Ghent, Belgium
micheline.lesaffre@ugent.be

Jukka Louhivuori
University of Jyväskylä
Department of Music, Art and Culture Studies
Seminaarinkatu 15
Jyväskylä, 40014, Finland
jukka.louhivuori@jyu.fi

Håkan Lundström
Lund University
Inter Arts Center
Bergsgatan 29
21422 Lund, Sweden
hakan.lundstrom@kanslik.lu.se

Pieter-Jan Maes
Ghent University
IPEM – Musicology, Department of Art, Music and Theatre Sciences
Sint-Pietersnieuwstraat 41
9000 Ghent, Belgium
pieterjan.maes@ugent.be

András Miklós
Steinbeis Transfer Center Applied Acoustics
Weilstetter Weg 36
70567 Stuttgart, Germany
akustikoptik@t-online.de

Emilio Molina
Universidad de Málaga
Departamento de Ingeniería de Comunicaciones, ETSI Telecomunicación
Campus de Teatinos s/n
29071 Malaga, Spain
emm@ic.uma.es

Thomas Moore
Rollins College
Dept. of Physics
1000 Holt Ave.
Winter Park, FL 32789, USA
tmoore@rollins.edu

Joaquin Mora
Escuela Superior de Ingenieros, Universidad de Sevilla
Departamento de Matemática Aplicada II
Camino de los Descrubrimientos, s/n
41092 Sevilla, Spain
mora@us.es

Robert Mores
University of Applied Sciences
Faculty of Design, Media & Information
Finkenau 35
22081 Hamburg, Germany
robert.mores@haw-hamburg.de

Andrew C. Morrison
Joliet Junior College
Dept. of Natural Sciences
1215 Houbolt Rd.
Joliet, IL 60431, USA
amorriso@jjc.edu

Meinard Müller
International Audio Laboratories Erlangen
Am Wolfsmantel 33
91058 Erlangen, Germany
meinard.mueller@audiolabs-erlangen.de

Yoshitaka Nakajima
Kyushu University
Dept. of Human Science
4-9-1 Shiobaru, Minami-ku
815-8540 Fukuoka, Japan
nakajima@design.kyushu-u.ac.jp

Tram Nguyen
The University of Western Ontario
The Brain and Mind Institute, Dept. of Psychology
London, N6A 5B7, Canada
tnguye95@uwo.ca

Luc Nijs
Ghent University
IPEM – Musicology, Department of Art, Music and Theatre Sciences
Sint-Pietersnieuwstraat 41
9000 Ghent, Belgium
luc.nijs@ugent.be

Giacomo Novembre
University College London
Gower St.
London, WC1E6BT, UK
giacomonovembre@gmail.com

Chiara Olcese
University of Ferrara
Dept. of Life Sciences and Biotechnology
Via Riccardo Selvatico 10
31100 Treviso, Italy
chiara.olcese@student.unife.it

Bryan Pardo
Northwestern University
Ford Engineering Design Center
2133 Sheridan Rd.
Evanston, IL 60208, USA
pardo@northwestern.edu

Marcus Pearce
Queen Mary University of London
School of Electronic Engineering and Computer Science
Mile End Rd.
London, E1 4NS, UK
marcus.pearce@qmul.ac.uk

Florian Pfeifle
University of Hamburg
Institute of Systematic Musicology
Neue Rabenstr. 13
20354 Hamburg, Germany
florian.pfeifle@uni-hamburg.de

Laurent Pugin
Swiss RISM Office
Hallwylstr. 15
3000 Bern, Switzerland
laurent.pugin@rism-ch.org

Zafar Rafii
Gracenote
2000 Powell St., Ste 1500
Emeryville, 94608, USA
zrafii@gracenote.com

Martin Rohrmeier
TU Dresden
Institute of Art and Music
August-Bebel-Str. 20
01219 Dresden, Germany
martin.rohrmeier@tu-dresden.de

Carles Roig
Universidad de Málaga
ATIC Research Group, Dep. Ingeniería de Comunicaciones, ETSI Telecomunicación
Campus de Teatinos s/n
29071 Malaga, Spain
carles@ic.uma.es

Thomas D. Rossing
Stanford University
Dept. of Music
541 Lasuen Mall, MC:3076
Stanford, CA 94305-3022, USA
rossing@ccrma.stanford.edu

Mark Sandler
Queen Mary University of London
School of Electronic Engineering and Computer Science
Mile End Rd.
London, E1 4NS, UK
mark.sandler@qmul.ac.uk

Gary Scavone
McGill University
Music Research, Schulich School of Music
555 Sherbrooke St. West
Montreal, QC H3A1E3, Canada
gary@music.mcgill.ca

Albrecht Schneider
University of Hamburg
Institute of Systematic Musicology
Neue Rabenstr. 13
20354 Hamburg, Germany
aschneid@uni-hamburg.de

Katrin Schulze
Heidelberg University
Dept. of Clinical Psychology and Psychotherapy, Institute of Psychology
Hauptstr. 47 – 51
69117 Heidelberg, Germany
katrin.schulze@psychologie.uni-heidelberg.de

Anthony Seeger
University of California Los Angeles (UCLA)
Dept. of Ethnomusicology
2 Kimber Ridge Ct.
Annapolis, MD 21403, USA
aseeger@arts.ucla.edu

Mohamed Sordo
University of Miami
Center for Computational Science
1320 S. Dixie Highway Suite 600.15, Loc: 2965
Coral Gables, FL 33146, USA
msordo@miami.edu

Lorenzo J. Tardón
Universidad de Málaga
Departamento de Ingeniería de Comunicaciones, ETSI Telecomunicación
Campus de Teatinos s/n
29071 Malaga, Spain
lorenzo@ic.uma.es

Mari Tervaniemi
University of Helsinki
Cicero Learning and Cognitive Brain Research Unit
Helsinki, 00170, Finland
mari.tervaniemi@helsinki.fi

Leslie Tilley
Massachusetts Institute of Technology
77 Massachusetts Ave.
Cambridge, 02139, USA
tilley@mit.edu

Alberto Torin
University of Edinburgh
Acoustics and Audio Group
Mayfield Rd.
Edinburgh, EH9 3JZ, UK
s1164558@sms.ed.ac.uk

George Tzanetakis
University of Victoria
Dept. of Computer Science
3801 Finnerty Rd.
Victoria, V8W 2Y2, Canada
gtzan@cs.uvic.ca

Edith Van Dyck
Ghent University
IPEM – Musicology, Department of Art, Music and Theatre Sciences
Sint-Pietersnieuwstraat 44
9000 Ghent, Belgium
edith.vandyck@ugent.be

Doug Van Nort
York University
Computational Arts and Theatre & Performance Studies
4700 Keele St.
Toronto, M3J1P3, Canada
dvnt.sea@gmail.com

Michael Vorländer
RWTH Aachen University
Institute of Technical Acoustics
Kopernikusstr. 5
52074 Aachen, Germany
mvo@akustik.rwth-aachen.de

Chris Waltham
University of British Columbia
Dept. of Physics & Astronomy
6224 Agricultural Rd.
Vancouver, BC V6T 1Z1, Canada
cew@phas.ubc.ca

Ron Weiss
Google Inc.
111 8th Ave.
New York, NY 10011, USA
ronw@google.com

Victoria Williamson
University of Sheffield
34 Leavygreave St.
Sheffield, UK
v.williamson@sheffield.ac.uk

Shigeru Yoshikawa
1-27-22 Aoyama
818-0121 Dazaifu, Japan
shig@lib.bbiq.jp

Tim Ziemer
University of Hamburg
Institute of Systematic Musicology
Neue Rabenstr. 13
20354 Hamburg, Germany
tim.ziemer@uni-hamburg.de

Contents

Part D Psychophysics/Psychoacoustics

Part E Music Embodiment

Part F Music and Media

List of Abbreviations

k-NN	*k*-nearest-neighbor
1-D	one-dimensional
1C	one's complement
2-D	two-dimensional
2C	two's complement
3-D	three-dimensional

A

A/D	analog-to-digital
A1	primary auditory cortex
AAF	anterior auditory field
AAF	advanced authoring format
ABR	auditory brainstem response
AC	autocorrelation
ACC	anterior cingulate cortex
ACE	acoustic conversion efficiency
ACF	autocorrelation function
AD	Alzheimer disease
ADC	analog-to-digital converter
ADM	adaptive delta modulation
ADPCM	adaptive differential pulse code modulation
ADSR	attack, decay, sustain, release
ADU	analog-to-digital unit
AEP	auditory evoked potential
AF	auditory filter
AI	active intensity
AI	artificial intelligence
AIFF	audio interchange file format
aKE	affected KE family (FOXP2 mutation)
ALM	adaptive logic module
AM	amplitude modulation
AN	auditory nerve
ANN	artificial neural network
AoIP	audio over IP
AoM	area of motion
AP	absolute pitch
APS	artifact-related perceptual score
AR	autoregressive
ARMA	autoregressive moving average
ARQ	automatic repeat request
ASBF	structured audio sample bank format
ASCII	American standard code for information interchange
ASIC	application-specific integrated circuit
ASSP	application-specific standard part
ASW	apparent source width
ATN	augmented transition network
AuP	auditory pathway
AVI	audio video interleaved
AWB	audio workbench

B

BA44	Brodmann area 44
BBS	Baseler Befindlichkeits-Skala
BC	boundary condition
BD	book-dependent
BEM	boundary element method
BER	bit error rate
BF	best frequency
BHAD	blind harmonic adaptive decomposition
BI	book-independent
BIAS	brass instrument analysis system
BLAS	basic linear algebra subprograms
BM	basilar membrane
BMF	best modulation frequency
BOLD	blood-oxygen-level-dependent
BPM	beats per minute
BRAM	block RAM
BRECVEMA	brain stem reflex, rhythmic entrainment, evaluative conditioning, contagion, visual imagery, episodic memory, musical expectancy, and aesthetic judgment
BS	beam splitter
BWF	broadcast wave format

C

CAD	computer-aided design
CAM	common amplitude modulation
CAP	central auditory processing
CB	critical band
CC	corpus callosum
CCD	charge-coupled device
CCF	cross-correlation function
CD	compact disc
CE	computational ethnomusicology
CF	characteristic frequency
CFL	Courant–Friedrichs–Levy
CGM	corpus geniculatum mediale/medial geniculate body
CI	cochlear implant
CLB	configurable logic block
CM	comparative musicology
CMR	comodulation masking release
CN	cochlear nucleus
CNS	central nervous system
codec	coder-decoder
CoM	centroid of motion
CP	cochlear partition
CPS	closure positive shift
cps	cycles per second
CPU	central processing unit
CR	critical ratio
CT	corpus trapezoideum

CT	center time
CTF	cochlear transfer function
CV	computer vision
CWMN	common Western music notation

D

DA	distribution amplifier
DAC	digital-to-analog converter
DARMS	digital alternative representation of music scores
DASH	digital audio stationary head
DAT	digital audio tape
DAW	digital audio workstation
DBN	dynamic Bayesian network
DCC	digital compact casette
DCN	nucleus cochlearis dorsalis
DCT	discrete cosine transformation
DDR	double data rate
DE	differential evolution
DES	differential emotions scale
DESPI	decorrelated electronic speckle pattern interferometry
DFT	discrete Fourier transformation
DirAC	directional audio coding
DIY	do-it-yourself
DL	difference limen
DM	delta modulation
DMIF	delivery multimedia integration framework
DML	distributed-mode loudspeakers
DOP	data-oriented parsing
DPCM	differential pulse code modulation
DPOAE	distortion product otoacoustic emission
DRM	digital rights management
DSP	digital signal processing
DTS	digital theatre system
DTT	distorted tunes test
DTW	dynamic time warping
DVB	digital video broadcast

E

EAN	early anterior negativity
ECG	electrocardiogram
EDT	early decay time
EEG	electroencephalogram/ electroencephalography
EEPROM	electrically erasable programmable read-only memory
EKF	extended Kalman filter
ELAN	early left anterior negativity
EMG	electromyogram
EMI	electromagnetic interference
EMI	experiments in musical intelligence
EPROM	erasable programmable read-only memory
ERAN	early right anterior negativity
ERB	equivalent rectangular bandwidth
ERF	event-related field
ERN	error-related negativity
EROSS	easily removable, wireless optical sensor system
ERP	event-related potential
EsAC	Essen associative code
ESM	experience sampling method
ESPI	electronic speckle pattern interferometry
ET	equal temperament

F

FA	factor analysis
FD	finite difference
FDDI	fiber-distributed digital interface
FDM	finite-difference method
FDTD	finite-difference time domain
FEC	forward-error correction
FEM	finite element method
FFR	frequency-following response
FFT	fast Fourier transform
FFTW	fastest Fourier transform in the West
FHT	fast Hadamard transformation
FIFO	first in/first out
FIR	finite impulse response
FLAC	free lossless audio codec
FM	frequency modulation
FMC	FPGA mezzanine card
fMRI	functional magnetic resonance imaging
FPGA	field programmable gate array
FRF	frequency response function
FSM	finite-state machine
FSR	force-sensing resistor
FT	Fourier transform
FTC	frequency-threshold curve
FTD	frontotemporal degeneration
FTP	file transfer protocol
FTT	Fourier time transformation
FWHM	full width at half maximum

G

GDIF	gesture description interchange file format
GEMS	Geneva emotional music scale
GKSO	Gustafsson Kreiss Sundstrom Osher
GME	middle-ear pressure gain
GMS	gesture motion signal
GOFAI	good, old-fashioned artificial intelligence
GPGPU	general purpose graphics processing unit
GPL	general public license
GSR	galvanic skin response
GTTM	generative theory of tonal music
GUI	graphical user interface

H

HCI	human–computer interaction
HDL	hardware description language

HDMI	high-definition multimedia interface
HLS	high-level synthesis
HMM	hidden Markov model
HMN	human mirror neuron
HNR	harmonic-to-noise ratio
HOA	higher-order ambisonics
HPC	high-performance computing
HRIR	head-related impulse response
HRTF	head-related transfer function
HRV	heart rate variability
HT	half tone
HTML	hyper-text markup language

I

I/O	input/output
I2S	inter-IC sound
IC	inferior colliculus
IC	information content
ICA	independent component analysis
ICC	central nucleus of the inferior colliculus
IF	instantaneous frequency
IFF	interchange file format
IHC	inner hair cell
IID	interaural intensity difference
IIR	infinite impulse response
IL	interface layer
ILD	interaural level difference
IM AF	interactive music application format
IMUTUS	interactive music tuition system
INA	Ideale Nierenanordnung
IOI	interonset interval
IP	intellectual property
IP	internet protocol
IPF	impulse pattern formulation
IPL	inferior parietal lobule
IPS	intraparietal sulcus
IPS	interference-related perceptual score
ISA	independent subspace analysis
ISI	interstimulus interval
ISIH	interspike interval histogram
ISO-BMFF	ISO based media file format
ISRC	international standard recording code
ISS	informed source separation
ITD	interaural time difference

J

JI	just intonation
JND	just-noticeable difference

K

K–H	Kirchhoff–Helmholtz
KdRV	Critique of Pure Reason
KF	Kalman filter
KL	Kullback–Leiber divergence
KOPRA-M	Entwicklung und empirische Validierung eines Modells musikpraktischer Kompetenzen

L

LA	line array
LAN	left anterior negativity
LDV	laser Doppler vibrometry
LEV	listener envelopment
LF	lateral energy fraction
LFE	low-frequency effect
LHA	latent harmonic allocation
LMA	Laban movement analysis
LP	linear prediction
LPC	late positive component
LTAS	long-term average spectrum
LTI	linear time-invariant
LTM	long-term memory
LUT	look-up table

M

MAC	medium access frames
MAC	multiply-and-accumulate
MADI	multichannel audio digital interface
MAP	multiactuator panel
MAP	maximum a posteriori
MCML	motion capture markup language
MDL	minimum description length
MDS	multidimensional scaling
MEG	magnetoencephalography
MEMS	micro-electric mechanical system
MEP	motor evoked potential
MF	missing fundamental
MFCC	Mel-frequency cepstral coefficient
mid-DLPFC	mid-dorsolateral prefrontal cortex
MIDI	musical instrument digital interface
MIR	music information retrieval
MIS	University of Iowa musical instrument samples
MKSA	meter, kilogramm, second, ampere
MLP	multilayer perceptron
MLS	maximum length sequence
MMN	mismatch negativity
MMNm	magnetic mismatch negativity
MMSE	mini-mental state examination
MMSE	minimum mean square error
MOP	maximal outerplanar graph
MoR	model routing layer
MPE	multipitch estimation
MPM	music paint machine
MPML	multimodal presentation markup language
MPO	maximum power output
MPT	music perception test
MRI	magnetic resonance imaging
MSD	mass-spring-damper
MTL	medial temporal lobe

MusicXML	music extensible markup language
MXF	material exchange format

N

NAH	near-field acoustic holography
NB	naive Bayes
NCD	normalized compression distance
NFS	network file system
NICA	nonnegative independent component analysis
NIFF	notation interchange file format
NIME	new interfaces for musical expression
NLL	N. Lemniscus lateralis
NMF	nonnegative matrix factorization
NN	neural network
NONCE	novel, optionally novel, nondeterministic process, constraint, existing element
NTF	nonnegative tensor factorization
NTT	number theoretic transform

O

OAE	otoacoustic emission
OCB	olivocochlear bundle
OCR	optical character recognition
OCT	optimized cardioid triangle
ODD	one document does it all
ODS	operating deflection shape
OHC	outer hair cell
OHCI	open host controller interface
OI	optical imaging
OMF	open media framework interchange
OMR	optical music recognition
OPS	overall perceptual score
OPSI	optimized phantom source imaging
OSC	open sound control
OSI	open standard interconnection
OSPL90	output sound pressure level with 90 dB SPL input
OW	oval window

P

PA	public address
PAF	posterior auditory field
PANAS	positive and negative affect scale
PC	personal computer
PCA	principal component analysis
PCAP	peristimulus compound action potential
PCIe	peripheral component interconnect express
PCM	pulse code modulation
PCP	pitch class profile
PD	Parkinson's disease
PDE	partial differential equation
PDF	probability density function
PEASS	perceptual evaluation methods for audio source separation
PET	positron emission tomography
pFMC	posterior frontomedial cortex
PLCA	probabilistic latent component analysis
PLD	programmable logic device
PM	phase modulated
PMC	premotor cortex
PML	performance markup language
POF	polymer optical fiber
POMS	profile of mood states
PP	planum polare
PPM	peak program meter
PROMS	programmable read-only memory
PSA	prior subspace analysis
PST	poststimulus time
PT	planum temporale
PTC	psychophysical tuning curve

Q

QEF	quasi-error-free
QoL	quality of life
QoM	quantity of motion
QoS	quality of service

R

R	referent
RA	radiatio acustica
RAAM	recursive auto-associative memory
RAM	random access memory
RANN	recurrent artificial neural network
RBM	restricted Boltzmann machine
RCD	radial convergence diagram
REPET	repeating pattern extraction technique
RESTFT	recursive exponential short-time Fourier transform
RI	reactive intensity
RIFF	resource interchange file format
RMS	root mean square
RP	relative pitch
RPCA	robust principal component analysis
RSSM	rhythm self-similarity matrix
RSV	reserved
RTCP	real-time transport control protocol
RTL	register transfer level
RTP	real-time protocol

S

SACD	super audio CD
SAM	self-assessment manikin
SAM	speckle-averaging mechanism
SAOL	structured audio orchestra language
SATA	serial advanced technology attachment
SBR	spectral band replication
SCR	skin conductance response
SD	standard deviation
SD	semantic dementia
SDI	serial digital interface

SDIF	sound description interchange format
SDM	sigma-delta modulation
SDS	sample dump standard
SF	spectral flux
SFS	sound field synthesis
SHS	subharmonic pitch summation
SI-PLCA	shift-invariant probabilistic latent component analysis
SM	systematic musicology
SMA	supplementary motor area
SMG	supramarginal gyrus
SMO	sequential minimal optimization
SMS	sensorimotor synchronization
SMS	spectral modeling synthesis
SMTP	simple mail transfer protocol
SOC	superior olivary complex
SOM	self organizing map
SPI	serial peripheral interface
SPINET	spatial pitch network
SPL	sound pressure level
SPOAE	spontaneous otoacoustic emission
Spt	Sylvian-parietal-temporal
SR	spontaneous rate
SSEP	steady-state-evoked potential
SSM	self-similarity matrix
SSR	state-space representation
SSSPV	stochastic state-space phase vocoder
STFT	short-term Fourier transform/short-time Fourier transform
STG	superior temporal gyrus
STM	short-term memory
SV	Sonic Visualiser
SVM	support vector machine
SVTF	stapes footplate velocity

T

TC	tonal center
TCP	transmission control protocol
TEN	threshold equalizing noise
TEN (HL)	threshold equalizing noise test in dB HL
TEOAE	transient evoked otoacoustic emission
TFLOPS	tera floating point operations per second
THD	total harmonic distortion
TIE	total intonation error
TMACS	tera multiply-and-accumulates per second
TMS	transcranial magnetic stimulation
TOJ	temporal order judgment
TPS	target-related perceptual score
TU	transmission unit

U

ucf	user constraint file
UCL	uncomfortable loudness level
UDC	uniform discrete cepstrum
UDP	user data protocol
UHF	ultra high frequency
USB	universal serial bus

V

VAME	verbal attribute magnitude estimation
VCN	nucleus cochlearis ventralis
VCO	voltage-controlled oscillator
VHDL	very high speed integrated circuit hardware description language
VLPFC	ventrolateral prefrontal cortex
VoIP	voice over IP
VU	volume unit meter

W

WAN	wide-area network
WAV	waveform audio file format
WCLK	word clock
WDF	wave digital filter
WDRC	wide dynamic range compression
WFA	wave field analysis
WFS	wave field synthesis
WM	working memory
WMA	Windows media audio

X

XMF	extensible music format
XML	extensible markup language

1. Systematic Musicology: A Historical Interdisciplinary Perspective

Albrecht Schneider

A brief account of the historical development of systematic musicology as a field of interdisciplinary research is given that spans from Greek antiquity to the present. Selected topics cover the rise of music theory from the Renaissance to modern times, the issue of *harmonic dualism* from Zarlino and Rameau to the 20th century and the controversy about *physicalism* versus *musical logic* in music theory. Sections of this chapter further relate to the notion of a *system* and the concept of *systematic* research (which is exemplified with respect to the work of Chladni, Helmholtz, Stumpf, and Riemann), and to the concept of Gestalt quality that spawned contributions to music perception from Gestalt psychology. In addition, some developments in music psychology outside the Gestalt movement as well as in the sociology of music are sketched, followed by a paragraph on modern research trends, which include semiotic, computational, and linguistic approaches to music perception and cognition as well as contributions from the neurosciences. A final paragraph provides some of the background that led to establishing systematic musicology as an academic discipline around 1870–1910, from where further disciplinary and scientific developments defined the field in the 20th century.

1.1 Systematic Musicology: Discipline and Field of Research

The term systematic musicology denotes a scientific discipline as well as a large and interdisciplinary area of research. While the academic discipline of systematic musicology was established only in the 20th century in universities, research in this field dates back far into history. In this chapter, a brief account of several stages in the historical development of research pertaining to systematic musicology will be given. It goes without saying that, within the frame of a chapter, this account cannot even attempt to cover every relevant fact and person. The focus here is on selected areas where certain problems have either been approached for the first time, or where a theoretical and methodological framework has gained the role of a *paradigm* for some period (as, for example, Gestalt theory or musical semiotics; see below). Moreover, there are problems that tend to resist a *final* solution (which seems a condition typical for science at large [1.1]).

Historical developments for some subfields of systematic musicology have been covered in comprehensive works (such as music theory in the ten vols of the *Geschichte der Musiktheorie* [1.2] and in a concise volume edited by *Christensen* [1.3]). However, there is a host of new concepts and methodologies issued in research related to sound and music production as

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_1

well as to music perception, cognition, and social reception due to developments in electronics, computer and media technology, and also in areas such as the neurosciences. Publications resulting from research efforts in various fields need to be evaluated in regard to theoretical foundations, methodology, and factual outcome (as in this Handbook). *Elschek* [1.4] provides a systematic overview that covers much of the research in musicology and neighboring disciplines from the 19th century to about 1970–80.

1.2 Beginnings of Music Theory in Greek Antiquity

Scholars of Greek antiquity conducted a huge amount of research on the fundamentals of tone systems and scales using an approach that combined mathematical reasoning with observation and measurement (see [1.5] and the collection of sources and commentaries in [1.6, 7]). This scientific effort, which probably had antecedents in Mesopotamia and Egypt (as well as possibly some parallels in early China), usually is traced back to Pythagoras of Samos and his pupils. The so-called Pythagoreans (as Platon addressed them) pursued studies in arithmetic, geometry, astronomy, and music theory (besides, they were interested in philosophical, religious and political issues). Though the works of Philolaus (6th century BCE) and of Archytas of Tarentum (who lived from around 420–355 BCE, see [1.8]) survived only in fragments, their central ideas apparently were known to the music theorist Aristoxenus of Tarentum (4th century BCE, see [1.6]) and were still referred to by scholars of the Hellenistic era like Ptolemy (Klaudios Ptolemaeus) and Nicomachus of Gerasa (as to their works, see [1.6, 7]). Essentials of Pythagorean mathematics and music theory are also found in the writings of Euclid. Of particular interest is Euclid's *Sectio canonis* in which he, among other issues, explains that (pairs of) consonant tones form a *mixture* while dissonant tones lack this quality [1.9]. Also of major interest is that Euclid treats intervals like line segments (several of which can be added up or, conversely, subtracted from each other). While Archytas probably explored pitches with a pan pipe-like instrument, a kanon or monochord apparently was in use as a device for measurements around 300 BCE [1.10].

Greek mathematics and music theory rested on numbers representing both ideal magnitudes and measurable quantities. Ratios of whole numbers in general were used to establish intervals and scales. Pairs of ratios form proportions like $6:8=9:12$ known as a Pythagorean Tetraktys, which contains the intervals of the octave $1/2$, the fifth $2/3$, the fourth $3/4$, and the whole tone $8/9$. Ratios considered as fundamental for harmony could either contain multiples of a basic number like $(2/1, 3/1, 4/1)$, or be formed according to $m+1/m$ (the class of superparticular intervals $3/2, 4/3, 5/4, 6/5, 7/6, 8/7, 9/8, \ldots$). In addition, there were so-called superpartientes (ratios of the form $(n+1+m)/n$ like $5/3$ considered in particular by a range of medieval scholars [1.11, pp. 64–87]). Besides such arithmetic there was also a geometrical approach based on the division of a line or string (or of several such strings arranged in parallel, as Ptolemy confirms) into sections. In this respect, considerations based on ratios of numbers could always be tested empirically by dividing strings so as to sound the basic musical intervals and scales. Plato (Pol. 530 C–531 C, ed. by *Burnet* [1.12]), who opposed empiricism, in fact let Socrates refer to the Pythagoreans as people who torture strings by working on their tension with tuning pegs. Archytas it seems already distinguished between the arithmetic, the geometric and the harmonic division (that can be applied either to an interval defined by numbers or to sectioning a line or string of given length). Though observation was used as a method for scientific investigations, mathematical reasoning among the Pythagoreans was regarded superior for its logical coherence and general applicability.

Since there evidently was music performed in various contexts in Greek societies (as many paintings and engravings demonstrate), the question of course is to what extent fundamental structures described by theorists (such as different divisions of the tetrachords, see Sect. 31.8 of this handbook) were actually found in musical practice. On the one hand, there are indications that theorists like Archytas were familiar also with musical practice [1.8]; on the other, a comprehensive treatise on music theory like the *Elementa harmonica* provided by Aristoxenus of Tarentum (4th century BCE [1.6]) gives many hints that musical practice at his time did not (or perhaps no longer) conform in every respect to norms and models that had been set up mainly by the Pythagorean school. One of the fundamental objections Aristoxenus made is that musical experience based on performance and listening may differ from, or may even contradict, certain models and propositions developed from arithmetic.

Much of classical Greek music theory survived in writings of the Hellenistic period, in particular in the works of Claudius Ptolemaeus and Nicomachus of

Gerasa (both lived in the 2nd century [1.6]). Boethius (around 480–528), who was familiar with a broad range of Greek and Hellenistic writers, condensed their works in his own treatise *De institutione musica* [1.13], which became the most important source for medieval music theory that, for the most part, was still close to Pythagorean tradition [1.11]. In the realm of musica speculativa, scholars throughout the Middle Ages and up to about the 15th century were contemplating correspondences between ratios and proportions of numbers and musical intervals; in addition, various divisions of the monochord were discussed (not just from a theoretical but sometimes from a clearly empirical point of view). An impressive survey of speculative music theory (where speculative has to be taken in the medieval sense) was provided by Jacob of Liège (*Jacobus Leodiensis*, around 1330; see the critical edition [1.14]). Phenomena of polyphonic music such as the organum were treated in writings on music theory as early as the *musica enchiriadis* (9th century).

1.3 From the Middle Ages to the Renaissance and Beyond: Developments in Music Theory and Growth of Empiricism

While much of medieval music theory reflects its Pythagorean origin and background, issues relevant to musical practice and organology gained increasing importance. One area where a transition from basically mathematical to empirically validated approaches can be shown is the mensuration of organ pipes [1.15, 16]; another is music theory itself where more *practical* views on musical syntax were included along with the time-honored considerations of consonance based on small integer ratios [1.17]. By the end of the 15th century, treatises on musica practica appear besides musica theorica. Within the domain of theory, where traditionally tone systems, scales and modes had been elaborated, an expansion of topics as well as a discussion adapted to developments in contemporary musical practice is observed (for a detailed account [1.18]). For example, expansion of the hexachord system as well as of the traditional modes (as in Glarean's *Dodecachordon* of 1547) should be mentioned. At about the same time, one had to provide tunings for keyboards and fretted instruments suited to performing music in acceptable temperaments (such as 1/4-comma meantone or similar tunings [1.19–21]). *Schlick* [1.22, Chap. 8] discusses how to tune an organ where the fifths have to be narrowed a little bit so as to get rid of the comma. Tempering the fifths is already controlled by audible beats. Also, composers and theorists in the 16th century (and well into the 17th) experimented with tone systems that covered the chromatic and the enharmonic genus. They had special instruments designed that offered 17, 19 or even 31 keys per octave [1.21, 23]. The intervals thus available were assessed not only in regard to their musical use but also with respect to sensory qualities such as consonance and dissonance. Zarlino's comprehensive works on music theory (including aspects of organology, tunings, etc.) must be regarded as milestones in the history of systematic musicology. *Zarlino* [1.24, Part III, Chap. 31] states that the arithmetic and the harmonic division of the fifth yield the (just) major and the (just) minor third respectively, from which he concludes that this distinction is the ultimate cause of harmony. Though *Zarlino* discusses intervals such as $2/1, 3/2, 5/4, 6/5$ [1.24, Part III, Chap. 32], one in fact can derive the major and the minor chord from the operations Zarlino proposes for the two divisions of the fifth, and later theorists like Rameau and Riemann therefore saw Zarlino's ideas as decisive steps on the way to modern tonality. As a matter of fact, Zarlino elaborated vertical harmony in his own polyphonic works (*Modulationes sex vocum*, 1566).

In the 16th century, progress was also made in musical acoustics when Fracastoro (1546) correctly explained the nature of resonance in strings (while the phenomenon as such, i. e., sympathetic vibration, was known long before [1.25]). Instrument building had reached a high level of invention and craftsmanship by about 1600, and many types were in use, for example of reed instruments (as is obvious from the account *Praetorius* [1.26, T. II] gives). As *Praetorius* [1.26, T. III, pp. 126f.] reports, in Italy organization of a range of string, woodwind and brass instruments into a Chorus instrumentalis alternating or playing together with the Chorus vocalis (in concert-like form) was part of musical practice.

With the rise of mechanics, the theory of vibration and the concept of frequency became areas of research based on empirical observation and mathematical scrutiny (for detailed accounts covering the period from around 1580–1750 see [1.27, 28]). *Mersenne* [1.29] found a formula for the calculation of frequencies of vibration and worked on absolute frequencies; he was familiar with a host of musical instruments including the enharmonic experiments Giambattista Doni and others were conducting in Italy. In the 17th century, logarithms for calculating musical intervals and temperaments were developed [1.30]. An early attempt to

expand the theory of consonance from ratios of numbers or string sections to actual sound was based on the coincidence of pulses assumed to be emitted into the air from vibrating strings (or air columns in pipes), and received in our ears as series of small *strokes* hitting the tympanum. This approach was taken, in about 1620, by Isaac Beeckman, who had formal training in medicine and understood the basic function of the tympanum and the middle ear ossicles. He apparently even knew the oval window that he, however, took as the entrance to the auditory nerve (the microanatomy of the cochlea was studied, by Alfonso Corti who in 1850 discovered the spiral structure of the receptor organ for hearing inside the bony part of the cochlea known since as Corti's organ). Though Beeckman's *atomistic* approach to sound was flawed (he opted for a corpuscular concept instead of a wave theory; for a detailed account of Beeckman see [1.27]), and his ideas, which he noted in his diaries, were not published before the 1940s (but had been shared with fellow scientists Descartes and Mersenne), the coincidence theory offered a physical explanation of consonance that could be tested as well as modified along with new observations. From this theory, Beeckman also understood the nature of roughness and beats. While it was known for a long time that each musical tone contained at least two components (in modern terminology, the fundamental and its octave, the second harmonic), Mersenne it seems could hear out several harmonics, which he identified as fundamental, octave, tenth and twelfth. *Mersenne* [1.29, L. VII, prop. XVIII] believed he could hear these partials also in the sound of bells, however, he recognized there were possibly some deviations from this scheme. Jacob van Eyck, an outstanding musician and musical scholar working at Utrecht, according to a note contained in the diary of Isaac Beeckman (24th of Sept. 1633) had been able to hear out some of the prominent partials in bells without touching the bell [1.31, pp. 310–11]. Apparently, van Eyck used to sing or whistle tones in order to excite a resonance in a vibrating body like a bell.

Once acoustics was introduced into musical science by Beeckman, Mersenne and also by Christiaan Huygens (who had more of a musical background than the aforenamed scholars), it was clear that sound as actually produced in performance and perceived by listeners was as important for music as were rules of counterpoint and harmony. However, one has to take developments in musical composition as well as in music theory into account. For example, intervals regarded as imperfect consonances in medieval treatises (the two thirds and the two sixths) were actually employed in composition already before 1500, and massively so in the 16th century when settings in four voices had become customary. Madrigals composed by Cipriano de Rore and, by the end of the 16th century, Luca Marenzio (in particular his fifth and sixth book of madrigals) demonstrate ample use of chords as vertical sonorities, and Monteverdi, in his fifth book of madrigals (1605), finally offered a major chord including the *dominant seventh* dissolving into the major third of the tonic (*Cruda amarilli*, m. 41–42). Soon, both works for keyboards (as, for example, the *Toccata quinta per il cimbalo cromatico* of Ascanio Mayone, published in his Secondo libro di diversi capricci per sonare, Napoli 1609) and the bold textures of sonorities Gesualdo created in his fifth book of madrigals (1611) became truly adventurous in exploring harmony. These explorations went along with experiments in tunings and temperaments as well as in building enharmonic instruments [1.20, 21, 32] designed by Mersenne (1636) and many other music scholars.

Though it is beyond the scope of this chapter to discuss developments in music and music theory in any detail, it seems safe to say that music theory from the 15th century onward comprised several large areas whereby the medieval division of *musica speculativa* and *musica practica* was maintained, in some respects (which can be traced in works such as Gaffori's and Zarlino's, and still around 1700 in, for example, Werckmeister's publications). While the theoretical and speculative approach was continued by scientists like Kepler and Mersenne [1.27], composition, performance practice and also teaching demanded works with a more *practical* orientation (for overviews covering developments in France, Italy, England, Germany and neighboring countries, see vols. 6, 7, 8/1, 8/2 and 9 of the *Geschichte der Musiktheorie*, 1986–2003 [1.2]). It was also from practical considerations that organ building had introduced pipe ranks an octave or two apart (like 16′ and 4′, see [1.22] and [1.16, pp. 89ff.]). Doubling of voices at the octave appears to have been in use not only in organ music but also in other musical genres (as testified by *Praetorius* [1.26, Vol. III, pp. 132ff.], see [1.33, pp. 445ff.]). Equivalence of octaves as tones and pitches thus was a common experience. Another such experience must have been that of harmonic triads (termed *trias harmonica* by *Lippius* [1.34]) especially when, not long after around 1600, musical genres with a basically homophonic structure were propagated and sequences of chords were arranged according to basso continuo rules. Inversion of chords is one of the topics elaborated in a number of 17th-century books on *practical theory*.

1.4 Sauveur, Rameau and the Issue of *Physicalism* in Music Theory

In 1687, Newton's *Philosophiae naturalis principia mathematica* appeared, which offered a coherent rational basis for mechanics. In 1677, John Wallis had published a paper on harmonic vibration and resonance in strings [1.28, pp. 44ff.]. By about 1700, *Sauveur* investigated harmonics in vibrating strings [1.35] and also discussed harmonics in organ pipes as well as the effect of superposition of several complex harmonic sounds [1.36]. He came to the conclusion that the use of several pipes from different organ stops gave the same basic harmony one could observe in vibrating bodies (*L'orgue ne fait qu'imiter par le mêlange des ses jeux, l'harmonie que la nature observe dans les corps sonores, qu'on appelle harmonieux*). Years after Sauveur had postulated that the harmonic series was the principle governing complex sounds, his findings were reported to *Rameau* (see [1.37, p. 3], [1.38, pp. 87f.], and [1.39, Chap. 6]). To be sure, *Rameau* had already stressed the unique role of the octave in regard to production and perception of pitches in his *Traité* [1.40, art. 3] where he points to overblowing a flute into the higher octave simply by increasing the blowing pressure. He also mentions a statement of Descartes (found in his *Musicae compendium*, written in 1618, and published in 1653 and 1656) according to which a (complex harmonic) tone is somehow always contained in his octave. Later, *Rameau* [1.41, p. 16] said:

> *je sçavois par experience que l'octave n'est qu'une replique, combien il y a d'identité entre les sons & leurs répliques, & combien il est facile de prendre l'un pour l'autre, ces sons même se confondant à l'oreille* [...]

Again, in a letter to Leonard Euler, which was published in part [1.42, pp. 167ff.], Rameau argued that tones one or several octaves apart blend perfectly since *la représentation d'un son dans ses octaves* gives a sensation of identity (*Identité*). This concept of functional and perceptual octave equivalence permitted Rameau to take the octave as a frame within which intervals can be inverted. Hinting at the commutative law (*Rameau* [1.40, p. 7] speaks of *raison ou comparaison renversée*), he pointed to a principle of *renversement* that became a major issue in his theoretical works since it could be used for deriving basic intervals as well as chords (e.g., [1.40, p. 10], [1.43, pp. 20f.]) from a given interval or triad. Octave equivalence and chord inversions played a distinctive role in functional harmony, and Rameau has been lauded to have made significant contributions to this field with his concepts of *son fundamental*, *basse fondamentale*, *dominante* and *sous-dominante*, *note sensible* (leading note) and other considerations [1.33, 38]. Moreover, he has been viewed as a proponent of so-called harmonic dualism for his derivation of both the major and the minor tonality from a set of basic principles.

Since harmonic dualism (a concept to be explained in brief below) was, and to some extent still is, a major issue in music theory [1.33, 44, 45], it seems apt to review at least some of the background. When *Zarlino* [1.24, Part III, Chap. 31] discussed the harmonic and the arithmetic division of the fifth, it was clear to him that the two thirds he obtained could be combined in two different ways, which, by implication, yielded a major and a minor triad, i. e., the basic building blocks of harmony *(...& da questa varietà dipende tutta la diversità, et la perfettione delle Harmonie*). The harmonic and the arithmetic division hence serve as generators of either a major or a minor triad. Apart from calculation (which shows that the two means between 2 and 3 are either $12/5 = 2.4$ or $5/2 = 2.5$), one may find the major and minor third intervals from division of an actual string.

The concept of harmonic dualism claims, in the first place, that major and minor are two distinct types of tonal organization for which principles for the generation of intervals and chords must be stated. One method suited to derive relevant intervals in fact could be arithmetic and harmonic division of a string. Arithmetic division proceeds from sectioning a string into n equal parts ($n = 1, 2, 3, \ldots$). For example, division of a string into five equal parts gives the fundamental ($5/5$ of the string) and the major third (at $4/5$ of the string) as well as the major sixth ($3/5$); division into six equal parts yields the fundamental note (at $6/6$), the minor third above this tone ($5/6$), the fifth ($4/6$), and so on [1.44, pp. 16ff.]. The tones from a division into five parts relative to C_4 ($5/5$), are E_4 ($4/5$), A_4 ($3/5$), E_5 ($2/5$), E_6 ($1/5$). One may regard this series as representing a chord in terms of *undertones* in particular if the series is centered at E_6, from which the other tones viewed downwards are in the intervals of an octave, fifth, fourth, and major third (it is easy to continue this series to the minor third A_3 etc.). What can be demonstrated is that a series of *undertones* derived from operations of equidistant sections of a string and a series of *overtones* found from sounding the fractional length $1/n$ of a string (where $n = 1, 2, 3, 4, 5, \ldots$) yields two series that are complementary and symmetric [1.44]. Since tones/notes representing the major chord are contained in the series of *overtones* $4 : 5 : 6$, and the tones/notes representing the minor chord in the series of *undertones* respectively, one might be inclined

to take both as equivalent in structure. In a formal approach, this is correct because of the symmetry of both series (Fig. 1.1).

The problem, however, is that *overtones* in fact are contained in each complex tone produced from a vibrating string or air column (as in an organ pipe or trumpet), while *undertones* (under normal conditions) are not (electronic instruments such as the Trautonium finally offered sounds comprising subharmonics which, as fractions of a fundamental frequency f_1, are representing series of *undertones* like $1/2f_1$, $1/3f_1$, $1/4f_1$, etc.). At the time of Sauveur and Rameau, *overtones* were of interest since it was understood (since Mersenne, Beeckman, van Eyck, see above) such tones are contained in single sounds produced from a *corps sonore*. As these partials were in harmonic ratios (like 2 : 4 : 5 : 6 as pronounced by Mersenne), a general idea emerged of harmonic superposition relevant not just for single sounds, but also for combinations thereof. For instance, *Rameau* [1.37, pp. 13, 16] reports small experiments where he experienced tones from several organ pipe ranks (activated one after another or simultaneously) blending and fusing while the single tones could still be distinguished (it was from quite similar experiments involving organ mixture stops that Carl Stumpf explored his concept of *Verschmelzung* around 1880–90; see below). To be sure, an instrument well known at the time of *Sauveur* and Rameau was the trompette marine [1.35, pp. 300f., 356] which, in particular in versions developed and played by the virtuoso Jean-Baptiste Prin around 1700–1720 [1.25, pp. 313ff.], employed sympathetic strings (as did the Baryton and the viola d'amore also in use at that time). *Sauveur* [1.35, p. 356] understood that the *trompette marine*, due to the peculiar regime of vibration in its long string *ne produit que les sons harmoniques*, and that, in long strings, there were knots (*nœuds*) where no motion was observed while motion was maximal at the antinodes (*ventres d'ondulations*). Since a long string vibrates in total as well as in parts, there is a *son fondamental* as well as *sons harmoniques*. *Sauveur* [1.35, pp. 300f.] said that during the night (when it was quiet) one could hear, besides the *son principal* (the main pitch of a long string) some small tones (*d'autres petits sons qui étoient à la douzième & à la dix-septième de ce son).* What is meant, from the table *Sauveur* [1.35, p. 350] includes in his essay, is partials no. 1, 3, 5 of a harmonic series.

In his system of tones and intervals, *Sauveur* [1.35, pp. 302ff.] offers a symmetric arrangement of tones relative to a son fondamental into upper *octaves ou octaves aigues* as well as lower ones (*sous-octaves ou octaves graves*). It may well be that this symmetric arrangement became known to Rameau. After *Rameau*

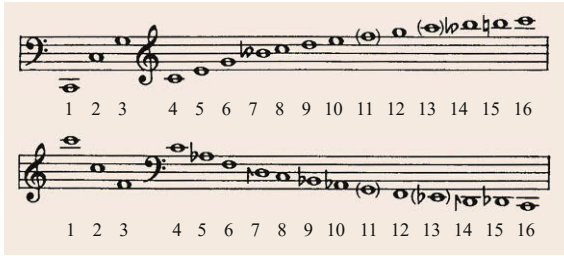


Fig. 1.1 Symmetry of *overtones* and *undertones*

had been informed about Sauveur's findings, he ([1.43, pp. 20ff.] and throughout [1.37]) set out his concept of harmony as a combination of elements known from Zarlino (the harmonic and arithmetic division of the fifth as a basis for the major and minor triad respectively) and Sauveur. Consequently, one finds the two proportions 1 : 3 : 5 and 1 : 1/3 : 1/5 as well as the concept of *son fondamental* and harmonic partials, and there are operations such as *renversement* and contraction of partials into one octave in order to establish symmetry needed for the juxtaposition of harmonic major and minor. Summing up his views, Rameau argued there are some fundamental intervals (unison, octave, fifth, major third) whose prominence is due to natural laws (namely, harmonic vibration), and that both harmonic major and minor can be grounded on proportions reflecting such laws. The harmonic proportion that *Rameau* [1.37, p. 29] states, in whole numbers, as 15 : 5 : 3 gives the *accord parfait* (put into one octave, the tones ut–mi–sol), but also the minor triad (viewed as the arithmetic proportion 6 : 5 : 4 . . . : 3; [1.37, p. 37], and examples I and II annexed to the book, also [1.41, pp. 20ff.]), though admittedly not just as perfect and *natural* than major [1.37, p. 32], because of its reciprocal relation to major is a construct backed at least by logical argument [1.37, Chap. IV]. As is well known (see [1.38], [1.39, Chap. 6]), Rameau's attempt at finding also an acoustical explanation for the minor triad in the pattern of sympathetic vibrations of strings tuned to harmonic ratios below a string actually set to motion, was flawed as far as sympathetic vibration of the lower strings in their full length (i. e., in their fundamental mode) is concerned. However, *Rameau* [1.37, p. 36f.] was cautious enough to suggest his reciprocal scheme of the major and minor triad seemed plausible for certain reasons; he did not claim it was verified by experimental observation. He later [1.41, pp. 21f. 63ff.] made it clear that strings tuned lower than that string actually played can only go into aliquot sympathetic vibration (a point underpinned also by [1.46] who had been supporting Rameau for a long time). *Rameau* [1.41, 47] maintained, though, that the major and minor mode should be conceptualized as forming

reciprocal relations (modeled after harmonic proportions).

Rameau offers a number of concepts that became essentials of music theory in the later 18th century and throughout the 19th century. One of course was his *basse fondamentale*; another the relation of a central tone (*le principe ut* [1.41]) to the major third and fifth above and below as well as the reciprocity of major and minor proportions. Further, there are tonal functions for certain tones such as the *dominante*, the *note tonique*, and the *note sensible* [1.40, p. 56]. The reciprocal construction of harmonic minor includes *la sousdominante* [1.37, pp. 132ff.] and another concept, termed by Rameau the double use (*double emploi*), affords that a chord f–a–c'–d' (played after c–e–g–c' as tonic), may be conceived, by inversion (*renversement*) also as d–f–a–c'. Hence, the chord of the *sousdominante* can appear in two different forms, and thereby can represent two different modes [1.41, 46]. However, taking f–a–c'–d' (relative to the initial C-major chord this is a major chord with sixte ajoutée) as equivalent to d–f–a–c' (which thus would be conceived as a minor chord with a minor seventh added on top) clearly is a reinterpretation (notwithstanding both resolve easily into a G-major chord, which in turn may be followed by the initial C-major to complete the cadence). Both *renversement* and reinterpretation of chords in regard to their tonal function played an important role in the theory of harmony ever since (the more so as equal temperament, proposed by Rameau in his later writings as well as by other music scholars around 1760–1800, became the standard tuning for pianos in the 19th century; see Sect. 31.7). Rameau's dualistic concept of major and minor was taken up by music theorists such as Moritz Hauptmann (who also adopted the concept of the octave, the fifth and the major third taken as *directly comprehensible* intervals) and Hugo Riemann. *Arthur von Oettingen* [1.48, 49] made a serious attempt to provide the acoustical and psychoacoustical foundations for harmonic dualism that Rameau could not offer [1.50, 51]. In order to give their approach to harmony foundations a base in natural laws, many theorists (like [1.52]) pointed to the harmonic series as a model for the major chord (e.g., c–e–g corresponds to the harmonics 4 : 5 : 6, while the first inversion e–g–c' would be equivalent to harmonics 5 : 6 : 8, and the second inversion G–c–e to harmonics 3 : 4 : 5) and also the minor triad (as represented by harmonic partials no. 10 : 12 : 15). A musicologist familiar with scientific methods [1.53] was critical of the adherents of *overtones* (he addressed them as *Obertönler*) and of *physicalism* in music theory in general since, as Hauptmann and Riemann had underpinned, harmony in music should be conceptualized in logical form (as an analogue to *well-formed* sentences or *propositions* [1.54, 55]) rather than as derived from sound structure. Though the logical and *propositional* approach to music theory is certainly relevant for the analysis of musical works (in particular as manifest in a symbolic notation but also as sound textures that are perceived by listeners more or less familiar with a musical syntax [1.56]), the disregard for actual sound properties of, for example, harmonic cadences and chord progressions was unfortunate. A new interest in psychoacoustic aspects of harmony [1.57–59] was spawned by empirical research in pitch perception including virtual pitch (Chap. 31) as well as by approaches to music cognition based on actual sound patterns. As far as the difference between major and minor is concerned, signal processing methods permit researchers to show that a significant difference in major and minor chords formed on the same fundamental (in just intonation) can be found in the subharmonic virtual pitches they give rise to [1.51]. Also, differences in the degree of spectral harmonicity can be found for major and minor chords from piano tones in ET12 (Fig. 1.2) when the periodicity of the signal is measured in the time domain.

1.5 Concepts of Systems and Systematic Research

It is obvious that the work of Greek and Hellenistic philosophers and scientists such as Aristotle, Euclid, and Ptolemy covered several if not many areas, and that they had both outstanding theoretical and factual knowledge. In certain respects, their knowledge gained what one may call systematicity since it was ordered according to first principles, integrated across individual observations, coherent and complete (as much as possible at a given time in the process of research). Aristotle discusses these criteria in regard to nature (physis) as well as science (theoria) directed toward the study of nature in the first place (see his lectures on Physics, in particular book II, and book X and XI of his Metaphysics). To understand the world (containing all of nature) as a whole, science also must be comprehensive. If nature in total is viewed as a system (or as a system of systems, as it is still in modern science) the research effort pursued by science will reveal more and more of the laws and categories governing nature. Thereby, a correspondence between nature (viewed as a large and

complex system) and science (viewed likewise as an increasingly complex and differentiated system) can be established. The science of nature by Aristotle *is seen as a systematic whole because it presents an account of the world that is similarly systematic* [1.60, p. 31].

The notion of a *system* and the principle of *systematicity* became constitutive for philosophy and the sciences to this day in a number of ways (for at least some of the background, see [1.50]). For example, if individuals are to acquire coherent and reliable knowledge, they proceed by integrating new information into a framework of (innate or learned) categories, and do so according to certain rules (as has been stressed, for language acquisition, by *Chomsky* [1.61]). One can draw a parallel between the process of cognitive systematization individuals effectively carry out over most of their lives in order to build up coherent knowledge of the *world* (as well as of themselves; the process is reflexive), and the collective endeavor of many scientists building up coherent knowledge of the *world* [1.1, 62]. Kant's epistemology (as set out in his Critique of Pure Reason (KdRV) and other writings, see *Rescher* [1.63]) permits such a parallel perspective since categories and principles fundamental to acquisition of coherent individual (ontogenetic) experience are fundamental also to structuring collective scientific knowledge. Kant himself, however, did not elaborate this approach for specific areas or disciplines of science. One of his few remarks on *system* (at the very end of KdRV) is that reason (Vernunft) is not additive yet integrative and rule-based so as to provide unity and coherence within the manifold of knowledge (KdRV B 860). Therefore, reason must rest on principles, and the scientific method must always be *systematic* (KdRV B 884) to achieve coherence.

It is of course outside the scope of this introductory chapter to delineate the concepts of *system* and *systematization* as they were developed in philosophy and in various areas of science within the 19th and 20th centuries. It is known that some philosophers, in particular in the 19th century, elaborated comprehensive *systems* of inquiry and reasoning (comprising logic, epistemology, philosophy of nature and other areas), and that Hegel's philosophy perhaps was the most advanced in regard to establishing a coherent yet in part speculative system derived from basic notions and principles. However, even in the 20th century, an approach from quite a different angle (that of so-called neopositivism [1.64], combined with elements from pragmatism and operationalism) explicitly called for a *Unified Science*. The movement, originating in Vienna in the 1920s and later centered in the United States, launched an encyclopedia as well as journals such as *Erkenntnis* and *Philosophy of Science*. Its members – for the most part philosophers, mathematicians, physicists, linguists and psychologists – published many works on theory and methodology of science. The idea still was to provide a framework suited to guide research in many areas so that coherent knowledge would be gathered and presented in an orderly fashion along the same basic principles. Another notable contribution to offering an integrative perspective was that of general system theory [1.65], which became influential in many fields (from the life sciences to economics).

Concepts of *system* and *systematization* played an increasingly important role in empirical science since, in the 18th and 19th centuries, factual and theoretical knowledge was expanding fast while a large number of fundamental principles governing both organic and inorganic nature were discovered (from mechanics and astronomy to organic chemistry and genetics). In addition, *evolution*, acknowledged first as a teleological principle [1.66] and then as biological fact [1.67], became accepted as perhaps the most fundamental *law* governing nature (including humans, their ways of living, as well as their arts and even their morals [1.68, Chaps. 4–7]). Growth of factual knowledge had led to ordering and classification of phenomena (e.g., the taxonomy of plants as worked out by Carl von Linné). A dynamic approach to systematization was established when, in the course of the 19th century, methods of classification and typology were combined with the principle of evolution. From this perspective, a number of *evolutionist* models were established not only in biology and physical anthropology, but also in prehistoric archeology and cultural anthropology. Humanities in general and musicology as an academic discipline, reestablished in Austrian and German universities around 1870, were to follow soon with concepts of so-called *Entwicklungsgeschichte* (developmental history) which, in most cases, offer Hegelian thoughts on progressive development mixed with ideas on *natural evolution* and with typological ordering of phenomena from simple to complex. Such series of types then are interpreted as representing *stages of evolution* (from *primitive* to *developed* and, historically, from *ancient* to *modern*). One has to remember, further, that in the 19th century the comparative method gained prominence in various areas of research (from anatomy of vertebrates to Indo-European linguistics) including, from about 1890 onward, comparative musicology [1.69].

1.6 Systematic Approaches: Chladni, Helmholtz, Stumpf, and Riemann

As outlined in the preceding paragraph, a systematic approach in scientific disciplines is one striving for coherent knowledge based on principles, categories, and empirical evidence needed for causal explanations. A serious attempt at presenting scientific as well as cultural knowledge that had been accumulated up to a certain time, in a concise format, was the *Encyclopédie* (1751–1772) edited by Diderot and d'Alembert with the support of famous collaborators. The *Encylopédie, ou dictionnaire raisonné des sciences, des arts et des métiers* includes an article on *system* (système, T. XV, Paris 1765, 777–781, written by d'Alembert, Le Blond, de Jaucourt and Rousseau) as well as an article on music (musique, T. X, 1751, 898–909, provided by Rousseau, de Jaucourt and Menuret), which stems from the established distinction between musique spéculative and musique pratique, and which is historical (with a focus on Greek antiquity) in its main orientation. The article on *system* covers areas from philosophy to astronomy and science in general as well as music (the musical part refers to structures of the Greek tonal system and later developments in a more precise and detailed manner than the article on music). While the authors maintain that the criteria of reasoning from principles and coherence are essential to a *system*, the article is clearly antispeculative while it underpins the need for experience and factual evidence *(Les experiences & les observations sont les matériaux des systems*). A system, in this sense, is an ordered presentation of facts along with principles suited to explaining their relationships as well as possible effects.

In 1802, Ernst Florens Chladni was the scholar that presented facts and principles for the field of acoustics in a truly *systematic* account. *Chladni*'s classic work [1.70] discusses generation of sound in vibrating bodies (such as rods and plates) as well as sound propagation in solids and in air. His book includes quite many chapters on auditory sensation (part 4) as well as on musical intervals, scales and temperaments (in part 1). He also invented a number of musical instruments suited to exploring sound properties in regard to pitch and timbre [1.71]. In 1852, Opelt issued a three-dimensional model of pitch (Chap. 31, Fig. 31.31), and at the same time *Marchese Corti* had made seminal discoveries of the anatomical structure of the cochlea and the auditory nerve [1.72]. Also, by about 1850 mathematical psychology [1.73] and psychophysics became an area of research that soon led to *Fechner*'s famous treatise [1.74].

Hermann (von) Helmholtz, a physician and physicist who, in 1855, was appointed professor of anatomy and physiology at the University of Bonn, picked up the findings of Corti in a public lecture where he elaborated on their implications for the theory of hearing and even for harmony in music [1.75]. This lecture discusses in a nutshell what a few years later would find extensive treatment in *Helmholtz*'s comprehensive book [1.76]: namely the interrelationship between fundamentals of vibration and sound, hearing as a sensory and perceptual process, and music as an art based on patterns of sound suited to communicate pitch structures, *sound colors* (German: *Klangfarben*, see Chap. 32 of this book) as well as temporal and dynamic parameters to subjects performing and/or listening to music (for an evaluation of Helmholtz's work in auditory theory and psychoacoustics see [1.77], [1.78, Part 1]). The impact of Helmholtz's *Tonempfindungen*, which saw six German editions and an English translation by Alexander John Ellis (*On the Sensations of Tone*) in two editions (the first, published in 1875, was based on the third German edition; the second, published in 1885, on the fourth German edition) as well as a number of reprints, can hardly be overstated. A decisive factor for the general acceptance and even popularity of the *Tonempfindungen* as well as of other works Helmholtz had published (e.g., his equally comprehensive books on physiological optics) was that Helmholtz reported a wealth of observations he had made himself, in experiments for which he even designed instruments for measurement. Moreover, his extensive research in central areas of physics, mathematics, and physiology allowed him to pursue an integrative approach to the sciences (as well as in related areas like musical acoustics and psychoacoustics). In addition, Helmholtz had a thorough understanding of epistemology and methodology, from which he developed his own theoretical framework in regard to sensation, perception, and cognition [1.79]. His reasoning combined fundamental categories of Kantian epistemology with realism and empiricism rooted in experiment and observation. Finally, like many scientists (from Huygens to Planck, Einstein, Fokker or Van der Pol) he had a musical background that let him study musical problems from his own experience and judgment. For experimental observations of pitch and interval perception and exploration of musical scales, he had a 53-keyed instrument (see [1.80] and the picture in [1.81, p. 76]) at his disposal. Taken together, all these factors made the *Tonempfindungen* a paradigm of music-related science as the book offered a coherent and systematic treatment progressing from fundamentals of vibration and properties of sound to auditory sensation and, in the final sections, to an assessment of concepts established in music theory (like tonality, harmony, voice leading) in

the light of empirical facts and explanations. Of course, later research has shown that some of Helmholtz's hypotheses were not confirmed by fact; this holds true in particular for the basilar membrane (BM) conceived as a chain of resonators tuned to certain *best frequencies*. (Later concepts still view the BM in regard to a tonotopic organization; the concept of *best frequency* also plays an important role in auditory theory; see Chap. 31.)

Carl Stumpf was a philosopher and psychologist with a thorough musical background (he learned to play a number of instruments and used to sing; he even pondered a musical career in his youth [1.82]). He studied philosophy with Brentano and then with Lotze (for his Ph.D. degree), the former known for his contributions to logic and studies on perception and cognition, the latter (with formal training also as a physician) likewise for books on logic and epistemology yet with a deep interest in psychology and aesthetics (for historical background, see *Boring* [1.83], *Pongratz* [1.84]). Combining theoretical analysis with empirical observation, Stumpf elaborated on the sensation and perception of tones (single tones as well as series of tones and intervals) in his *Tonpsychologie*, originally conceived in four volumes of which two appeared, in 1883 and 1890, under this title, and most of what was planned for vol. 3, separately as *Konsonanz und Dissonanz* [1.85]. It was with this latter monograph that Stumpf opened a series, *Beiträge zur Akustik und Musikwissenschaft* (H. 1–9, 1898–1924), which offered many articles on sensation and perception of musical sound as well as on topics from comparative musicology.

Stumpf acknowledged some of Helmholtz's achievements in psychoacoustics but was critical of his concept of consonance and dissonance as it lacked, as Stumpf saw it, a psychological explanation. Taking Helmholtz's criteria (beats, roughness, coincidence of partials) as objective features inherent in the sound structure, *Stumpf* [1.86, 87] argued these are in fact antecedent conditions relevant for acoustics and sensory physiology, whereas a psychological explanation must look for the effects sound structures have in the perception of listeners. Perception, for Stumpf as well as for his teacher, *Franz Brentano* [1.88], includes cognitive acts and functions such as noticing certain objects and features, identification and comparison of such features, judgment, and apprehension of chords and melodies in music as configurations of elements. Stumpf therefore offered a different concept he believed was fundamental to the experience of consonance in both complex harmonic tones as well as in multiplicities of sound. This concept he addressed as *Verschmelzung* ([1.89] and Chap. 31 of this handbook).

With his background ranging from sensory physiology and acoustics to foundations of psychology as related to theory of knowledge [1.82, 90], Stumpf pursued research that integrated systematic and comparative musicology [1.50, pp. 29ff.] since he was of the opinion that perceptual and musical phenomena, such as formation of tone systems and scale types, consonance and dissonance, cannot be studied appropriately if confined to observations from European art music (similar concerns had been expressed also by *Helmholtz* [1.76]). Stumpf's Psychological Institute (founded in 1894 as a seminar, expanded in 1900) soon included a collection of non-Western music recorded in the field or in Berlin from ensembles visiting the capital. One such recording session took place in September 1900 with a *Siamese* (Thai) ensemble [1.91]. From the recordings as well as from additional information *Stumpf* had gathered in interviews and small experiments he took with the musicians, he prepared an in-depth study of the tone system and music of this culture [1.92]. This study can be called exemplary in that it offers a description of the instruments played by the Thai ensemble along with data from measurements of their tuning, followed by a discussion of how the seemingly equiheptatonic scale of the Thai and a similarly equipentatonic scale assumed for the Javanese *gamelan slendro* might be explained in regard to perception. To be sure, sectioning an octave (2 : 1) into two equal parts involves calculating the geometric mean, which is $\sqrt{2} = 1.41421$, the frequency ratio of the tritone in modern ET12. Sectioning an octave into five, seven, or twelve equal parts (as in ET12) is even more demanding (one has to calculate the basic step as $\sqrt[5]{2}$, $\sqrt[7]{2}$, or $\sqrt[12]{2}$ respectively, and then take their multiples). *Stumpf* [1.92, p. 89], therefore, asked: How could one establish such scales like the *Siamese* and slendró without calculating roots and logarithms? Stumpf drew on *Fechner's law* (which relates stimulus magnitudes to sensation, see Chap. 30 of this book) and considered also distance estimates for a tentative explanation. As a trained musician, he was able to transcribe and analyze a substantial portion of the music as well [1.92].

A device that facilitated Stumpf's investigation of Thai (and other non-Western) music was Edison's *perfected* phonograph that became available in 1888. *Stumpf* [1.92, p. 136] also used a brand-new *Telephonograph* (also *Telegraphon*, invented by the Danish physicist, Valdemar Poulsen), the first machine to offer magnetic recording on a steel wire, in a much improved sound quality. Recording of actual sound for analysis and transcription was a fundamental requirement for systematic and comparative musicology as well as for phonetics [1.93]. For a period of more than two decades, Stumpf engaged himself (as well as his

coworkers) in analysis and synthesis of sound as well as in experiments on perception. He condensed his many observations into a monograph addressing speech sound but also phenomena relevant for music such as *sound color* and intonation [1.87].

Hugo Riemann, like Stumpf, took his doctorate with Lotze. He submitted a thesis [1.94] on music perception, which he relates to cognitive foundations of functional harmony. In his thesis and some other early writings, Riemann tried to integrate concepts from *Rameau* [1.37, 41], *Hauptmann* (1853) [1.95], *Helmholtz* (1870) [1.96], and *von Oettingen* (1866) [1.48]. At the core of his considerations is harmonic dualism, which brings up the problem of symmetry between *overtones* and *undertones* tackled by Rameau (see above). *Riemann* [1.94, pp. 12ff.], referring to Helmholtz's BM resonance theory, made some rather speculative remarks where he asserts that *undertones* could be evoked in auditory sensations since fibers in the BM corresponding to subharmonics of the frequency of a tone actually presented to the ear could be set to partial sympathetic vibration. From such a pattern of partial BM fiber responses he concluded we would have *an implicit perception of undertones*. Riemann did not, as has been erroneously presumed at times, assert that *undertones* were part of the sound wave reaching our ear, and hence an acoustic phenomenon (see [1.94, pp. 12f.], [1.51]). His argument by and large follows *Rameau* [1.41] yet is beefed up with a hypothesis of sympathetic resonance in BM fibers in the light of Corti's discovery and Helmholtz's resonance theory (strange enough, *Rameau* [1.37, p. 7] made reference to *Fibres qui tapissent de fond de la Conque de l'Oreille*, which he believed would resonate like *corps sonores*).

Riemann, like Rameau before him, needed *undertones* (even if only perceived indirectly) for a symmetric model of major and minor. He had one remark on harmonic minor, evaluated as *blurred major* even by Helmholtz, saying that, in a minor chord, the harmonic partials contained in each complex tone (e.g., of a piano) would interfere with the intervals between the tones of the minor chord while they would fit well to a major triad. This is indeed a point that can be proven empirically, for if a minor chord on a grand piano tuned to ET12 is played (in *root* position with the bass note doubled in the lower octave) by simply moving one finger one key down from the major to the minor third, the degree of harmonicity (measured as harmonic-to-noise (HNR) ratio from cross-correlation and expressed in dB) drops significantly (Fig. 1.2). This is why minor chords in *root* position with tones condensed into one octave in fact are *less perfect* (as Rameau had it) than major chords. Composers aware of this fact often have

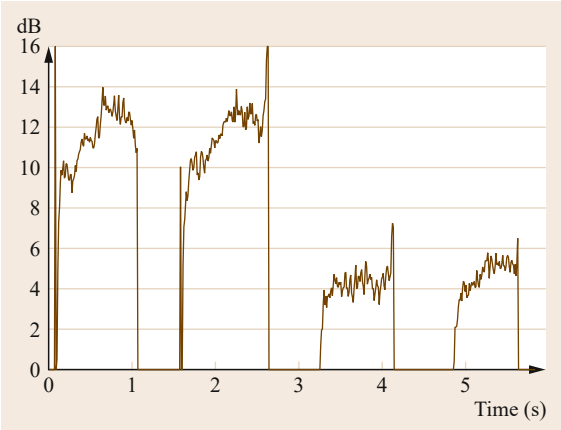


Fig. 1.2 C-major (*left*) and c-minor (*right*) played in *root* position, grand piano (ET12), HNR (dB) over time

set minor chords so that their tones are distributed over more than one octave.

Riemann apparently had hoped for some cooperation with scholars experienced in acoustics and sensation (like Helmholtz, Oettingen, Stumpf) that unfortunately never materialized [1.50]. At the same time, his aim was to continue the *logical* approach to tonality outlined by *Hauptmann* [1.54] who viewed tonal and chord functions in terms close to Hegelian dialectics. Hauptmann argued that the chord of the tonic is the basic unit (equivalent to a thesis) and that harmonic tension is generated from the fifth above and below a tonic that *disunites* with the tonic. However, the dominant and subdominant lead back to the tonic, and the three chords thereby are considered as an *organic unit*. The scheme *Hauptmann* [1.54, p. 26] gives for this unit (the elementary cadence) is symmetric (he also was a dualist)

F a C e G C e G h D

Since Hauptmann conceived his tonal relations in just intonation pitches (like Helmholtz, von Oettingen, and also Riemann in his early years), one may write the scheme as follows

−1 a e h
0 f c g d
(−1 = 1 syntonic comma flat relative to 0)

It is plain to see that there are four pure fifths at the bottom, three of which have a pure major third above. The tones a, e, h again are in relations of just fifths among each other, however, the interval d–a is not a pure fifth but is narrowed by one comma to 680 cents. One may use the seven notes as a just major scale [1.44] for

melodic intervals. Like Rameau before him, *Hauptmann* struggled with harmonic minor [1.54, 32ff.] and modeled the minor triad as the inverse (6 : 5 : 4) of major. Applying his logical arguments, he considered minor a *negation* of major.

Riemann took up Hauptmann's idea of a *musical logic* in a range of writings, of which his *Musikalische Syntaxis* (1877) sketched the conceptual framework while his books on harmony [1.97], on rhythm and metrics [1.98] as well as on other musical subjects expanded and detailed his ideas. The broad spectrum of his publications (from instruments and musical acoustics to performance practice, composition, analysis, and aesthetics) established him as perhaps the most versatile music scholar of his time. He played a key role in the development of musicology as an academic discipline around 1880–1910 in that he alone covered almost every field of historic and systematic musicology in a large number of publications, and moreover was acknowledged as a musician and teacher.

1.7 Gestalt Quality and Gestalt Psychology

In much of 19th century psychology [1.83], association served as a fundamental principle. It basically suggests an *additive* mode of experience, in that it assumes subjects combine elements such as elementary sensations into a complex perceptual object, in a sequence of associations. In a quite different approach, one of the problems tackled by *Brentano* [1.88] was if, and perhaps how, several complex objects may be perceived in one act of cognition. The dispute whether sequential association or integrative acts might offer appropriate models for perception was still under way when, in 1890, Ehrenfels published his famous article in which he argued that a melody is not the sum of tones sensed in a row yet probably a new entity, perceived as such. *Mach* had already pointed to the fact that subjects identify two melodies as representing the same *Tongestalt* if they share the same consecutive intervals [1.99, Chap. XIII]. As a parallel, he mentions geometric figures that we can recognize as similar in shape if their parts have identical relations. Mach ascribed the experience of convergence to sensations as well as to memory. *Ehrenfels* took another approach since, in a melody transposed by a certain interval up or down, all tones in the transposed version differ from the original in their pitch, which implies tone sensations must be different while transposition does not affect the profile of intervals making up this melody [1.100]. He suggested hearing a melody would produce a chain of tonal images or *impressions* (*Eindrücke*), which are integrated into one complex conception in the listener's consciousness (the German term he uses is *Bewusstsein*). The elements integrated into the complex can still be identified as such in perceptual analysis, however, the specific quality of the complex (which makes a melody or a geometrical shape identifiable against others) cannot be derived from the row of elements (e.g., tones) in a temporal Gestalt, or from the sum of elements (such as points, lines, angles) in a spatial Gestalt. While the elements (tones, lines, etc.) are real in terms of sensory data, the complex conception (in modern terminology: the perceptual object) is constituted in cognitive acts. Gestalt qualities according to Ehrenfels' definition are *positive conceptual contents* (*positive Vorstellungsinhalte*) that require perception of ordered complexes of elements. These complexes are fundaments (*Grundlagen*) for Gestalt qualities. Hence, the scheme Ehrenfels suggests contains three stages: first, sensation of elements; second, integration of elements into one complex conception; and third, apperception of a Gestalt quality. Though the process is sequential in regard to the first two steps, Ehrenfels argued the Gestalt quality would be *given* as a psychic fact immediately and concurrently with its *Grundlage* (the conceptual complex derived from temporal and structural integration of elements).

A few decades later, when Gestalt psychology became a paradigm [1.83, Chap. 23], this aspect of immediacy at times has been emphasized to the point that Gestalten should be *given* as wholes, and that these *complex wholes* (as Kurt Koffka called them) would be perceived as such, not as an additive sum of sensations and not even as a perceptual structure integrated over time and/or space. In a radical way, this view claims primacy of a Gestalt over its constituents, in regard to their occurrence in perception as well as their hierarchy. A more moderate view was that temporal structures such as melodies or rhythmic patterns give rise to perception of a Gestalt quality in that listeners recognize a well-formed, stable configuration of elements (and may mentally *complete* a temporal configuration after perceiving just the first few elements ordered into a structure). A Gestalt thereby is an emergent quality assigned to a perceptual object. To be sure, *Stumpf* [1.90, § 15] was quite critical of some tenets of the Gestalt school (spearheaded by his former students Wertheimer, Koffka, and Köhler), and in particular of Gestalten as immediately given, complex wholes.

He maintained that, for most Gestalt-like configurations (such as chords in music) perception includes an evaluation of relations between structural elements (otherwise, at least a musically trained listener may hear a complex, but may not perceive a Gestalt). Stumpf's approach was that of a cognitivist who would not expel an analytical stance for perceiving Gestalten. He [1.87, 90] admitted, though, that there were complexes in musical sound that, in certain respects, were Gestalt-like in quality yet indeed difficult to analyze perceptually into their constituents. Musical timbres (or sound colors) were among such complexes. *Robert Lach* [1.101], in a scholarly paper on indeterminacy and ambiguity of Gestalten in music, argued there were four basic parameters of tone sensation (pitch, intensity, sound color, duration) and three Gestalt qualities in tonal music (sound color, harmony, melody). Though sound color could be explained in physical terms (according to the Ohm–Helmholtz concept of additive Fourier synthesis) it would not be perceived (different from melody and harmony) as elementary, coherent and unambiguous. Notwithstanding such ambivalence, sound color would be perceived as a complex that in itself has a primary Gestalt quality in terms of aesthetic appreciation.

The Gestalt concept, which had been developed first in regard to temporal and spatial structures (melody, geometric figures) derived from perception of ordered complexes, was soon extended to various *wholes* (including, for example, emotions). Also, from Ehrenfels' concept of Gestalt quality an expansion was made (by Meinong) toward *higher order* Gestalten. The configuration that is perceived as a Gestalt by itself of course can be used as an element in a complex of Gestalten (such as a melodic theme in a polyphonic setting). *Hornbostel* [1.102], from his background in phenomenology and Gestalt psychology, surveyed auditory phenomena where he, among other topics, discussed consonance in terms of Gestalt criteria such as unity, simplicity, stability, and closedness. In a contribution to the Festschrift for Guido Adler (known for his book on the concept of *style* in music), *Hornbostel* [1.103] suggested musical works might be apprehended as representing a certain *style* from the perceptions a number of musically trained subjects have in common by *centering* on the same structural features contained in particular works.

Since the experience of a subject performing music or listening to music is perception of a multitude of Gestalt-like formations, the concept lends itself to music analysis as well as to music psychology and aesthetics. For example, *Wellek* [1.104], who had academic training both as musicologist and psychologist, employed Gestalt concepts in his studies and gave criteria (like closedness, stableness, degree of inner structuredness) for the evaluation of Gestalt-like formations in music. In the United States, *Leonard Meyer* [1.105] discussed pattern perception in music in regard to well-known Gestalt *laws* (of which *Helson* [1.106] provided a list that contained more than one hundred items) like good continuation, completion and closure.

From the side of experimental psychology, rhythm was one area investigated with respect to perception already before 1900 (in studies by Bolton and by Meumann). *Koffka* [1.107] took up rhythm in his dissertation (at Berlin under Stumpf) and found subjects tend to structure isochronous pulse sequences into perceptual patterns. One may regard subjective groupings of (sonic or optical) pulses that have the same intensity, quality, and duration in a sequence the result of cognitive activity (that is, grouping is viewed as a *psychic function* in line with Stumpf's concept). However, such grouping processes often take place almost involuntarily if a stimulus (in this case, an isochronous pulse sequence) is presented for some time. A similar observation is that subjects tend to instantly complete ensembles of graphic elements so as to make up a geometric figure. As to rhythm perception, *Schmidt* [1.108] found that subjects structure the elements within a rhythmic pattern from a larger unit containing a number of beats (a period, which can be taken as a temporal Gestalt), that is, the Gestalt as a whole governs the arrangement of the parts. In addition, a range of experiments conducted in France confirmed that rhythm is a temporal Gestalt phenomenon that, however, involves motor behavior [1.109, 110].

Another experiment from Gestalt psychology relevant for music perception was to offer the initial part of a melody to subjects (with some musical background) who were asked to *continue* as well as to *complete* the fragment [1.111]. Such experiments explored the actual range of melodic patterns that subjects regarded as a *good continuation* of a given melodic shape. In discussing temporal organization of tones in melodic patterns, *Koffka* [1.112, p. 434] had the following example (Fig. 1.3) for which he argued that, in the two scales (played staccato, one ascending from C_4 to C_5, the other descending from C_5) the common tone A_4 *will not be heard as one long note, but as two notes of normal length*. Koffka concludes that *the factor of uni-*

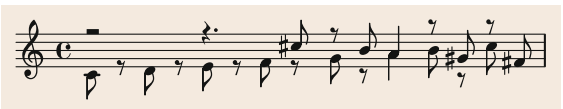


Fig. 1.3 Ascending and descending scales, *uniformity* versus *good continuation*

formity or homogeneity (as realized in the two scales) *has been overcome by the factor of good continuation*.

A quite advanced research project was conducted, by *Truslit* [1.113], on gestural motion in and along with music (where musical Gestalten induce a number of *typical* gestural movements condensed in trajectories that can be shown as curves). One may regard such gestural movements as spatiotemporal Gestalten (see articles in [1.114]).

1.8 Music Psychology: Individual and Sociocultural Factors

Gestalt psychology addressed, for the most part, perception as observed in individuals. However, there were also issues such as musical ability and aptitude as well as actual musical performance that became topics in psychology of music. *Carl Seashore* developed a test for musical talent and published a monograph on the same subject [1.115]. His lab in Iowa became a center of music psychology, a field covered in one of his textbooks [1.116]. Seashore's seminal work on musical *talent* spawned revised versions of his test (from his own lab) as well as some tests (Bentley, Wing, Gordon) more or less similar in design (but expanding from elementary to more *musical* items). He also edited collections of experimental studies on topics in music psychology, for example intonation in singing and violin playing (see contributions by *Harold Seashore*, Ray Miller, Arnold Small, and Paul Greene in [1.117]). As to intonation practice, this had been investigated earlier by *Stumpf* and *Max Meyer* [1.118]; the latter then went to the United States to continue his work at Iowa [1.83]. *Seashore* [1.119] also devoted himself to empirical aesthetics of music. Musical enjoyment had been studied experimentally by *Weld* [1.120] who employed psychological as well as psychophysiological methods. *Hevner* [1.121–123] employed her well-known cycle of adjectives to investigate how musical parameters such as the major and minor mode as well as pitch and tempo can influence the perceived expressiveness and the affective value of music. *Gundlach* [1.124], in a study noteworthy for its statistical techniques including factor analysis, presented 40 short musical phrases to more than 100 subjects who were asked to mark those of 17 preselected *categories* (adjectives such as exalted, mournful, glad, melancholy) or add some of their own choice that would characterize each phrase (and the presumed attitude of its composer) best.

In the studies published by Gundlach and by Hevner, the effect of music on subjects listening to music was investigated. Since their subjects typically were college students, reasonable homogeneity of subjects in their samples can be assumed in regard to mental capacity and other parameters. However, responses of subjects of course depend on their musical training and background as well as on preferences. To study a possible dependence of aesthetic judgments on intelligence, age, and musical training, *Rubin-Rabson* [1.125] presented 24 examples of music covering the period of around 1750–1925 from phonograph recordings to 70 adult men and women who could express their liking or dislike for each item on a rating scale. From the intercorrelation matrices she interpreted three parameters (age, intelligence, training) as positively or negatively related to groups of musical examples regarded as representing three historical and stylistic periods (classic, transition, modern).

One has to remember that in the 1930s the issue of *like* and *dislike* as well as factors relevant for listening preferences of audiences had become of interest to the industry behind the media (radio, phonogram records, publishing) as well as for social scientists. In the years 1938–1941, Th. W. Adorno was involved in the Princeton Radio Project (led by Paul Lazarsfeld), for which he wrote several critical papers that include proposals for empirical research focusing on individual listeners of radio programs. The essence of his criticism in regard to *popular* music as presented in radio and other media is condensed in the chapter on *culture industry* in the book he wrote jointly with *Horkheimer* [1.126]. *Adorno*'s contributions to social psychology and sociology of music [1.127] are based on his observations and critical judgment but do not attempt to provide empirical data or similar factual evidence in support of arguments. However, a more formal approach to the social psychology of music including *musical taste* was soon established by, in particular, *Farnsworth* [1.128] who had measured, among other issues, the interest school children had in either *serious* or in *popular* music. Empirical studies on the reception of music among school children and adolescents [1.129, 130] confirmed the dominant role pop and rock music play in the life of young people. Empirical and theoretical studies were also carried out in music sociology with the aim of studying uses and functions of various types of music in societal groups [1.131]. The *International Review of the Aesthetics and Sociology of Music* (founded in 1971) publishes scholarly papers from both areas including aspects of media. Journals such as *Popular Music* are devoted to various musical, technical, and sociocultural aspects of the field.

1.9 Some Modern Developments

Summing up developments briefly sketched in the preceding paragraphs, one can say that, by about 1960, areas such as musical acoustics, music theory, and music psychology were established and productive areas of research that yielded many articles in specialized journals as well as a number of monographs and textbooks. For example, the *Journal of Music Theory* was issued in 1957, and a comprehensive book on music perception by Robert Francès appeared in 1958 (for an English translation, see *Francès* [1.132]). While experimental research on perception of musical intervals [1.133] and rhythmic structures [1.110] was continued (largely from a Gestalt psychological perspective), several new areas gained interest, among them sound as a medium for creative work in primarily electronic music, and timbre as a *dimension* relevant for both musical production and perception (see [1.134] and Chap. 32 of this handbook). Further, with the rise of communication and information theory, computers were employed for *experimental music* [1.135] beginning in the 1950s, and information theory soon reached sound and music research [1.136–139]. Since about 1970, computers were widely employed for sound generation as well as composition and analyses of music, often in combination with signal processing methodology (for an overview, see articles in [1.140–143]). The *Computer Music Journal* (founded in 1977) and the *Journal of New Music Research* (formerly *Interface*, founded in 1972) serve an international community of scientists and musicians studying sound and music with computers. Signal processing of course became central in musical acoustics and sound research [1.144], and in particular in physical modeling and design of virtual instruments (there is a large body of relevant publications, many of which are referenced in [1.145, 146]). Further, computers along with special codes have been employed in modeling perception and production of music with artificial intelligence (AI) methodology (e.g., contributions in [1.147, 148]) and by means of artificial neural networks (ANNs) [1.149–151]. In addition, parameters of musical performance, such as tempo and dynamic changes adapted to musical textures and relevant for artistic expression, have been modeled with special software [1.152, 153]. More recently, areas of machine learning, automated music transcription as well as music retrieval and data mining have become large fields of research, partly with practical applications (see chapters in *Klapuri* and *Davy* [1.154] and *Ystad* et al. [1.155] and articles in a special issue edited by *Conklin* et al. [1.156]).

Around 1970, concepts from structural linguistics and semiotics were adapted more frequently to music research, which implies some parallels between natural languages and music. Of course, there are certain analogies that have been stressed long ago (as in concepts of *musical rhetoric* in Renaissance and Baroque music theory and practice [1.157]). Quite many books on *practical* music theory provide rules for creating *well-formed* musical textures (see, for example, [1.158]). Well-formedness of sentences in regard to grammar and syntax of course is a criterion for accurate speech as much as well-formed musical settings should follow the rules of harmony and counterpoint. The analogical relation between speech and musical textures gained a more formal status when *Hauptmann* [1.54] and in particular *Riemann* [1.55, 97] conceived of harmony and musical structure in terms of syntax and grammar, implying musical *sentences* or *propositions* have to conform to *musical logic*. However, the parallel between language and music gained new ground when music was indeed viewed, beginning in about 1950, as a *communication system* and as a *sign phenomenon* that would lend itself to an adaptation of categories developed in linguistics and semiotics. Adaptation of course means one has to establish correspondences in a plausible way between structures in language and music in regard to syntactic, semantic, and pragmatic categories, and it calls also for a definition of sign categories. The discussion of obvious parallels as well as of distinctive differences between language and music was intensive in the 1970s and 1980s (of the many publications, see [1.159–164]). Though it is possible to demonstrate, in a plausible way, how (indexical and other) signs are used to *make sense* in certain genres and pieces of music [1.165], it was found rather difficult to establish semiotics for music in general. One of the reasons probably is that music in most of its forms and contexts in essence seems gestural rather than propositional (see [1.161] and articles in [1.114]). Quite a different matter is that musical analysis cannot avoid language to conceptualize music and to communicate its findings (since it could be a problem to sing or whistle such results to colleagues in music and musicology), a circumstance that *Charles Seeger* [1.166] has called the *linguocentric predicament*.

The linguistic paradigm of generative-transformational grammar [1.167] clearly has been the inspiration for the generative theory of tonal music (GTTM) worked out by *Lerdahl* and *Jackendoff* [1.56]. Their endeavor has been lauded for the vigor with which rules have been set up for an analysis of music into segments pertaining to hierarchical structures, but there has been criticism arguing, for example, that listeners would rather perceive music as a sequence of mean-

ingful units arranged to temporal order than mentally *represent* musical structures like hierarchical tree diagrams. However, GTTM in fact attempts to formulate rules that reflect perceptual and cognitive strategies of the *experienced listener* (in a similar vein, *Riemann*'s musical logic, with a mental representation of music as tonal conceptions (*Tonvorstellungen*, see [1.168]), aimed at subjects with a thorough musical background).

Another problem that became apparent in cognitive music research during the 1980s was that the distinction of *symbolic* and *subsymbolic* in regard to music production and perception may be misleading. The notion of *symbol* and concepts of *symbolic* processing of course have played a major role in semiotics and computer science, from where they were transferred into cognitive sciences including areas of *cognitive musicology* (for a detailed account of concepts, see *Seifert* [1.148]). To be sure, there are definitions of cognition in terms of computation such as: *cognition is symbolic computation* or *cognition is massively parallel numerical computation* [1.169, p. 46], which express a metaphor according to which the *mind* and/or the brain functions in ways equivalent to a computer. Though the *mind-as-computer* idea has been questioned as a suitable model for cognition [1.170], and other metaphors (e.g., *cognition as intuitive inferential statistics* [1.171]) have been shown to reflect methodological tools available to scientists at a certain time turned into explanatory models, this does not necessarily impair their heuristic value. The point of interest here rather is that, in cognitive musicology of the 1980s, descriptions and analyses often refer to scores viewed as a *symbolic* representation of music. By comparison, music as organized sound has been approached as *subsymbolic* so as if it were the mere fundament for a *higher* level of processing: that of cognition directed to analyzing musical structures from scores. This in fact is the time-honored approach of the musical expert who, according to *Riemann* [1.55, 168], apprehends music from carefully *reading* scores. While this might in fact be enough to understand syntactic and formal structures, it is obvious that the score of a work falls short of representing timbral and dynamic parameters in sufficient detail, and that one needs *ears* to perceive and apprehend music in full (in addition, musical sound is suited to eliciting emotional qualities, on a psychophysiological basis). Thus, if music reduced to a score can be called a *symbolic* representation, a live or studio performance including a range of timbral and dynamic shadings as well as subtleties of intonation, phrasing, etc. should be taken as *suprasymbolic* rather than *subsymbolic* (for it yields more than just reading symbols). The signification process in music in fact relates to sound, in the first place [1.172].

With the upswing of the *cognitive sciences* (including a range of computational, psychological, neurological, etc. approaches [1.173]) in the decades beginning by around 1960, a fresh interest in music perception including conceptualizations of music, musical memory, and musical imagery has led to an impressive range of empirical plus theoretical studies. Topics of research included basic structures and schemata of tonal organization, pitch perception and scale formation, melodic contours, memory for pitch and melodic shapes, perception of timbre, rhythm and meter, the role of time and *timing* in the production, performance and perception of music, music as related to bodily movement and gestures, musical form, musical imagery, musical creativity, improvisation (also in jazz and in non-Western music), learning of music (with a focus on cognitive development as well as actual musical training), performance practice (in particular, parameters relevant for expression), emotion and *mood* as related to music, etc. In addition to a large number of articles published in journals such as *Psychology of Music* (founded 1977), *Psychomusicology* (founded 1981), *Music Perception* (founded 1983), and *Musicae Scientiae* (founded 1997), there have been many monographs and anthologies covering specific areas of music perception and cognition or providing an overview of relevant research close to a textbook format. The following list is but a small selection (given the space available for this chapter) from a much broader range of relevant publications. In chronological order, the selection includes [1.174–184], [1.150], [1.185–193], and [1.114].

Among the many developments in experimental and cognitive music psychology, a revival of Gestalt psychology principles (in a modern and expanded approach termed auditory scene analysis [1.194]) seems noteworthy. Also of interest is the more recent rediscovery of the body as part not only of music perception but also cognition [1.193]. As a matter of fact, people tend to synchronize bodily functions such as heart and pulse rate or breathing as well as their movements in walking or dancing to periodic stimuli, and there is entrainment on various levels (see [1.195] and a range of peer commentaries, also [1.196]).

This brings us to a final point in this section, a research area sometimes labeled neuromusicology. It was clear to Helmholtz, Stumpf and other scholars that brain structures were the ultimate stage where sound and music perception as well as cognition (*apperception*) takes place. It was not possible, though, to investigate brain functions in a noninvasive technique before the invention of the electroencephalogram (EEG) (around 1930). Anatomical, neurophysiological and functional studies in regard to sound and music were intensified

in the 1960s and 1970s when handedness and hemispheric dominance as well as music-related disorders and clinical applications of music in therapy became major issues for research (some of which was summarized in [1.197]; see also articles in [1.198, 199]). In the course of the 1980s, measurement of auditory evoked potentials (AEP) was a standard technique, and brain imagery by means of positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) had become available also for research related to sound and music perception [1.200]. Since the 1980s, an enormous amount of research has been done in experimental neuroscience and neuropsychology including studies on music perception and cognition (see articles in [1.201–203], and Part C as well as Chap. 31 of this handbook).

1.10 Systematic Musicology as a Musicological Discipline

Though musicology as an area of research and as a body of knowledge has been established in antiquity, to be continued in modern times (see above), it was only in the 19th century that a discipline under this name became part of European universities. In Germany and Austria, the movement known as *Historismus* (historicism) led to incorporating subjects such as prehistoric archeology, art history and musicology into the philosophical faculties. Musicology had a strong historical and philological orientation (with a focus on European music history, biographies of composers, and scholarly edition of works from various epochs and musical genres according to philological principles), which is obvious from the first professorships that were installed at Prague, Vienna, and Berlin (all in the 1870s). However, by that time a somewhat broader view of musicology had already been pursued by the Belgian composer and musicologist, François-Joseph Fétis (who published also on music theory, and who, in addition to studying European art music, had an interest in genres of folk and even non-Western music traditions). Also, as outlined above (Section 1.6), areas such as musical acoustics and psychoacoustics had been firmly established by about 1860–70, so that the philological and historical approach to music history and performance was paralleled by the scientific study of sound and fundamentals of music including perception and appreciation (Fechner was one of the scholars who had propagated empirical foundations not only for psychology but also for aesthetics [1.204]). So, by about 1880 two different orientations existed side by side, and research in a third area, namely folk and vernacular music traditions in Europe as well as the study of so-called *primitive* music in areas of Africa, Asia, the Americas, and Oceania, was also underway.

When *Guido Adler*, in 1885, opened a new journal devoted to musicology with his programmatic article [1.205], he included already comparative musicology into his synopsis of musicological research fields and subdisciplines. Adler's bipartite scheme of *historic* and *systematic* (the latter including comparative musicology) entailed conceptual decisions [1.50] that relate to ideas of music history viewed in terms of development and *evolution*. In short, the historical branch of research should furnish factual evidence for a *progressive* development of music from modest beginnings to full art as evident in European genres, while the task of finding *governing laws* (as contained in fundamental structures of sound and music) for that *evolution* was assigned to the systematic branch. To complete the picture in regard to geographical and ethnic diversity as well as developmental stratification, comparative musicology would have to provide specimens of music from all around the world, and to put them into some order in regard to musical structures and functions as well as tentative musical development ([1.206] and [1.207] are synopses written from this point of view; see also [1.69]). From Adler's scheme, a tripartite organization of musicology into historic, systematic, and comparative resulted, which in fact had practical consequences far into the 20th century, if not to the present (the unifying ideas behind Adler's scheme were abandoned when, in the 20th century, the all-embracing evolutionary perspective was given up for, first, historical indifference and cultural relativism, and then for various trends of *postmodernism* and emphatic *multiculturalism*).

A significant difference between historic and systematic orientations in musicology in fact was and is that the former worked (and continues to work) predominantly with music as *text* (conventional notations or similar *symbolic* representation), while the latter from the very beginning had a focus on sound and perception [1.93, 150, 172]. Further, systematic musicology, like acoustics, psychoacoustics, and psychology, made use of experimental methodology including statistics. Even *Stumpf*, who argued that it is unreasonable to believe that dozens of unmusical laymen could judge musical items more reliably than any single expert musician or musically experienced researcher, included statistics in his studies [1.118]. In the 20th century, experiments and statistics became an integral part

of music psychology [1.208, 209]. Controlled observation, field work and statistics of course are fundamental also to sociocultural studies conducted by researchers in comparative musicology and ethnomusicology as well as in the field of popular music and jazz.

The division of musicology into historic and systematic subdisciplines, consequent to conceptual and methodological differences [1.50, 210], was put into practice early when, at the University of Vienna, a professorship in musicology was established that, in essence, combined areas of music psychology and comparative musicology. Vienna has maintained a combination of *scientific* and *cultural* research efforts in a specific concept of *Vergleichend-Systematische Musikwissenschaft* [1.211, 212] for a long time, and at Berlin, Stumpf's associate Erich M. von Hornbostel also covered systematic and comparative musicology in his teaching and in his publications. With the decline of comparative musicology due to political and other factors [1.69], and the rise of ethnomusicology as a discipline with a descriptive and *culturalist* rather than a systematic orientation (see [1.213]), a tripartite division into historic, systematic, and ethnomusicology was carried out at a number of universities (e.g., Cologne, UCLA). In academic institutions of the US, music theory also gained relative independence as a musicological subdiscipline (that typically includes analysis of works from European art music, and may include also genres such as electronic and computer music).

The idea of systematic musicology as *fundamental science of music* stems from a combination of several factors. One (that relates to basic concepts of numbers, *counting* and *measurement* as well as directed observation) certainly is of antique heritage; another relates to developments in music theory from around 1400–1750 (as outlined above) that gradually introduced empirical observation and, in particular with Rameau, musical acoustics and perception into music theory. Musical acoustics and psychoacoustics expanded greatly in the 19th century, and with their groundbreaking results demonstrated theoretic and methodological independence of *systematic* areas of research from historic-philological orientations. In a wider sense, musicology thus was one particular field marking the general split of disciplines into sciences and *humanities* (that took place, roughly, between 1800 and 1860). In regard to music research, this split was perhaps unavoidable, though not total since of course music theory and also studies in music psychology often combine aspects from both areas (the *integrative* approach is still evident in Helmholtz's seminal book, where its author offers chapters on music theoretical topics in the main body and – not to embarrass a more general readership – puts the differential equations into appendices to his work).

Notwithstanding trends of professional specialization that led to establishing new disciplines (such as psychology in about 1870–80 and sociology also in the second half of the 19th century), systematic musicology as an area of interdisciplinary research benefited from the many contributions offered by scholars who were located in departments other than music or musicology. Perhaps the closest connections were with acoustics and psychology (but also with experimental phonetics, linguistics, anthropology as well as with medical subjects). A common denominator (which prevails to this day) was that many scientists had training and even thorough experience also as musicians, which often led them into research projects on various aspects of sound and music (e.g., [1.214]).

The division of musicology into subdisciplines and fields of research (as envisaged already by Adler, see above) gained new impetus in the years after WWII, for methodological and practical reasons. First, the differences in perspective (e.g., music as *text* versus music as *sound*, historical versus synchronous [1.215, 216]) as well as in research topics and methodology were too obvious to be overlooked. Second, in order to integrate somehow theoretical and factual knowledge accumulated in the large number of studies on musical sound, music perception and performance as well as on sociocultural contexts that had been published during the past decades, one needed scholars capable of dealing with this matter (and qualified to continue research on a professional level). Therefore, establishment of *systematic musicology* as *fundamental and integrative music science* was proposed [1.217]. While the scientific status of systematic musicology as *fundamental science* was acknowledged by many if not most of the scholars working in musicology, there were opinions that stressed the interdisciplinary relation of historic, systematic and comparative research [1.218], which implied a common disciplinary and institutional umbrella. Third, even if one wanted to maintain a single academic discipline of *musicology*, it was clear that professorships for systematic and comparative musicology/ethnomusicology had to be set up at least in a number of universities, as in fact was done in Germany and in the US (plus the chair that had been established at Vienna decades earlier).

With ongoing specialization in areas such as musical acoustics and music psychology and new areas of research coming up with the advance of electronics and studio technology (and when, not much later, also computers became available to musicologists), historic and systematic musicology had been drifting far apart (with ethnomusicology becoming a largely independent subject that, in many institutions in particular in the US, was closer to anthropology and ethnology than

to the somewhat old-fashioned historic musicology). In view of factual and methodological developments reached in the 1970s, it was only natural to give former musicological *branches* independent disciplinary and organizational status.

At Hamburg, comparative musicology had been installed by about 1930, and systematic musicology was part of regular teaching and research since 1948. In 1975, the University of Hamburg decided to make systematic musicology a *sovereign* discipline and subject that can be studied, with a special curriculum (the introductory textbook in 1976 was from *Vladimir Karbusicky* [1.219]) for a *major*. In addition, students can choose *minor* subjects ranging from the sciences to law and economics as well as to humanities. There are some more universities now that offer systematic musicology as a *major* to be studied in a combination with other subjects, and there are also quite many institutions that offer ethnomusicology in some form. It should be noted that major parts of systematic musicology in some institutions are incorporated into programs with more *modern* designations (referring to *computer*, *sound* or *cognition* in their respective titles). As one may expect, with the development of scientific investigations in areas including sound and music, new designations for special segments have been proposed, for example *psychomusicology* (see journal of same title) or *cognitive musicology* [1.220], in order to delimit a certain approach and field of research (see articles in [1.184] and [1.150]). Others have simply opted for a broad label called *empirical musicology*, which may cover many things (given the range of different meanings *empirical* had in psychology alone since the 19th century [1.83]). Systematic musicology in fact almost always included *empirical* observation and experimental research, however, it also includes theoretical reasoning on natural and cultural foundations of music [1.221, 222] and certainly mathematics and modeling [1.223–225] as part of an interdisciplinary tradition that spans from Archytas to the present.

References

1.1 N. Rescher: *Nature and Understanding. The Metaphysics and Method of Science* (Clarendon, Oxford 2000)

1.2 F. Zaminer (Ed.): *Geschichte der Musiktheorie*, Vol. 1–10. (Wissenschaftliche Buchgesellschaft, Darmstadt 1985–2003)

1.3 T. Christensen (Ed.): *The Cambridge History of Western Music Theory* (Cambridge Univ. Press, Cambridge 2002)

1.4 O. Elschek: *Die Musikforschung der Gegenwart, ihre Systematik, Theorie und Entwicklung*, Vol. 1/2 (Stiglmayr, Wien–Föhrenau 1992)

1.5 A. Szabó: *The Beginnings of Greek Mathematics* (Reidel, Dordrecht, Boston 1978)

1.6 A. Barker: *Harmonic and Acoustic Theory*, Greek Musical Writings, Vol. 2 (Cambridge Univ. Press, Cambridge 1989)

1.7 A. Barker: *The Science of Harmonics in Classical Greece* (Cambridge Univ. Press, Cambridge 2007)

1.8 C. Huffman: *Archytas of Tarentum. Pythagorean, Philosopher and Mathematician King* (Cambridge Univ. Press, Cambridge 2005)

1.9 O. Busch: *Logos Syntheseōs. Die Euklidische Sectio Canonis, Aristoxenos, und die Rolle der Mathematik in der antiken Musiktheorie* (Staatliches Institut für Musikforschung, Berlin 1998)

1.10 D. Creese: *The Monochord in Ancient Greek Harmonic Science* (Cambridge Univ. Press, Cambridge 2010)

1.11 B. Münxelhaus: *Pythagoras musicus. Zur Rezeption der pythagoreischen Musiktheorie als quadrivialer Wissenschaft im lateinischen Mittelalter* (Verlag für syst. Musikwiss., Bonn 1976)

1.12 Platon: Res publica. In: *Platonis Opera*, Vol. 4, ed. by J. Burnet (Clarendon, Oxford 1912)

1.13 A. Heilmann: *Boethius' Musiktheorie und das Quadrivium. Eine Einführung in den neuplatonischen Hintergrund von De institutione musica* (Vandenhoeck Ruprecht, Göttingen 2007)

1.14 J. Leodiensis: *Speculum musicae Lib. I–VII*, ed. by R. Bragard (American Institute of Musicology, Rome 1955–1973)

1.15 K.J. Sachs: *Mensura fistularum. Die Mensurierung der Orgelpfeifen im Mittelalter*, Vol. 2 (Musikwissenschaftliche Verlags-Gesellschaft, Murrhardt 1980)

1.16 H. Klotz: *Die Orgelkunst der Gotik, der Renaissance und des Barock*, 2nd edn. (Bärenreiter, Kassel 1975)

1.17 M. Haas: Die Musiklehre im 13. Jahrhundert von Johannes de Garlandia bis Franco. In: *Die Mittelalterliche Lehre von der Mehrstimmigkeit*, Geschichte der Musiktheorie, Vol. 5 (Wissenschaftliche Buchgesellschaft, Darmstadt 1984) pp. 89–159

1.18 F. Rempp: Elementar- und Satzlehre von Tinctoris bis Zarlino. In: *Italienische Musiktheorie im 16. und 17. Jahrhundert*, Geschichte der Musiktheorie, Vol. 7 (Wissenschaftliche Buchgesellschaft, Darmstadt 1989) pp. 39–220

1.19 J. Barbour: *Tuning and Temperament* (Michigan State Univ. Press, East Lansing 1951), repr. (Da Capo, New York 1972)

1.20 M. Lindley: Stimmung und Temperatur. In: *Hören, Messen und Rechnen in der frühen Neuzeit*, Geschichte der Musiktheorie, Vol. 6 (Wissenschaftliche Buchgesellschaft, Darmstadt 1987) pp. 109–331

1.21 P. Barbieri: *Enharmonic Instruments and Music 1470–1900* (Levante, Latina 2008)

1.22 A. Schlick: *Spiegel der Orgelmacher und Organisten* (Peter Schoeffer, Mainz 1511), repr. (Bärenreiter, Kassel 1951), also (Knuf, Buren 1980)

1.23 N. Vicentino: *L'antica musica ridotta alla moderna prattica* (A. Barre, Rome 1555)

1.24 G. Zarlino: *Le Istitutioni Harmoniche* (G. Zarlino, Venice 1558) repr. (Broude, New York 1965)

1.25 E. Küllmer: *Mitschwingende Saiten. Musikinstrumente mit Resonanzsaiten* (Verlag für systematische Musikwissenschaft, Bonn 1986)

1.26 M. Praetorius: *Syntagma musicum*, T. II, *De organographia*. T. III: *Termini musici*. (Holwein, Wolffenbüttel 1619)

1.27 H.F. Cohen: *Quantifying Music. The Science of Music at the First Stage of the Scientific Revolution, 1580–1650* (Reidel, Dordrecht, Boston 1984)

1.28 S. Dostrovsky, J. Cannon: Entstehung der musikalischen Akustik (1660–1750). In: *Hören, Messen und Rechnen in der frühen Neuzeit*, Geschichte der Musiktheorie, Vol. 6 (Wissenschaftliche Buchgesellschaft, Darmstadt 1987) pp. 7–79

1.29 M. Mersenne: *Harmonie universelle, contenant la théorie et la pratique de musique* (Charlemagne, Ballard, Paris 1636–1637) facs. Éd. T. I–III (CNRS, Paris 1965)

1.30 P. Barbieri: Juan Caramuel Lobkowitz (1606–1682): Über die musikalischen Logarithmen und das Problem der musikalischen Temperatur, Musiktheorie **2**, 147–168 (1987)

1.31 C. de Waard (Ed.): *Journal tenu par Isaac Beeckman*, Vol. III (Nijhoff, La Haye 1945)

1.32 F.J. Ratte: *Die Temperatur der Clavierinstrumente* (Bärenreiter, Kassel 1991)

1.33 H. Riemann: *Geschichte der Musiktheorie im IX.–XIX. Jahrhundert*, 2nd edn. (Hesse, Berlin 1920)

1.34 H. Lippius: *Synopsis musicae novae omnino verae atque methodicae universae* (Ledertz, Straßburg 1612) repr. (Olms, Hildesheim 2004)

1.35 J. Sauveur: Système général des intervalles des sons… Mémoires de Mathematique & de Physique, 299–366 (1701). In: *Collected Writings on Musical Acoustics*, ed. by R. Rasch (Diapason, Utrecht 1984) pp. 472–488

1.36 J. Sauveur: Sur l'application des sons harmoniques à la composition des jeux d'orgues, Mémoires de L'Academie…, 316–336 (1702). In: *Collected Writings on Musical Acoustics*, ed. by R. Rasch (Diapason, Utrecht 1984) pp. 168–192

1.37 J.P. Rameau: *Génération harmonique ou traité de musique* (Prault fils, Paris 1737)

1.38 H. Pischner: *Die Harmonielehre Jean-Philippe Rameaus* (Breitkopf Haertel, Leipzig 1963)

1.39 T. Christensen: *Rameau and Musical Thought in the Enlightenment* (Cambridge Univ. Press, Cambridge 1993)

1.40 J.P. Rameau: *Traité de l'harmonie*. (Ballard, Paris 1722)

1.41 J.P. Rameau: *Démonstration du principe de l'harmonie* (Durand Pissot, Paris 1750)

1.42 J.P. Rameau: Extrait d'une reponse de M. Rameau à M. Euler sur l'identité des octaves. (Durand, Paris 1753). In: *J.P. Rameau: Complete Theoretical Writings*, Vol. V, ed. by E. Jacobi (American Institut of Musicology, Dallas 1969) pp. 168–188

1.43 J.P. Rameau: *Nouveau Système de musíque théorique*. (Ballard, Paris 1726)

1.44 M. Vogel: *On the Relations of Tone* (Verlag für systematische Musikwissenschaft, Bonn 1993)

1.45 D. Harrison: *Harmonic Function in Chromatic Music. A Renewed Dualist Theory and an Account of its Precedents* (Univ. of Illinois Press, Chicago, Urbana 1994), paperback edn. 2010

1.46 J.R. d'Alembert: *Élémens de musique théorique et pratique, suivant les principes de M. Rameau* (Durand, Paris 1752), German transl. by Fr. Marpurg (Breitkopf Haertel, Leipzig 1757)

1.47 J.P. Rameau: *Nouvelles Réflexions sur le principe sonore*. (Durand, Paris 1760)

1.48 A. von Oettingen: *Harmoniesystem in dualer Entwicklung* (Gläser, Dorpat, Leipzig 1866)

1.49 A. von Oettingen: *Das duale Harmoniesystem* (C. Siegel, Leipzig 1913)

1.50 A. Schneider: Foundations of systematic musicology: A study in history and theory. In: *Systematic and Comparative Musicology: Concepts, Methods, Findings*, ed. by A. Schneider (P. Lang, Frankfurt/M., Berne 2008) pp. 11–61

1.51 A. Schneider: Music theory: Speculation, reasoning, experience. In: *Musiktheorie / Musikwissenschaft. Geschichte – Methoden – Perspektiven*, ed. by T. Janz, J.P. Sprick (Olms, Hildesheim, New York 2010) pp. 53–97

1.52 P. Hindemith: *Unterweisung im Tonsatz: Theoretischer Teil* (Schott, Mainz 1940)

1.53 J. Handschin: *Der Toncharakter. Eine Einführung in die Tonpsychologie* (Atlantis, Zürich 1948)

1.54 M. Hauptmann: *Die Natur der Harmonik und Metrik* (Breitkopf Haertel, Leipzig 1873), 1st ed. 1853

1.55 H. Riemann: *Musikalische Syntaxis. Grundriß einer harmonischen Satzbildungslehre* (Breitkopf Haertel, Leipzig 1877)

1.56 F. Lerdahl, R. Jackendoff: *A Generative Theory of Tonal Music* (MIT Press, Cambridge 1983)

1.57 E. Terhardt: The concept of musical consonance: A link between music and psychoacoustics, Music Percept. **1**, 276–295 (1984)

1.58 E. Terhardt: Methodische Grundlagen der Musiktheorie, Musicol. Austr. **6**, 107–126 (1986)

1.59 R. Parncutt: *Harmony. A Psychoacoustic Approach* (Springer, Berlin 1989)

1.60 A. Falcon: *Aristotle and the Science of Nature. Unity Without Uniformity* (Cambridge Univ. Press, Cambridge 2005)

1.61 N. Chomsky: *Rules and Representations* (Columbia Univ. Press, New York 1980)

1.62 N. Rescher: *Cognitive Systematization. A Systemstheory Approach to a Coherentist Theory of Knowledge* (Blackwell, Oxford 1979)

1.63 N. Rescher: *Kant and the Reach of Reason. Studies in Kant's Theory of Rational Systematization* (Cambridge Univ. Press, Cambridge 2000)

1.64 R. Haller: *Neopositivismus. Eine historische Einführung in die Philosophie des Wiener Kreises* (Wissenschaftliche Buchgesellschaft, Darmstadt 1993)

1.65 L. von Bertalanffy: *General System Theory* (Braziller, New York 1968)

1.66 I. Kant: *Kritik der Urteilskraft* (Lagarde, Berlin 1793)

1.67 C. Darwin: *Origin of Species by Means of Natural Selection* (Murray, London 1859)

1.68 M. Harris: *The Rise of Anthropological Theory* (Routledge Kegan Paul, London 1969)

1.69 A. Schneider: Comparative and systematic musicology in relation to ethnomusicology. A historical and methodological survey, Ethnomusicology **50**, 236–258 (2006)

1.70 E.F. Chladni: *Die Akustik* (Breitkopf Haertel, Leipzig 1802)

1.71 D. Ullmann: *Chladni und die Entwicklung der Akustik von 1750–1860* (Birkhäuser, Basel 1996)

1.72 A. Corti: Recherches sur l'organe de l'ouïe des mammifères, Z. Wiss. Zool. **III**(1), 1–63 (1851), plus several pages with graphics

1.73 M. Drobisch: *Erste Grundlehren der Mathematischen Psychologie* (Voss, Leipzig 1850)

1.74 T.G. Fechner: *Elemente der Psychophysik*, Vol. 1/2 (Breitkopf Haertel, Leipzig 1863)

1.75 H. von Helmholtz: Über die physiologischen Ursachen der musikalischen Harmonien (Lecture at Univ. of Bonn 1857). In: *Vorträge und Reden*, Vol. 1, 4th edn. (Vieweg, Braunschweig 1896) pp. 119–155

1.76 H. von Helmholtz: *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (Vieweg, Braunschweig 1863), 3rd edn. 1870, 4th edn. 1877, 5th edn. 1896

1.77 S. Vogel: Sensation of tone, perception of sound, and empiricism. In: *Hermann von Helmholtz and the Foundations of Nineteenth-Century Science*, ed. by D. Cahan (Univ. of California Press, Berkeley 1993) pp. 259–287

1.78 E. Hiebert: *The Helmholtz Legacy in Physiological Acoustics* (Springer, Cham 2014)

1.79 H. von Helmholtz: Die Thatsachen in der Wahrnehmung (revised version of a speech delivered in 1878 at the University of Berlin). In: *Vorträge und Reden*, Vol. 2 (Vieweg, Braunschweig 1906) pp. 213–247

1.80 R. Bosanquet: *An Elementary Treatise on Musical Intervals and Temperament* (Macmillan, London 1876)

1.81 J. Fauvel, R. Flood, R. Wilson (Eds.): *Music and Mathematics. From Pythagoras to Fractals* (Oxford Univ. Press, Oxford 2003)

1.82 C. Stumpf: Carl Stumpf. In: *Die Philosophie der Gegenwart in Selbstdarstellungen*, Vol. 5, ed. by R. Schmidt (Meiner, Leipzig 1924) pp. 205–265

1.83 E. Boring: *A History of Experimental Psychology*, 2nd edn. (Appleton-Century-Crofts, New York 1950)

1.84 L. Pongratz: *Problemgeschichte der Psychologie* (Francke, Bern 1967)

1.85 C. Stumpf: Konsonanz und Dissonanz. In: *Beiträge zur Akustik und Musikwissenschaft*, Vol. 1 (Barth, Leipzig 1898)

1.86 C. Stumpf: *Tonpsychologie*, Vol. 1/2 (Barth, Leipzig 1883)

1.87 C. Stumpf: *Die Sprachlaute* (Springer, Berlin 1926)

1.88 F. Brentano: *Psychologie vom empirischen Standpunkte*, Vol. 1 (Duncker Humblot, Leipzig 1874), repr. (Meiner, Hamburg 1955, 1971)

1.89 A. Schneider: "Verschmelzung", tonal fusion, and consonance: Carl Stumpf revisited. In: *Music, Gestalt, and Computing. Studies in Cognitive and Systematic Musicology*, ed. by M. Leman (Springer, Berlin 1997) pp. 117–143

1.90 C. Stumpf: *Erkenntnislehre*, Vol. 1 (Barth, Leipzig 1939)

1.91 A. Simon, U. Wegner (Eds.): *Music! 1000 Recordings. 100 Years of the Berlin Phonogramm-Archiv 1900–2000* (4 CDs and booklet) (Wergo, Mainz 2000) (Schott Music & Media)

1.92 C. Stumpf: Tonsystem und Musik der Siamesen, Beitr. Akustik Musikwiss. **3**, 69–138 (1901)

1.93 A. Schneider: Change and continuity in sound analysis: A review of concepts in regard to musical acoustics, music perception, and transcription. In: *Sound – Perception – Performance*, ed. by R. Bader (Springer, Cham 2013) pp. 71–111

1.94 H. Riemann: *Musikalische Logik. Hauptzüge der physiologischen und psychologischen Begründung unseres Musiksystems* (Kahnt, Leipzig 1873), As Ph.D. diss. at Univ. of Göttingen: *Über das musikalische Hören*

1.95 M. Hauptmann: *Die Natur der Harmonik und Metrik*, 1st edn. (Breitkopf Haertel, Leipzig 1873)

1.96 H. von Helmholtz: *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*, 3rd edn. (Vieweg, Braunschweig 1870)

1.97 H. Riemann: *Vereinfachte Harmonielehre oder die Lehre von den tonalen Funktionen der Akkorde*, 2nd edn. (Augener, London 1893)

1.98 H. Riemann: *System der musikalischen Rhythmik und Metrik* (Breitkopf Haertel, Leipzig 1903)

1.99 E. Mach: *Die Analyse der Empfindungen und das Verhältnis des Physischen zum Psychischen* (G. Fischer, Jena 1886), 9th edn. 1922

1.100 C. von Ehrenfels: Über Gestaltqualitäten, Vierteljahrsschr. Wiss. Philos. **14**, 249–292 (1890)

1.101 R. Lach: Gestaltmehrdeutigkeit und Gestaltunbestimmtheit, Sitzungsber. Akad. Wiss. Wien, Phil.-hist. Kl. **191**, 95–149 (1920), 3. Abhandl.

1.102 E. Hornbostel: Psychologie der Gehörserscheinungen. In: *Handbuch der normalen und pathologischen Physiologie*, Vol. XI,1, ed. by A. Bethe (Springer, Berlin 1926) pp. 701–730

1.103 E. Hornbostel: Gestaltpsychologisches zur Stilkritik. In: *Studien zur Musikgeschichte. Festschrift für Guido Adler zum 75. Geburtstag* (Universal, Wien 1930) pp. 12–16

1.104 A. Wellek: *Musikpsychologie und Musikästhetik* (Akademische Verlagsgesellschaft, Frankfurt/M. 1963)

1.105 L. Meyer: *Emotion and Meaning in Music* (Univ. of Chicago Press, Chicago 1956)

1.106 H. Helson: The fundamental propositions of Gestalt psychology, Psychol. Rev. **40**, 13–32 (1933)

1.107 K. Koffka: Experimental-Untersuchungen zur Lehre vom Rhythmus, Z. Psychol. **52**, 1–109 (1909)

1.108 E. Schmidt: Über den Aufbau rhythmischer Gestalten, Neue psychol. Studien **XIV**(2), 1–98 (1939)

1.109 P. Fraisse: *Les Structures Rythmiques* (Publ. Univ. de Louvain, Louvain 1956)

1.110 P. Fraisse: *Psychologie du Rythme* (Pr. Univ. de France, Vendôme 1974)

1.111 F. Sander: Experimentelle Ergebnisse der Gestaltpsychologie. In: *Bericht über den 10. Kongress für experimentelle Psychologie, Bonn 1927* (G. Fischer, Jena 1928) pp. 23–88

1.112 K. Koffka: *Principles of Gestalt Psychology* (Routledge Kegan Paul, London 1936)

1.113 A. Truslit: *Gestaltung und Bewegung in der Musik* (Vieweg, Berlin 1938)

1.114 R.I. Godøy, M. Leman (Eds.): *Musical Gestures. Sound, Movement, and Meaning* (Routledge, New York, London 2010)

1.115 C. Seashore: *The Psychology of Musical Talent* (Silver Burdette, Boston 1919)

1.116 C. Seashore: *Psychology of Music* (McGraw-Hill, New York 1938), repr. (Dover, New York 1967)

1.117 C. Seashore (Ed.): *Studies in the Psychology of Music*, Vol. 4 (Univ. of Iowa, Iowa City 1936)

1.118 C. Stumpf, M. Meyer: Maassbestimmungen über die Reinheit consonanter Intervalle. In: *Beiträge zur Akustik und Musikwissenschaft*, Vol. 2 (Barth, Leipzig 1898) pp. 84–167

1.119 C. Seashore: *In Search of Beauty in Music. A Scientific Approach to Musical Esthetics* (Ronals, New York 1947)

1.120 H. Weld: An experimental study of musical enjoyment, Am. J. Psychol. **23**, 245–308 (1912)

1.121 K. Hevner: The affective character of the major and minor modes in music, Am. J. Psychol. **47**, 103–118 (1935)

1.122 K. Hevner: Experimental studies of the elements of expression in music, Am. J. Psychol. **48**, 246–268 (1936)

1.123 K. Hevner: The affective value of pitch and tempo in music, Am. J. Psychol. **49**, 621–630 (1937)

1.124 R. Gundlach: Factors determining the characterization of musical phrases, Am. J. Psychol. **47**, 624–643 (1935)

1.125 G. Rubin-Rabson: The influence of age, intelligence, and training on reaction to classic and modern music, J. Gen. Psychol. **22**, 413–429 (1940)

1.126 M. Horkheimer, T.W. Adorno: *Dialektik der Aufklärung* (S. Fischer, Amsterdam 1947), rev. edn. S. Fischer, Frankfurt/M. 1971

1.127 T.W. Adorno: *Einleitung in die Musiksoziologie. Zwölf theoretische Vorlesungen* (Suhrkamp, Frankfurt/M. 1962), also (Rowohlt, Reinbek 1968)

1.128 P. Farnsworth: *The Social Psychology of Music* (Dryden, New York 1958), 2nd edn. (Iowa State Univ. Press, Ames 1969)

1.129 P. Brömse, E. Kötter: *Zur Musikrezeption Jugendlicher. Eine psychometrische Untersuchung* (Schott, Mainz 1971)

1.130 K.E. Behne: *Hörertypologien. Zur Psychologie des jugendlichen Musikgeschmacks* (Bosse, Regensburg 1986)

1.131 V. Karbusicky: *Empirische Musiksoziologie* (Breitkopf Härtel, Wiesbaden 1975)

1.132 R. Francès: *The Perception of Music* (Erlbaum, Hillsdale 1988), translated from R. Francès: *La Perception de la Musique* (Vrin, Paris 1958)

1.133 H.P. Reinecke: *Experimentelle Beiträge zur Psychologie des musikalischen Hörens* (Sikorski, Hamburg 1964)

1.134 P. Schaeffer: *Traité des objets musicaux* (Ed. du Seuil, Paris 1966)

1.135 L. Hiller, L. Isaacson: *Experimental Music. Composition with an Electronic Computer* (McGraw-Hill, New York 1959)

1.136 A. Moles: *Théorie de l'information et perception esthetique* (Flammarion, Paris 1958)

1.137 J. Pierce: *Signals, Symbols and Noise. The Nature and Process of Communication* (Harper Row, New York 1961)

1.138 F. Winckel: Die informationstheoretische Analyse musikalischer Strukturen, Musikforschung **17**, 1–14 (1964)

1.139 L. Hiller, C. Bean: Information theory analyses of four sonata expositions, J. Music Theory **10**, 96–137 (1966)

1.140 C. Roads, J. Strawn (Eds.): *Foundations of Computer Music* (MIT Press, Cambridge 1985)

1.141 M. Mathews, J. Pierce (Eds.): *Current Directions in Computer Music Research* (MIT Press, Cambridge 1989)

1.142 G. De Poli, A. Piccialli (Eds.): *Representations of Musical Signals* (MIT Press, Cambridge 1991)

1.143 C. Roads, S. Pope, A. Piccialli, G. De Poli (Eds.): *Musical Signal Processing* (Swets Zeitlinger, Lisse, Abingdon 1997)

1.144 J. Beauchamp (Ed.): *Analysis, Synthesis, and Perception of Musical Sound* (Springer, New York 2007)

1.145 J. Smith III: Virtual acoustic musical instruments: Review and update, J. New Music Res. **33**, 283–304 (2004)

1.146 R. Bader: *Nonlinearities and Synchronization in Musical Acoustics and Music Psychology* (Springer, Cham 2013)

1.147 M. Balaban, K. Ebcioğlu, O. Laske (Eds.): *Understanding Music with AI: Perspectives on Music Cognition* (AAA/MIT Press, Cambridge, Menlo Park 1992)

1.148 U. Seifert: *Systematische Musiktheorie und Kognitionswissenschaft* (Verlag für systematische Musikwissenschaft, Bonn 1993)

1.149 J. Bharucha, N. Todd: *Music and Connectionism* (MIT Press, Cambridge 1991)

1.150 M. Leman: *Music and Schema Theory* (Springer, Berlin 1995)

1.151 P. Toiviainen: *Modelling Musical Cognition with Artifical Neural Networks*, PhD. Thesis (Univ. of Jyväskylä, Jyväskylä 1996)

1.152 P. Zanon, G. de Poli: Estimation of time-varying parameters in rule systems for music performance, J. New Music Res. **32**, 295–315 (2003)

1.153 A. Friberg, R. Bresin, J. Sundberg: Overview of the KTH rule system for musical performance, Adv. Cogn. Psychol. **2**, 145–161 (2006)

1.154 A. Klapuri, M. Davy (Eds.): *Signal Processing Methods for Music Transcription* (Springer, New York 2006)

1.155 S. Ystad, R. Kronland-Martinet, K. Jensen (Eds.): *Computer Music Modeling and Retrieval. Genesis of Meaning in Sound and Music* (Springer, Berlin 2009)

1.156 D. Conklin, R. Ramirez, J. Iñesta (Eds.): Music and machine learning. J. New Music Res. (Special issue) **43**(3/4), 251–399 (2014)

1.157 W. Braun: *Deutsche Musiktheorie des 15. bis 17. Jahrhunderts*, Geschichte der Musiktheorie, Vol. 8/II (Wissenschaftliche Buchgesellschaft, Darmstadt 1994)

1.158 J. Kirnberger: *Die Kunst des reinen Satzes in der Musik. Aus sicheren Grundsätzen hergeleitet*, Vol. I/II (Lagarde, Berlin, Königsberg 1776–1779) repr. (Olms, Hildesheim 2010)

1.159 N. Ruwet: *Langage, musique, poésie* (Ed. du Seuil, Paris 1972)

1.160 J.J. Nattiez: *Fondements d'une sémiologie de la musique* (Union gén. d'éditions, Paris 1975)

1.161 M. Bierwisch: *Musik und Sprache. Überlegungen zu ihrer Struktur und Funktionsweise*, Jahrbuch Peters (1978) (Ed. Peters, Leipzig 1979) pp. 9–102

1.162 R. Schneider: *Semiotik der Musik* (Fink, München 1980)

1.163 P. Faltin: *Bedeutung ästhetischer Zeichen. Musik und Sprache* (Rader, Aachen 1985)

1.164 V. Karbusicky: *Grundriß der musikalischen Semantik* (Wissenschaftliche Buchgesellschaft, Darmstadt 1986)

1.165 V. Karbusicky: The experience of the indexical sign: Jakobson and the semiotic phonology of Leoš Janáček, Am. J. Semiot. **2**, 35–58 (1983)

1.166 C. Seeger: On the moods of a music-logic, J. Am. Mus. Soc. **13**, 224–261 (1960)

1.167 N. Chomsky: *Aspects of the Theory of Syntax* (MIT Press, Cambridge 1965)

1.168 H. Riemann: Ideen zu einer 'Lehre von den Tonvorstellungen', Jahrbuch Peters **21/22** (1914/15), 1–26

1.169 P. Smolensky, G. Legendre: *Cognitive Architecture*, The Harmonic Mind. From Neural Computation to Optimality-Theoretic Grammar, Vol. 1 (MIT Press, Cambridge 2006)

1.170 R. Penrose: *The Emperor's New Mind. Concerning Computers, Minds, and the Laws of Physics* (Oxford Univ. Press, Oxford 1989)

1.171 G. Gigerenzer, D. Murray: *Cognition as Intuitive Statistics* (Erlbaum, Hillsdale 1987)

1.172 M. Leman: Naturalistic approaches to musical semiotics and the study of causal musical signification. In: *Music and Signs. Semiotic and Cognitive Studies in Music*, ed. by I. Zannos (ASCO Art Science, Bratislava 1999) pp. 11–38

1.173 H. Gardner: *The Mind's New Science* (Basic Books, New York 1987)

1.174 M. Clynes (Ed.): *Music, Mind, and Brain* (Plenum, London 1982)

1.175 D. Deutsch (Ed.): *Psychology of Music* (Academic, Orlando, San Diego 1982)

1.176 J. Sloboda: *The Musical Mind. The Cognitive Psychology of Music* (Clarendon, Oxford 1985)

1.177 J. Sloboda (Ed.): *Generative Processes in Music. The Psychology of Performance, Improvisation, and Composition* (Clarendon, Oxford 1988)

1.178 P. Howell, I. Cross, R. West (Eds.): *Musical Structure and Cognition* (Academic, Orlando 1985)

1.179 J. Dowling, D. Harwood: *Music Cognition* (Academic, Orlando 1986)

1.180 S. Handel: *Listening. An Introduction to the Perception of Auditory Events* (MIT Press, Cambridge 1989)

1.181 C. Krumhansl: *Cognitive Foundations of Musical Pitch* (Oxford Univ. Press, New York 1990)

1.182 P. Howell, R. West, I. Cross (Eds.): *Representing Musical Structure* (Academic, Orlando 1991)

1.183 S. McAdams, E. Bigand (Eds.): *Thinking in Sound. The Cognitive Psychology of Human Audition* (Clarendon, Oxford 1993)

1.184 J. Louhivuori, J. Laaksamo (Eds.): *Proc. 1st Int. Conf. Cogn. Music.* (Univ. of Jyväskylä, Jyväskylä 1993)

1.185 M. Leman (Ed.): *Music, Gestalt, and Computing. Studies in Cognitive and Systematic Musicology* (Springer, Berlin 1997)

1.186 O. Elschek, A. Schneider (Eds.): *Ähnlichkeit und Klangstruktur / Similarity and Sound Structure*, Syst. Musikwiss. – Syst. Musicol. IV, Vol. 1–2 (ASCO, Bratislava 1996) pp. 1–376

1.187 I. Zannos (Ed.): *Music and Signs. Semiotic and Cognitive Studies in Music* (ASCO Art Science, Bratislava 1999)

1.188 R.I. Godøy, H. Jørgensen (Eds.): *Musical Imagery* (Swets Zeitlinger, Lisse, Abingdon 2000)

1.189 B. Snyder: *Music and Memory* (MIT Press, Cambridge 2000)

1.190 P. Desain, L. Windsor (Eds.): *Rhythm Perception and Production* (Swets Zeitlinger, Lisse, Abingdon 2000)

1.191 P. Juslin, J. Sloboda (Eds.): *Music and Emotion. Theory and Research* (Oxford Univ. Press, Oxford 2001)

1.192 I. Deliège, G. Wiggins (Eds.): *Musical Creativity. Multidisciplinary Research in Theory and Practice* (Psychology Press, Hove, New York 2006)

1.193 M. Leman: *Embodied Music Cognition and Mediation Technology* (MIT Press, Cambridge 2008)

1.194 A. Bregman: *Auditory Scene Analysis. The Perceptual Organization of Sound* (MIT Press, Cambridge 1990)

1.195 M. Clayton, R. Sager, U. Will: In time with the music: The concept of entrainment and its significance for ethnomusicology. In: *European Meetings in Ethnomusicology*, Vol. 11, ed. by U. Will (Durham Univ., Durham 2005) pp. 1–143

1.196 P. Toiviainen, G. Luck, M. Thompson: Embodied meter: Hierarchical eigenmodes in music-induced movement, Music Percept. **28**, 59–70 (2010)

1.197 M. Critchley, R.A. Henson (Eds.): *Music and the Brain. Studies in the Neurology of Music* (W. Heinemann, London 1977)

1.198 R. Spintge, R. Droh (Eds.): *Musik in der Medizin. Music in Medicine* (Springer, Berlin 1987)

1.199 H. Petsche (Ed.): *Musik – Gehirn – Spiel* (Birkhäuser, Basel 1989)

1.200 J. Sergent: Human brain mapping. In: *MusicMedicine*, Vol. 2, ed. by R. Rebollo Pratt, R. Spintge, R. Droh (MMB, St. Louis 1996) pp. 24–49

1.201 G. Avancino (Ed.): *The Neurosciences and Music*, Annals of the N.Y. Acad. of Sciences, Vol. 999 (N.Y. Acad. of Sciences, New York 2003)

1.202 G. Avancino (Ed.): *The Neurosciences and Music II: From Perception to Performance*, Annals of the N.Y. Acad. of Sciences, Vol. 1060 (N.Y. Acad. of Sciences, New York 2005)

1.203 A. Patel: *Music, Language, and the Brain* (Oxford Univ. Press, New York 2008)

1.204 C. Allesch: *Geschichte der psychologischen Ästhetik* (Hogrefe, Göttingen 1987)

1.205 G. Adler: Umfang, Ziel und Methode der Musikwissenschaft, Vierteljahrsschr. Musikwiss. **1**, 5–20 (1885)

1.206 C. Stumpf: *Anfänge der Musik* (Barth, Leipzig 1911)

1.207 C. Sachs: *The Wellsprings of Music*, ed. by J. Kunst (Nijhoff, The Hague 1962)

1.208 H. Böttcher, U. Kerner: *Methoden in der Musikpsychologie* (Ed. Peters, Leipzig 1978)

1.209 J. Beran: *Statistics in Musicology* (Chapman Hall, CRC, Boca Raton 2004)

1.210 A. Schneider: Systematische Musikwissenschaft: Traditionen, Ansätze, Aufgaben, Syst. Musikwiss. – Syst. Musicol. **1**, 145–180 (1993)

1.211 W. Graf: *Vergleichende Musikwissenschaft* (Stiglmayr, Wien 1980), ed. by F. Födermayr

1.212 F. Födermayr, W. Deutsch: Zur Forschungsstrategie der vergleichend-systematischen Musikwissenschaft, Musicol. Austr. **17**, 163–180 (1998)

1.213 B. Nettl: *The Study of Ethnomusicology* (Univ. of Illinois Press, Champaign, Chicago 1983), 3rd edn. 2015

1.214 A.D. Fokker: *New Music with 31 Notes* (Verlag für systematische Musikwissenschaft, Bonn 1975)

1.215 C. Seeger: Systematic musicology: Viewpoints, orientations, and methods, J. Am. Musicol. Soc. **4**, 240–248 (1951)

1.216 C. Seeger: *Studies in Musicology*, Vol. 1 (Univ. California Press, Berkeley 1977)

1.217 A. Wellek: Begriff, Aufbau und Bedeutung einer systematischen Musikwissenschaft, Musikforschung **1**, 157–171 (1948)

1.218 W. Wiora: Historische und systematische Musikforschung. Thesen zur Grundlegung ihrer Zusammenarbeit, Musikforschung **1**, 171–191 (1948)

1.219 V. Karbusicky: *Systematische Musikwissenschaft* (Fink, München 1976)

1.220 O. Laske: Introduction to cognitive musicology, J. Music. Res. **9**, 1–22 (1989)

1.221 N. Wallin: *Biomusicology. Neurophysiological, Neuropsychological, and Evolutionary Perspectives on the Origins and Purposes of Music* (Pendragon, Stuyvesant 1991)

1.222 M. Dobberstein: *Musik und Mensch. Grundlegung einer Anthropologie der Musik* (Reimer, Berlin 2000)

1.223 A. Tanguine: *Artificial Perception and Music Recognition* (Springer, Berlin 1993)

1.224 A. Marsden: *Representing Musical Time. A Temporal-Logic Approach* (Swets Zeitlinger, Lisse, Abingdon 2000)

1.225 G. Mazzola, S. Göller: *The Topos of Music. Geometric Logic of Concepts, Theory, and Performance* (Birkhäuser, Basel 2002)

Part A Musical Acoustics and Signal Processing

Ed. by Rolf Bader

In the Hornbostel/Sachs classification of musical instruments first published in 1914, the instrument families were ordered according to their physical driving mechanism, such as bowed, plucked, or struck instruments. This scheme reflects the nature of systematic musicology, which aims to search for universals in music that govern all musical genres around the world.

Still, musical acoustics itself is much older. Joseph Sauveur (1653–1716), who coined the term acoustics in the sense we use it today, was already sorting vibrating systems in terms of strings, membranes or plates at this time, well in accordance with musical instruments. This also reflects the close correspondence between music, physics, geometry and arithmetic, which was present in ancient times with the Pythagoreans or in the music theory of Archytas of Tarrent (428–347). Music was part of the Quadrivium in Renaissance times, for example in the music theory of Johannes Kepler in his *Harmonices Mundi* from 1619. Music was also often close to mathematics and geometry in the work of mathematicians like Mersenne, Zarlino, Euler or Gauss, philosophers like Rene Descartes or physicians like Hermann von Helmholtz, who in his *On the sensations of tone as a physiological basis for the theory of music* from 1863, was able to derive the Western tonal system from acoustical, physiological and psychological findings.

However, investigations of musical instruments in detail were only prominently carried out first by Felix Savart (1791–1841) on the violin. Savart, known to physicists due to the Biot–Savart law of magnetostatics, also invented an octobass: a huge double bass so high that it needed two floors and two players, one bowing on the first floor and one pressing down the strings on the second. Later, Helmholtz was the first to visualize the sawtooth motion of a violin string by stroboscopy. He and Lord Rayleigh were also among the first to investigate the organ theoretically as well as experimentally.

After World War II research on musical instruments intensified, as is reflected in the work of the *Cutgut Acoustical Society* in England, *the Physikalisch-Technische Bundesanstalt* in Braunschweig or the *Institut de Recherche et Coordination Acoustique/Musique* (*IRCAM*) in Paris and by many individuals mainly in the US, Canada, Europe, Australia and Japan. Other parts of the world where musicological research is performed were concentrating not so much on musical acoustics but more on music ethnology, such as in China or India, although an Indian researcher, Chandrasekhara Venkata Raman (1888-1970) was the first to develop a theory on the bowed string.

Today musical acoustics, which has been growing tremendously over the last decades, is a very active and lively scene with increasing numbers of researchers. The instrument industry, which has always been working in this field in very close contact with researchers, is profiting from the results in musical acoustics, which is also reflected in patents and applications improving instrument quality and design. Indeed many myths about musical instruments are around among instrument builders, and musical acoustics may help here too. Related industries like the software-developing market, which has been building synthesizers and virtual musical instruments in recent years, have increasingly profited from models of musical instruments developed in musical acoustics to realize more realistic-sounding plugins or synthesizers, mainly in the domain of physical modeling.

After about two hundred years of research in musical acoustics, many open questions are still present in the field. The present models, although often sounding very realistic, are still quite restricted to certain instrument types or articulations. A basic understanding of the role of turbulence in wind instruments is an ongoing debate since many researchers, starting from Helmholtz and Rayleigh, found the organ to be a linear system that does not correspond to our understanding of turbulence. The role of forced oscillations in musical instruments and the difference between the eigenmodes of guitar and violin bodies and the forced oscillation patterns found empirically when driving the instrument with strings are also an ongoing debate. Furthermore the radiation of a musical instrument, its dependency on the driving point, or its transient nature are not fully understood yet. The role of material, its stiffness and, maybe even more importantly, its damping, is not yet understood at all. This list could be continued. So not only are details of musical instruments still under debate, but often the basic process of tone production is still discussed.

The reason for these ongoing debates may be seen in the extreme sensitivity of the human ear. We judge the quality of an instrument not by just producing sound at all; we are interested and fascinated by the details of the sound, the range of articulatory possibilities of an instrument, its character and *musicality*. Therefore musical acoustics is a discipline with many open questions and hopefully many fascinating new results in the future.

The present section gives a systematic overview on basics in the field as well as addressing open questions and providing insight into ongoing debates. The choice of topics is also related to the position of musical acoustics in the field of systematic musicology: opening links to signal processing and applications via mathematical

modeling of instruments, music ethnology through discussion of wood properties, basic driving mechanisms or non-Western instruments, or music psychology by discussing timbre- or articulation-related aspects of the instruments. In the end, musical instruments are the source of a musical performance and provide the possibilities a musician might use to express themself. Very often tools become music and improvements to the instruments only triggered new musical styles, sounds and extended techniques.

In Chap. 2 *Wilfried Kausel* gives a comprehensive and systematic mathematical introduction to musical acoustics starting with one-dimensional vibrating systems. After introducing a mass-spring system and deriving complex numbers he considers the wave equation on strings and in ducts including reflections, forced oscillations and impedance. He also discusses stiffness in one-dimensional media and longitudinal vibrations in bars, explaining many mathematical functions in detail as used in musical acoustics on an everyday basis.

Wilfried Kausel then enlarges the one-dimensional picture of Chap. 2 into a two- and three-dimensional one in Chap. 3. He derives and discusses the mathematics of membranes as used in drums or string instruments like the banjo and turns over to stiff two-dimensional geometries like rectangular and circular plates as basic components of stringed instruments. He then gives the basic equations for three-dimensional geometries as cavities or rooms and discusses modes and resonances.

Switching from abstract mathematical models to basic material, in Chap. 4 *Chris Waltham* and *Shigeru Yoshikawa* discuss wood used for musical instruments, mainly for stringed, wind and percussion instruments. After a summary of common tonewoods they discuss many aspects of the relation between wood and the sound of the instrument. They conclude that even if knowledge about wood does not guarantee a high quality sound from the instrument, it still helps to avoid many mistakes and errors in instrument building.

Measuring musical instrument vibrations is nontrivial and *Thomas Moore* in Chap. 5 discusses the techniques used in the field. He discusses the use of microphone arrays in terms of acoustic holography, where the vibrations of radiating musical instrument surfaces are measured using multiple microphones recording the radiated sound, which is then back-propagated to the instrument surface. Another powerful tool introduced is laser Doppler interferometry, where deflections of the instrument surface are measured as phase shifts in a split laser beam. Finally, accelerometer measurements are discussed using piezoelectric crystals attached to the instruments.

Turning then to stringed instruments, *Nicholas Giordano* in Chap. 6 gives a brief overview of pianos, guitars and violins. He introduces the basic differential equations of the instruments while discussing the driving mechanisms, the energy transfer and the modes of vibrations. The paper is rich with many complex and interesting phenomena associated with the instruments, which cannot always be discussed in detail but where the appropriate literature is referenced.

Turning to wind instruments, *Benoit Fabre*, *Avraham Hirschberg* and *Joël Gilbert* in Chap. 7 give an overview of the clarinet, the oboe, the harmonica, the trombone, and the modern transverse flute. The instruments are classified along a general model system consisting of a nonlinear generator and a linear resonator holding as a general principle for the very diverse systems of single and double reed and wind jet instruments. The paper presents models that can easily be implemented as physical models, which can end in sound-producing software tools.

In Chap. 8, *András Miklós* and *Judit Angster* present measurements and a model of the organ pipe. They emphasize the role of the attack transient, the very first beginning of the sound, which perceptually is the most salient part of musical tones in general. The instrument is discussed as a coupled system of a hydrodynamically oscillating air jet and a resonating pipe with wall losses. The measurements of the attack show a complex transient phase known to organ builders as *chiff*, which is known to be a quality criterion for organ pipes.

In Chap. 9, *Andrew Morrison* and *Thomas Rossing* give an introduction to percussion instruments with many examples, as well as theoretical and experimental findings. The drums of a rock or jazz drum set with snare, bass drum and tom-toms are discussed as well as instruments from other parts of the world like Caribbean steelpans, Indian tablas, Chinese stone chimes, Indonesian gongs or Japanese drums. Many laser interferometry measurements of the instruments' eigenmodes are shown and systematic investigations in terms of tuning systems and strike notes are discussed.

Discussing the role of nonlinearities in musical instruments in Chap. 10, *Rolf Bader* distinguishes between nonlinearities that lead to an enhancement of the brightness of musical instrument timbres and those nonlinearities that are crucial for musical tone production. He finds synchronization in harmonic overtone structures in wind instruments as a result of synchronization

within these instruments driven by turbulent damping. Applying this finding to other instrument families, a general understanding of musical instruments as self-organized systems is possible.

The section closes in Chap. 11 with *Michael Vorländer* presenting room acoustics as the link between musical instruments and the audience in a concert hall. He discusses basic aspects of room acoustics while introducing ray-tracing algorithms for modeling concert spaces. Using the resulting impulse response as a filter function for a dry musical signal in a fast convolution algorithm, he shows that the acoustic space can be reproduced in silico, making it possible for architects to estimate the room acoustics during the process of designing a space.

2. Vibrations and Waves

Wilfried Kausel

This chapter deals with vibration and wave propagation under the general assumption that amplitudes are sufficiently small in order to neglect nonlinear effects when vibrations or waves are superimposed. It will be shown how wave equations can be derived for strings, bars and air columns and how analytic results can be obtained for some boundary conditions. This chapter will also review techniques for the calculation of resonance frequencies. Finally an introduction into the analysis of real musical instruments in the frequency domain will be given.

2.1 Vibrations

Musical instruments are built to create sound. The musician typically sets some part of it into vibration and this vibration is coupled to other parts radiating sound waves perceived by a listener. For example, a violinist uses his bow to set a string into vibration. This vibration is coupled to the bridge and from there to the body of the violin. The vibrating plates are large enough to initiate a sound wave propagating in the surrounding air towards the ears of listeners.

The terms *vibration* and *wave* are often not used in a consistent way. Vibration obviously describes an oscillating state, a physical quantity like displacement, velocity, air pressure, or a spatial profile or a deformation, varying periodically around an equilibrium state, which will be reached when no stimulating force is present and after any existing vibration has decayed completely.

The most simple vibrating system is hypothetical. It consists of a point mass and an ideal spring. The vibrating characteristics are displacement, velocity and acceleration of the mass as well as force and displacement of the spring, the latter two being perfectly proportional in an ideal spring. The vibration itself is characterized by the period, which is the time duration of one complete cycle, and its reciprocal value, the frequency, which is the number of cycles per second. Each oscillating quantity has its amplitude, which can be indicated as a peak value or as an RMS value (root mean square integrated over an integer number of periods), which is related to the average power of that quantity.

Waves are characterized by the fact that they propagate at a certain speed. They occur if a medium (or space itself in the case of electromagnetic waves) provides a continuum of infinitesimal small oscillatory systems, which are locally coupled in a way to pass on their own vibrational state to adjacent systems. Waves are characterized by their propagation speed, their amplitude and their wavelength.

In reality vibrations and waves are strongly connected. Looking at a wavy water surface one can ob-

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_2

serve numerous floating particles or froth oscillating up and down. Observing carefully one will recognize that a vibrating string is actually a nearly one-dimensional continuum where perturbances propagate as a wave in both directions until they are reflected at the ends. Vibrating reeds and lips of brass players consist of spatially distributed mass and stiffness and their apparent *vibration* is in fact based on wave propagation along their length and respectively across their surface.

Nevertheless, physicists often successfully treat such systems as hypothetical mass–spring systems, especially if effects due to their vibration-like properties are to be studied and their wave-like behavior can be ignored.

2.1.1 Mass–Spring Systems

A mass–spring system is governed by an equation of motion where any inertial force $F_\mathrm{i} = ma = m\mathrm{d}v/\mathrm{d}t = m\mathrm{d}^2z/\mathrm{d}t^2$ of a point mass m (displacement z, velocity v, acceleration a) is always balanced by a spring force $F_\mathrm{s} = -Ez$, which acts against the inertial force and which has a magnitude proportional to the displacement z with proportionality constant E, known as the *spring constant*. We obtain the linear second-order ordinary differential equation

$$m\frac{\mathrm{d}^2z}{\mathrm{d}t^2} + Ez = 0\,. \tag{2.1}$$

With some good intuition and perhaps some trial and error we find a solution for z (or look it up in any book)

$$\begin{aligned} z_1(t) &= \sin\omega t \\ \frac{\mathrm{d}z_1}{\mathrm{d}t} &= \omega\cos\omega t \\ \frac{\mathrm{d}^2z_1}{\mathrm{d}t^2} &= -\omega^2\sin\omega t\,. \end{aligned} \tag{2.2}$$

Substituting this solution into (2.1) proves the correctness of our initial formulation and yields an expression for the angular frequency $\omega = 2\pi f$

$$\begin{aligned} -m\omega^2\sin\omega t + E\sin\omega t &= 0 \\ \omega &= \sqrt{\frac{E}{m}}\,. \end{aligned} \tag{2.3}$$

As it works with $\sin\omega t$ we are tempted to try the same with $\cos\omega t$

$$\begin{aligned} z_2(t) &= \cos\omega t \\ \frac{\mathrm{d}z_2}{dt} &= -\omega\sin\omega t \\ \frac{\mathrm{d}^2z_2}{\mathrm{d}t^2} &= -\omega^2\cos\omega t\,. \end{aligned} \tag{2.4}$$

We can see that both solutions work in a similar way. Knowing that any linear combination of possible solutions is again a solution, we can formulate a more general solution according to

$$z(t) = A\sin\omega t + B\cos\omega t\,. \tag{2.5}$$

It is often convenient to rewrite (2.5) as

$$z(t) = C\cos(\omega t + \phi) \tag{2.6}$$

thus replacing the linear combination $A\sin x + B\cos x$ by the single harmonic function $C\cos(x+\phi)$ with an amplitude $C = \sqrt{A^2+B^2}$ and a phase angle ϕ satisfying $\tan(\phi + \pi/2) = B/A$.

This result tells us that the displacement of the mass (or the length of the spring) oscillates harmoniously with an angular frequency ω, which only depends on the stiffness E of the spring and the mass m. The amplitude C and the initial phase ϕ can be determined when the initial conditions $z(0)$, the initial displacement, and $z'(0)$, the initial velocity, are known.

Differentiating the solution for $z(t)$ from (2.6) twice gives us solutions for the velocity $v(t)$ and acceleration $a(t)$, both again harmonic functions with the same angular frequency ω but different amplitudes.

$$\begin{aligned} z'(t) = v(t) &= -\omega C\sin(\omega t+\phi) \\ &= \omega C\cos\left(\omega t + \phi + \frac{\pi}{2}\right) \\ z''(t) = a(t) &= -\omega^2 C\cos(\omega t+\phi) \\ &= \omega^2 C\cos(\omega t+\phi+\pi)\,. \end{aligned} \tag{2.7}$$

The amplitude of a sinusoidal velocity at a given frequency ω can be derived by multiplying the amplitude C of the displacement by ω. To get the acceleration a multiplication with ω^2 is required. Each differentiation step shifts the phase by $\pi/2 = 90°$.

Using Euler's formula $\mathrm{e}^{\mathrm{i}x} = \cos x + \mathrm{i}\sin x$ (imaginary unit $\mathrm{i} = \sqrt{-1}$) and $\hat{C} = C\mathrm{e}^{\mathrm{i}\phi}$ (2.6) can be transformed into exponential form (with Re being the real part of a complex number, and complex numbers marked by the superscript symbol ˆ) according to

$$z(t) = \mathrm{Re}(C\,\mathrm{e}^{\mathrm{i}(\omega t+\phi)}) = \mathrm{Re}(\hat{C}\,\mathrm{e}^{\mathrm{i}\omega t})\,. \tag{2.8}$$

It can easily be verified that not only is the real part of the complex exponential from (2.8) a solution of our differential equation but so too is the imaginary part, and even the complex exponential itself

$$\begin{aligned} z(t) &= \hat{C}\,\mathrm{e}^{\mathrm{i}\omega t} \\ \frac{\mathrm{d}z}{\mathrm{d}t} &= \mathrm{i}\omega\hat{C}\,\mathrm{e}^{\mathrm{i}\omega t} \\ \frac{\mathrm{d}^2z}{\mathrm{d}t^2} &= -\omega^2\hat{C}\,\mathrm{e}^{\mathrm{i}\omega t}\,. \end{aligned} \tag{2.9}$$

This makes differentiation and integration very easy and allows for dealing with constant phase differences in a very elegant way. The magnitude of the complex coefficient $\hat{C}$ represents a scale factor for the unit amplitude of $e^{i\omega t}$ and its time-invariant phase angle ϕ adds like an initial phase to any momentary phase of $e^{i\omega t}$.

It should be remembered here that complex numbers are multiplied by multiplying their magnitudes and adding their phases. To only scale an amplitude the complex exponential has to be multiplied with a real number having 0 phase. To only phase shift a complex exponential it has to be multiplied with another complex exponential $e^{i\phi}$, which represents the phase angle ϕ and has a magnitude of unity.

Imaginary Exponents

It is important to understand Euler's correspondence because later on complex exponentials will be used very often to represent harmonic functions of time or space. From the Taylor series expansions we see

$$
\begin{aligned}
e^{z} &= 1 + \frac{z}{1!} + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{z^n}{n!} \\
\cos z &= 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \cdots = \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n}}{(2n)!} \\
\sin z &= z - \frac{z^3}{3!} + \frac{z^5}{5!} - \cdots = \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{(2n+1)!} \, .
\end{aligned}
\tag{2.10}
$$

Euler discovered that the series representing e^{ix} could also be obtained when the series expansions of $\cos x$ and $i \sin x$ are added (published in 1748 [2.1]). This can be easily verified by substituting $z \to ix$ in the first and $z \to x$ in the second and third row of (2.10).

The current notion that a complex number represents a point in the complex plane was developed more than half a century after Euler's formula, and it was *Gauss* in 1831 [2.2] who recognized the importance of that interpretation.

Each complex number has a real part and an imaginary part, which is a multiple of the imaginary unit i. The complex plane has a real and an imaginary axis corresponding to the x and y-axis of other coordinate systems.

The complex term $e^{i\theta}$ with its real part $\cos\theta$ and its imaginary part $\sin\theta$ describes one point on the unit circle in the complex plane (Fig. 2.1). As $\theta = \omega t + \phi$ increases with time t by an interval of 2π the point will describe one full cycle on the unit circle. At time $t = 0$ the momentary phase angle θ is the initial phase angle ϕ.

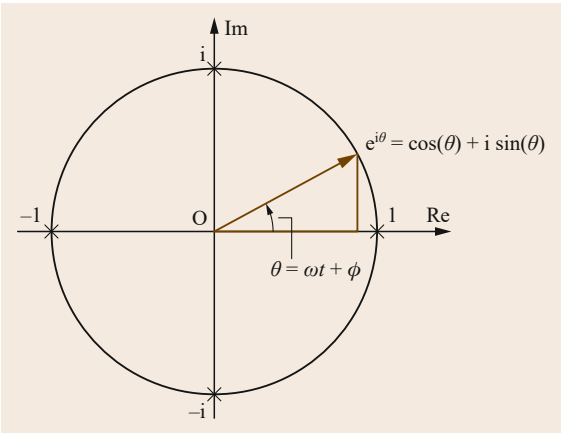


Fig. 2.1 Complex exponentials

We can also say, $\hat{C}\,e^{i\omega t}$ is a rotating pointer in the complex plane. It rotates around the origin as time t increases. Its length $|\hat{C}|$ corresponds to the amplitude of the harmonic function and the pointer's angle with the positive real axis is the momentary phase. The momentary signal displacement is the real part of the pointer. The imaginary part of the results can be simply disregarded. Its mere purpose was to allow the use of exponential terms, which can be differentiated, integrated, and multiplied very nicely.

If $\hat{C}$ is complex and not real then an initial phase at $t = 0$ can be specified. The constant phase angle of the complex amplitude $\hat{C}$ is simply added to the phase of the rotating pointer. A signal with amplitude 10 and 0 phase at $t = 0$ like $10\sin(\omega t)$ can be written as $\mathrm{Re}(-10ie^{i\omega t})$. The complex amplitude $-10i$ scales the signal amplitude from unity and shifts the initial phase by $-\pi/2$. This is necessary because for this case we want to represent $\sin(\omega t)$ and not $\cos(\omega t)$, the real part according to Euler.

The term $e^{-i\omega t}$ itself is a unity pointer, rotating with the same speed but in the opposite direction. If we add $e^{i\omega t}$ and $e^{-i\omega t}$ the imaginary parts always cancel out while the real parts add up. The transition from the complex domain to the domain of real physical quantities and harmonic functions according to (2.8) can therefore also be made according to

$$
z(t) = \frac{C}{2}\left(e^{i(\omega t+\phi)} + e^{-i(\omega t+\phi)}\right) = C\cos(\omega t+\phi) \, .
\tag{2.11}
$$

One very nice outcome of using this transformation is that $de^{i\omega t}/dt = i\omega e^{i\omega t}$ and $d^2e^{i\omega t}/t^2 = -\omega^2 e^{i\omega t}$. This translates differential equations in the time domain into simple algebraic equations.

2.1.2 Forced Vibration

Now we extend the equation of motion for a single mass system (2.1) by adding an external driving force F and a damping force that acts against the inertial force, which is proportional to the velocity $v = \mathrm{d}z/\mathrm{d}t$ with a proportionality constant D

$$m\frac{\mathrm{d}^2 z}{\mathrm{d}t^2} = F - D\frac{\mathrm{d}z}{\mathrm{d}t} - Ez\,. \tag{2.12}$$

The equation of motion (2.12) has been reordered here to make it more clear that any change of momentum (mass times acceleration) is due to a resulting force acting on the mass. The resulting force is the sum of an external driving force F and two other forces, a frictional and an elastic force, both of them resisting the movement.

Assuming a harmonic driving force $F = \hat{F}\,\mathrm{e}^{\mathrm{i}\omega t}$ we can again try the solution given by $z = \hat{Z}\,\mathrm{e}^{\mathrm{i}\omega t}$, $\hat{F}$ and $\hat{Z}$ being complex coefficients determining the amplitude and initial phase of the oscillating force F and displacement z. This leads to an equation for $\hat{Z}$

$$-\omega^2 m\hat{Z} = \hat{F} - \mathrm{i}\omega D\hat{Z} - E\hat{Z} \tag{2.13}$$

with a solution

$$\hat{Z}(\omega) = \frac{\hat{F}}{E + \mathrm{i}\omega D - m\omega^2}\,. \tag{2.14}$$

By substituting $D \to |\hat{F}|/(\omega_0 Q)$ and $\omega \to \omega_0$ into (2.14) it is easy to show that the damping factor D is related to a resonance quality factor Q, which is the amplitude near resonance where $\omega = \omega_0$, according to $Q = |F|/(\omega_0 D)$ with $\omega_0 = \sqrt{E/m}$, the eigenfrequency of the undamped free system according to (2.3).

Amplitude and phase of $\hat{Z}$ have been plotted in Fig. 2.2 for $m = 1, F = 1, E = 1$ and $Q = 5$. It can be seen that there is a maximum near the eigenfrequency ω_0 of the undamped system. At this frequency there is also a phase transition from 0 to $-\pi$. If the damping is low, Q and therefore the resonance amplitude can become very high. On the other hand, when there is high damping and low Q, no resonance occurs but there is still a phase of $-\pi/2$ and an amplitude of Q at $\omega = \omega_0$.

2.1.3 Linearity

In solving the differential equation above we assumed that any resulting displacement had to be a sinusoidal motion at a single frequency, if the stimulus force is sinusoidal at the same frequency. This way all time-dependent terms canceled out and we obtained a resulting amplitude and phase as a function of the stimulus frequency.

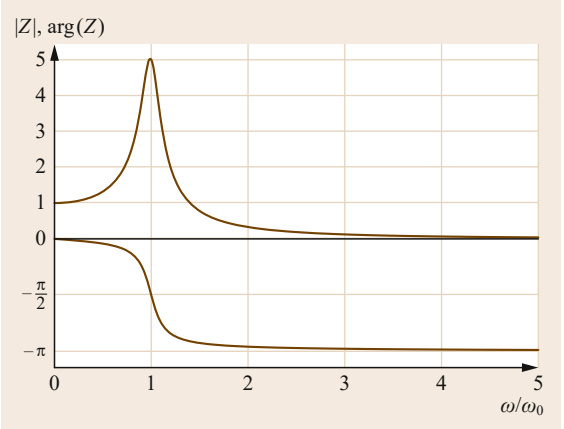


Fig. 2.2 Resonance magnitude and phase

This result seems to be very limited because we know that real systems are not stimulated at a single frequency by a sinusoidal force. The violin bow vibrates a string in a saw-tooth manner, which is far from being sinusoidal. The reed instruments are stimulated by an airflow that is cut off whenever the reed closes the mouthpiece. The same is true for a brass player's lips. In reality there are no sinusoidal stimulus forces! Are our above solutions therefore useless?

Fortunately most real systems are linear, at least, if amplitudes are not too high. This means, that the effect of different stimulus signals can be additively superimposed. As long as any real stimulus signal can be represented by a sum of sinusoidal signals, it is possible to calculate the response to the stimulus components separately and subsequently add them in order to get the system response to a real stimulus.

This is exactly how a frequency domain solution is applied. Any periodic signal F (e.g., sawtooth, rectangular, pulse train or any other repetitive shape) can be represented by a weighted sum of strictly harmonic sine or cosine components (partials) at integer multiples of the fundamental frequency $\omega = 2\pi f$.

This is called Fourier decomposition and can be described according to

$$F(\omega t) = \sum_{n=0}^{\infty} A_n \sin(n\omega t) + B_n \cos(n\omega t) \tag{2.15}$$

or

$$F(\omega t) = \sum_{n=0}^{\infty} C_n \cos(n\omega t + \phi_n) \tag{2.16}$$

which is equivalent if $\tan(\phi_n + \pi/2) = B_n/A_n$ and $C_n = \sqrt{A_n^2 + B_n^2}$.

With Euler's formula and $D_n = C_n e^{i\phi_n}$ we can also write

$$F(\omega t) = \mathrm{Re}\left(\sum_{n=0}^{\infty} D_n e^{in\omega t}\right). \tag{2.17}$$

Knowing the coefficients D_n of the frequency components of any periodic stimulus signal, it is possible to use the frequency domain solution to determine the response to all the partial stimulus components and add them together to obtain the real world system response. The coefficients D_n of the Fourier-series representation of any periodic signal $s(t)$ with a period T can be calculated according to

$$D_n = \frac{1}{T}\int_{t_0}^{t_0+T} s(t) e^{\frac{-i2\pi nt}{T}} dt \tag{2.18}$$

or obtained from a mathematics handbook.

We mentioned before that linearity can be assumed as long as amplitudes are not too high. But what are the limits? Let us look at the single mass–spring system. The spring was characterized by the proportionality between force and length change ΔL of the spring. But this proportionality no longer exists when the spring is significantly deformed. At the extreme when the spring wire is completely straightened, the length will no longer change at all. If we still increase the force the wire will yield and eventually break. The stiffness of the spring obviously depends on how much it has already been elongated or compressed.

Figure 2.3 shows the dependency between length and force of a real spring. It starts to become nonlinear at a certain compression and elongation. Similar curves are valid for elastic fluids or solids. Wood, for instance, is only elastic between certain limits. Beyond these limits it becomes very stiff until it breaks.

Even air loses its ideal elasticity if the pressure becomes so high that the resulting local temperature increase can no longer be neglected. The latter effect contributes to the nonlinear wave steepening in brass wind instruments, which is known to even create shock waves at the fortissimo level [2.3].

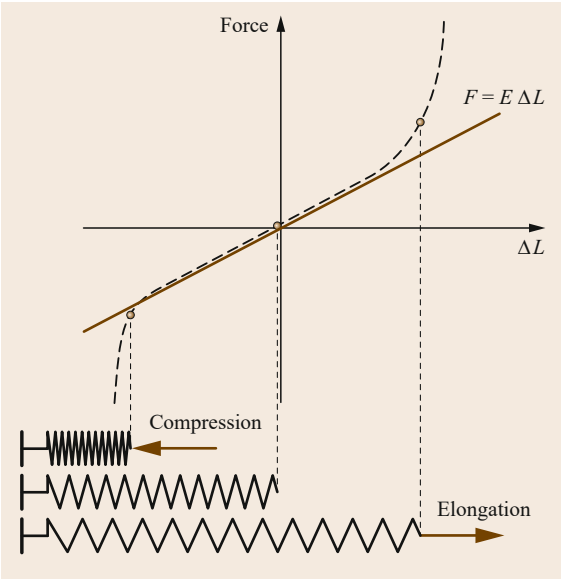


Fig. 2.3 Linearity range of spring

Linearity is also jeopardized when there are collisions with other solid objects, for instance, when amplitudes exceed a certain threshold. This can occur when strings collide with the fingerboard or reeds with the mouthpiece.

Once linearity is lost simple frequency domain solutions can no longer be obtained. Systems will respond with new frequency components that are not part of the stimulus signal, but which typically form higher harmonics of the stimulus partials. This effect is called nonlinear distortion. Subharmonics and other combination tones can also be created, which is called intermodulation distortion.

2.2 Waves

As previously explained, waves are coupled vibrations, distributed and propagating in one, two or three dimensions. If these distributed mass points oscillate in a direction parallel to the direction of wave propagation then this kind of wave is called a longitudinal wave. Waves of this type are often referred to as compressional waves. Sound waves in a medium with low viscosity like air are typical examples for this wave type.

In materials where significant shear forces can occur, for example in solids, local oscillations may also be perpendicular to the direction of wave propagation. Such waves are referred to as transverse waves or shear waves. Waves on strings, membranes or, most typically, on a water surface are examples of such types of waves.

Electromagnetic waves, such as radio waves or light waves, are also transverse waves because oscillating electric and magnetic field vectors are both perpendicular to the overall propagation direction. Nevertheless, there remains another degree of freedom due to the ax-

ial symmetry of a light beam. If there is a single plane containing all oscillating field vectors then the electromagnetic wave is said to be linearly polarized. The acoustical equivalent is a vibrating string with all string segments oscillating in the same direction, for instance vertically to the soundboard, such as the case for a piano string having just been struck by the hammer. During decay this polarization plane can rotate after some time and change from the vertical to the horizontal. This often happens in pianos or guitars and causes the very characteristic shape of their decay curve.

For the sake of completeness it should be mentioned that light or other electromagnetic waves can also exhibit circular polarization. This means that the field vectors no longer oscillate but rotate around the axis at a certain angular velocity depending on the wave length. Such a wave phenomenon can also be observed on vibrating strings, however, it does not seem to have much significance in musical acoustics. The more general elliptic polarization is somewhere between linear and circular polarization.

In addition to longitudinal and transverse waves, as described above, there is also another type of wave. In these waves, the microscopic oscillation has the characteristic of a rotation around an axis. If this rotary oscillation propagates along this axis, this kind of wave is usually referred to as torsional wave. Such kind of waves can be observed on bowed violin strings although their acoustic effects are clearly of higher order compared to dominant transverse waves.

Huygens [2.4] formulated his well-known principle that every point of such a continuum reached by a wave front is a source of a new bidirectional (1-D), circular (2-D) or spherical (3-D) wave front. The overall wave pattern can always be obtained as the superposition of all such secondary wavelets.

Superposition requires linearity as already discussed in the previous section. If sound waves – just like waves in general – have very small amplitudes, which means that the periodic fluctuations of the related physical quantities are small compared to their quiescent value, then oscillations can be treated as minor perturbations of the equilibrium state of a medium. On that condition second and higher-order terms can be neglected and a purely linear system can be assumed.

In such a system waves caused by different stimuli can be superimposed additively and complex wave shapes can be analyzed by decomposing them into single simpler shapes, called modes, attributed to a single stimulus. Linear systems also exhibit the property of scalability, that means, if a stimulus is scaled in amplitude then the response will be scaled by the same factor.

The superposition principle can therefore be applied to wave shapes in space and not only to stimulus and response signals in the time domain.

If we take a spatial snapshot of a steady state wave excited by a nonsinusoidal stimulus at any given time point t, the superposition principle allows us to decompose the observed deformation profile into a weighted sum of profiles, which would occur if the system had been separately stimulated by the sinusoidal frequency components of the composite stimulus.

Even transient wave phenomena can be additively superimposed if the system is linear. A stone thrown into the water will trigger circular wave fronts propagating radially from the center. A second stone thrown nearby will create a completely independent wave pattern. Here where the two patterns overlap, the total displacement of the water surface will simply be the sum of the wave displacements caused by the two independent events.

2.2.1 Reflection

Usually waves do not propagate in an infinite medium. Sooner or later they will reach a boundary where a certain condition of the wave is enforced. For instance, any transverse displacement of a vibrating string will be forced to 0 if the end of the string is rigidly mounted. In the same way longitudinal oscillation of air molecules will be forced to 0 at any rigid wall hit by the wave.

On the other hand, transverse water waves can also be observed to double in amplitude at a quay wall, which throws them back or reflects them, to use the terminology of physics. The same will happen to an acoustic sound pressure wave striking against a sufficiently large wall perpendicular to the direction of sound propagation.

This can be qualitatively and quantitatively understood when the superposition principle is considered. What we physically observe is the superposition of a forward traveling wave and a backward traveling wave, which is created at the boundary of the wave medium by reflection.

Obviously there are two kinds of reflection: one that extinguishes the original wave at the boundary and another one that doubles its amplitude. The first one is called *reflection at the closed or fixed end* and the second is called *reflection at the open end*.

The difference between the two types is the sign of the reflected wave. In the first case the reflected wave has the opposite direction and opposite sign from the impinging wave. Both components will therefore compensate each other at the boundary. In the *open end* case the reflected wave has the opposite direction but same

sign. Both components will interfere constructively and therefore double the amplitude at the boundary.

In reality there are many more than just these two types. Everything in between these extremes is possible. The realistic case is that a part of the incoming wave will pass through the boundary while only the remaining part will be reflected.

This so-called scattering at boundaries or discontinuities is commonly described by a complex factor known as the *reflectance*. Its magnitude, the reflection factor, as well as its phase are frequency dependent. The phase can be anything between 0 (reflection with same sign) and π (reflection with opposite sign).

2.2.2 Standing Waves

For establishing a resonance at least two reflections are needed. Primary resonances usually occur between two reflecting boundaries. Waves can recirculate between the boundaries until their energy is dissipated by non-ideal reflection conditions or propagation losses due to inner or outer friction usually subsumed as *damping*.

If the distance between those boundaries is such that a wave after a complete round trip has again the original phase then constructive interference takes place during recirculation and a standing wave of maximum amplitude is observed. The formation of standing waves therefore requires a certain relationship between the round trip distance, the wavelength, and the type of reflection on both sides. A standing wave is characterized by alternating regions of minimum and maximum wave amplitude, so-called nodes and antinodes, which do not move in space.

As wavelength $\lambda = c/f$ is a function of frequency f the condition for the occurrence of standing waves will – for a given wave propagation or phase speed c – only be met at a certain number of frequencies called resonance frequencies. As reflection at a fixed end inverts the wave, it has the same effect as an increase of the round trip distance by half a wavelength.

This leads us to the lowest air resonance f_0 (*fundamental frequency*) of a tube with length L, which is closed on one end and open on the other. A clarinet with all tone holes closed, for instance, corresponds to this case, as it is a roughly cylindrical tube, closed by the reed on the mouthpiece side and open on the other side. Reflection at one of the ends already inverts the wave, so we need to invert just one more time during a round trip in order to satisfy the reinforcement condition required by resonance.

The round trip distance $2L$ must therefore be equal to half a wavelength. The length therefore corresponds to a quarter wavelength or $\lambda = 4L$, from which the fundamental frequency $f_0 = c/4L$ can be calculated. Higher resonances at frequencies f_n will require that $2L = \lambda/2 + n\lambda$ or $f_n = c(2n+1)/4L$. This leads to a sequence of higher resonances with frequencies of 3, 5, 7, ... times the fundamental frequency f_0 and explains why clarinets overblow into the twelfth (three times f_0) and why their sound does not contain many harmonic components at even multiples of f_0.

Flutes are cylindrical tubes with two open ends. The round trip distance $2L$ must therefore correspond to any multiple of a complete wavelength λ, which results in a fundamental frequency $f_0 = c/2L$, one octave higher than for a clarinet of the same length. The frequencies of higher resonances can be calculated from $f_n = (n+1)c/2L = (n+1)f_0$, which leads to a sequence of overtones at 2, 3, 4, ... times the fundamental frequency. This explains why flutes overblow into the octave and why their sound does contain odd and even multiples of the fundamental.

The same equations apply to vibrating strings fixed at both ends, which enforces displacement nodes at the bridge and nut. However, the propagation speed c of a transverse wave on a taut string is much different from the sound speed in air, and it strongly depends on string tension and mass. Higher resonances of vibrating strings are called *flageolet notes* and they will also sound at 2, 3, 4, ... times the fundamental frequency.

The assumption that strings are fixed at both ends or that sound waves are perfectly reflected at an open end or at the mouthpiece side of trumpets or clarinets, where those instruments are more or less closed by the player's lips or by the reed, is not at all justified in real instruments. Anyhow, it is good to understand the idealized case before starting to analyze more realistic conditions.

Resonances in two-dimensional objects like circular membranes (timpani, banjo resonator) are important but nevertheless much more difficult to model analytically. In order to analyze the quality of real sound boards and complete resonators of stringed instruments or even concert halls, numerical methods and adequate computational capacity are required.

2.2.3 Linear Regime

Whether the linearity assumption is valid for any real world system can be examined by stimulating that system with a sinusoidal signal at a certain frequency. If frequencies different from the stimulus frequency can be observed in steady state – typically higher harmonics generated by a kind of nonlinearity like collision or saturation effects – then the system is not linear.

In the nonlinear case system responses will also no longer be perfectly proportional to the stimulus amplitude. In acoustics and structural mechanics linearity

will usually be observed for small amplitude stimulus signals and will gradually degrade when stimulus amplitudes tend to exceed a certain *linear* range often called the *linear regime* of a system.

However, it has to be noted that the linear regime has very blurred boundaries, and even at small amplitudes nonlinear effects like wave steepening can accumulate over greater distances and eventually become notable [2.5]. Wave steepening occurs when the wave propagation speed depends on the wave variable. It is responsible for the steepening and breaking of water waves near the shore.

2.3 Wave Equations 1-D

Wave propagation is omnipresent in music acoustics. In this section it will be shown how different mechanisms related to sound generation and propagation in musical instruments can be described analytically.

The derivation of practically useful analytic expressions could (and should) start from the fundamental laws of fluid dynamics or structural mechanics. The many terms of these conservation laws need to be carefully inspected in order to find out which ones may safely be neglected under certain conditions and which ones need to be taken into account. The symmetry of the problem and its degrees of freedom have to be considered and an appropriate coordinate system has to be chosen. With this approach a system of much simpler equations can usually be obtained which sometimes, but not always, has a chance of being solved analytically.

However, this systematic approach is not followed here because it requires much prior knowledge in physics and mathematics and it does not necessarily lead to an intuitive understanding. Instead, it will be demonstrated here, how the same results can be obtained by studying simple one-dimensional problems and applying obvious fundamental laws in a heuristic way.

It will be shown that the analytic treatment of vibrating strings, oscillating air columns, longitudinal waves in solids and even torsional waves will lead to structurally very similar expressions. This class of differential equations are called *wave equations*.

The analytic and numerical solution of the obtained wave equations and its practical implications will be dealt with in a following chapter of this book.

2.3.1 Transverse Waves on Strings

Chordophones are characterized by a sound generation mechanism based on vibrating strings. An analytic description of vibrating strings does not primarily depend on how the string is excited. Whether the string is plucked, struck by a hammer or bowed, string resonances and possible vibration modes can be calculated in the same way.

There are differences between thin guitar and heavy piano strings due to different stiffness and mass distribution but for now we are going to ignore such effects to keep things as simple as possible. We even ignore the primary purpose of vibrating strings in musical instruments, which is to transmit sound energy through the bridge to the resonator in order to radiate sound.

We will first study an infinitely thin, perfectly flexible string with no inner or outer friction losses and with a constant mass distribution along its length. The string displacement is assumed to be very small compared to the string length so that the tension does not change when the string is displaced.

Figure 2.4 shows a string taut between two fixed points on the x-axis. The highlighted segment represents an infinitesimal small segment of length $\mathrm{d}s \approx \mathrm{d}x$ located at segment position x. The big arrows labeled with T are the stretching and restoring forces, which are caused by the tension of the string. The local string displacement at the left side of the segment is y and at the right side $y + \mathrm{d}y$.

The segment slope gradually changes from $y' = \partial y/\partial x$ at the left side to $y' + \partial^2 y/\partial x^2 \mathrm{d}x$ at the right side of the segment. The curvature $\partial^2 y/\partial x^2$ multiplied with the segment length $\mathrm{d}x$ gives the change in slope across the segment.

The slope is important because it determines the y-axis component of the tension force T on both sides of the segment. The resulting force in the y-direction is the sum of the y-axis components of the tension forces T left and right of the segment, and this is T times the difference in slope due to the opposite direction of T at both sides of the segment.

The *equation of motion* (2.19) shows how the resulting driving force on the right side counteracts the

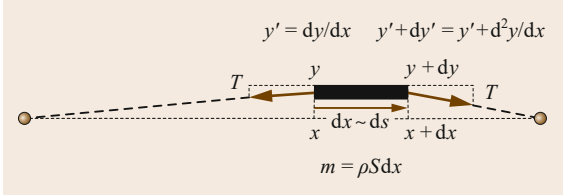


Fig. 2.4 Transverse wave on a string

inertial force on the left side,

$$m\frac{\partial^2 y}{\partial t^2} = T\frac{\partial^2 y}{\partial x^2}\mathrm{d}x\,. \tag{2.19}$$

The assumption that tension T does not depend on the displacement y is only justified for very small displacements. It is known from experience that a guitar string after having been plucked strongly will drop in pitch as the amplitude decays. The initially higher pitch is because of the increased tension that comes with large displacements.

With segment mass $m = \rho S\mathrm{d}s \approx \rho S\mathrm{d}x$, ρ being the mass density and S the cross-sectional area, we obtain (2.20), which is the wave equation for the ideal string. $T/\rho S = c^2$ determines the phase speed of a transverse wave along the string,

$$\rho S\frac{\partial^2 y}{\partial t^2} = T\frac{\partial^2 y}{\partial x^2}\,. \tag{2.20}$$

2.3.2 Plane Waves in Air

Sound waves in air with plane or at least almost plane wave fronts can be observed at large distances from a point source or inside a straight duct with constant cross-sectional area. The latter case corresponds to the interior of cylindrical sections of brass wind instruments and to certain cylindrical wind instruments, like flutes and clarinets.

The importance of the plane wave model is due to the fact that wind instruments with arbitrary bore profiles can be treated as a sequence of short cylindrical slices with discontinuities in the cross-sectional areas in between them. This is called *stepped tube* discretization and allows to model wave propagation in any real wind instrument in a relatively realistic way.

It turns out that the influence of bends on acoustical characteristics is small and even the actual cross-sectional shape does not largely matter, so it is usually sufficient to only deal with the cross-sectional area. It may seem surprising, but even the vocal tract of singers, the cavity between vocal folds and lip orifice and respectively nostrils, is usually modeled by an approximately 17 cm long straight stepped tube consisting of cylindrical slices.

Figure 2.5 shows an infinitesimal small segment of an arbitrary acoustical wave guide. As we are dealing with a longitudinal wave now, the particle displacement ξ and the fluid velocity v are parallel to the x-axis, the direction of wave propagation. The length of the segment is $\mathrm{d}x$ and the corresponding air mass $m = \rho S\mathrm{d}x$, ρ being the air density, S the cross-sectional area, and $\mathrm{d}x$ the infinitesimal length of the segment.

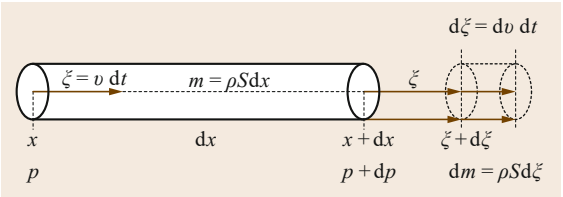


Fig. 2.5 Longitudinal plane waves in air

This air mass is accelerated by the pressure force $-S\mathrm{d}p$, which is the difference between the pressure forces at the left and right side of the segment. The negative sign comes from the fact that a positive $\mathrm{d}p$ (right side pressure is higher than left side pressure) will accelerate the air towards the left side. The equation of motion is therefore

$$m\frac{\partial^2 \xi}{\partial t^2} = -S\mathrm{d}p\,. \tag{2.21}$$

With $v = \partial\xi/\partial t$ we can write

$$\rho\frac{\partial v}{\partial t} = -\frac{\partial p}{\partial x}\,. \tag{2.22}$$

Now we need to focus on the pressure p. The fundamental law of fluid dynamics $pV^\kappa = \text{const.}$ (with volume V and $\kappa = C_\mathrm{p}/C_\mathrm{V}$, the ratio of heat capacitances C at constant pressure and constant volume), which is valid for adiabatic processes and ideal gases, gives us a relation between pressure p and mass density $\rho = m/V$,

$$pV^\kappa = p_0 V_0^\kappa\,, \tag{2.23}$$

$$\frac{p}{\rho^\kappa} = \frac{(p_0+\hat{p})}{(\rho_0+\hat{\rho})^\kappa} = \frac{p_0}{\rho_0^\kappa}\,. \tag{2.24}$$

With p_0, V_0 and ρ_0 we designate the atmospheric equilibrium conditions while $\hat{p}$ and $\hat{\rho}$ represent small time varying quantities, treated as minor perturbations of the equilibrium state hence

$$\frac{p_0+\hat{p}}{p_0} = \left(\frac{\rho_0+\hat{\rho}}{\rho_0}\right)^\kappa\,. \tag{2.25}$$

A first-order approximation $\hat{p} = \hat{\rho}\partial\hat{p}/\partial\hat{\rho}$ for $\hat{p} << p_0$ and $\hat{\rho} << \rho_0$ yields the proportionality between p and ρ

$$\hat{\rho} = \hat{p}\frac{\rho_0}{\kappa p_0}\,. \tag{2.26}$$

We need this proportionality because the continuity equation describing the conservation of mass relates a density change with a velocity gradient $\partial v/\partial x$ according to

$$\frac{\partial\rho}{\partial t} = -\rho_0\frac{\partial v}{\partial x}\,. \tag{2.27}$$

Equation (2.27) states that any increase or decrease in density $\rho = mV^{-1}$ per second must be due to a velocity gradient, which decreases or increases the fluid volume $V = S\mathrm{d}x$ over time by $\mathrm{d}V = S\Delta v\mathrm{d}t$ because its right boundary travels faster or slower than its left boundary. The mass in that fluid element must be constant (for continuity). With $\Delta v = \partial v/\partial x\,\mathrm{d}x$ we obtain the volume change rate $\mathrm{d}V/\mathrm{d}t = S\partial v/\partial x\,\mathrm{d}x$. Using $\mathrm{d}\rho/\mathrm{d}t = (\mathrm{d}\rho/\mathrm{d}V)(\mathrm{d}V/\mathrm{d}t)$ and $\mathrm{d}\rho/\mathrm{d}V = -mV^{-2}$, (2.27) can be derived without difficulty.

Differentiating (2.22) with respect to x and differentiating (2.27) with respect to t will allow to eliminate the v related terms. Substituting ρ according to (2.26) leads to the wave equation for the sound pressure p

$$\frac{\partial^2 p}{\partial t^2} = \frac{\kappa p_0}{\rho_0}\frac{\partial^2 p}{\partial x^2}\,, \tag{2.28}$$

where $\kappa p_0/\rho_0 = c^2$ determines the phase speed of a longitudinal plane wave.

Differentiating (2.22) with respect to t and differentiating (2.27) with respect to x and substituting ρ according to (2.26) will allow to eliminate the p related terms. This leads to the wave equation for the sound velocity v

$$\frac{\partial^2 v}{\partial t^2} = \frac{\kappa p_0}{\rho_0}\frac{\partial^2 v}{\partial x^2}\,. \tag{2.29}$$

The phase speed of the sound velocity wave is, as expected, the same as for the sound pressure wave.

Equation (2.22) gives us a relation between the pressure p and fluid velocity v. If we assume the pressure $p(x,t) = A\,\mathrm{e}^{-\mathrm{i}kx}\mathrm{e}^{\mathrm{i}\omega t}$ to be a forward traveling harmonic wave function then the partial derivative with respect to x is $p_\mathrm{x} = -\mathrm{i}kp$. If the pressure is harmonic then the velocity must be harmonic also. From $v(x,t) = C\,\mathrm{e}^{-\mathrm{i}kx}\mathrm{e}^{\mathrm{i}\omega t}$ follows $v_\mathrm{t} = \mathrm{i}\omega v$. Substituting both p_x and v_t in (2.22) yields $\rho\omega v = kp$. With $k = \omega/c$ we obtain

$$\rho c = \frac{p}{v}\,. \tag{2.30}$$

A backward traveling pressure wave according to $p(x,t) = B\,\mathrm{e}^{\mathrm{i}kx}\mathrm{e}^{\mathrm{i}\omega t}$ would yield $p_\mathrm{x} = \mathrm{i}kp$, eventually resulting in $\rho\omega v = -kp$. This reminds us to draw opposite velocity arrows for the forward and backward traveling waves in order to keep the characteristic impedance of the medium $Z_0 = p/v = \rho c$ positive in both cases.

2.3.3 Longitudinal Waves in Solids

Longitudinal waves in piano strings are widely known to be an issue as they are clearly audible and generally not related to the normal harmonic spectrum of the string. Piano makers had to spend a lot of effort to control the pitch of longitudinal string modes in a way to let them match one of the harmonic overtones.

Longitudinal waves are stimulated by the hammer because any local transverse displacement is inevitably connected to a local length strain, which propagates as a longitudinal wave. The nonlinear coupling of transverse and longitudinal waves generates a wealth of so-called *phantom partials* contributing to the specific piano timbre especially in the lower registers [2.6, 7].

The difference between longitudinal waves in fluids and solids is small, so Fig. 2.5 may as well serve as a template for the treatment of longitudinal waves in solids, like strings and thin bars. Equation (2.21), the equation of motion, is again the starting point for the derivation. Instead of the air pressure p and the air pressure difference $\mathrm{d}p$ we have the stress due to the relative length change given by $f/S = E\partial\xi/\partial x$, f being the force and E the Young's modulus or a spring constant.

A difference between the right and left-sided stress $\mathrm{d}f/S = E\partial^2\xi/\partial x^2\,\mathrm{d}x$ will accelerate the mass $m = \rho S\mathrm{d}x$ like $\mathrm{d}p$ does in fluids.

This leads again to a wave equation for the displacement ξ according to

$$\frac{\partial^2 \xi}{\partial t^2} = \frac{E}{\rho}\frac{\partial^2 \xi}{\partial x^2} \tag{2.31}$$

with a phase propagation speed $c = \sqrt{E/\rho}$.

2.3.4 Torsional Waves in Bars

Torsional vibrations are not very common in musical acoustics but there are certain cases where they have some significance. Bowed strings, for instance, are stimulated into torsional vibrations especially during transients like onset and bow change. This has considerable influence on minimum bow force and other parameters of the bow–string interaction [2.8].

It is known that free reeds and even beating reeds, like clarinet or saxophone reeds, do have a certain inclination to exhibit torsional vibration modes. However, players try to avoid those modes during normal playing because they do not sound nice. Investigations imply that resistance against torsional vibrations is related to subjective quality measures [2.9].

Torsional waves in bars or rods can be treated very similarly to compression waves. Mass m translates to a polar moment of inertia $\rho I_\mathrm{p}\mathrm{d}x$, acceleration translates into angular acceleration $\partial^2\theta/\partial t^2$, and the driving force is the net torque $\partial T/\partial x\,\mathrm{d}x$ acting on an infinitesimal small segment of length $\mathrm{d}x$ (refer to Fig. 2.6).

The torque T is related to the twist angle θ by $T = J\partial\theta/\partial x$, with $J = GK_T$ being the torsional rigid-

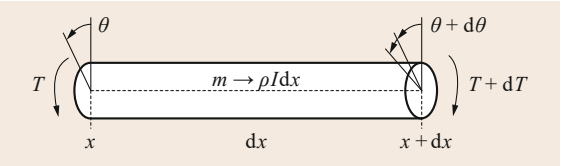


Fig. 2.6 Torsional waves in a rod

ity, a product of the shear modulus G and a geometry factor K_T.

The wave equation for the angular displacement, the twist angle θ of torsional waves becomes

$$\frac{\partial^2\theta}{\partial t^2} = \frac{G K_T}{\rho I_\mathrm{p}} \frac{\partial^2\theta}{\partial x^2} \,. \tag{2.32}$$

The phase propagation speed is $c = \sqrt{G K_T/\rho I_\mathrm{p}}$.

The shear modulus G is often related to the Young's modulus E and Poisson's ratio ν according to $G = E/(2+2\nu)$. K_T/I_p is related to the geometric shape of the cross-section. According to [2.10] it is approximately 1 for a circular cross-section or a circular tube, 0.92 for a square, 0.74 for a rectangle with width $w = 2h$ and $2h/w$ for plates with a width w more than 6 times the height h.

Moment of Inertia

The polar moment of inertia of any plane shape with respect to a certain point is its resistance against being twisted around an axis through that point perpendicular to the plane. Each area element $\mathrm{d}A$ inside the boundaries of that shape contributes a certain inertial force $\mathrm{d}F$ causing an inertial moment $\mathrm{d}M = r\mathrm{d}F$, r being its distance from the rotation center.

The inertial force $\mathrm{d}F$ depends on the mass of that element and its acceleration $\mathrm{d}v/\mathrm{d}t$. A certain angular acceleration $\mathrm{d}^2\theta/\mathrm{d}t^2$ will accelerate a point mass in a tangential direction, but the physical acceleration will not be equal to the angular acceleration because the tangential displacement is proportional to the radius r. Therefore we obtain the relation $\mathrm{d}v/\mathrm{d}t = r\mathrm{d}^2\theta/\mathrm{d}t^2$. Physical and angular acceleration will only be identical for a radius of 1 m because there the arc length in meters and the angle θ in radians are the same.

If the mass distribution over area, the area density ρ_A, is constant it makes sense to just consider the shape. The polar moment of inertia of a massless plane shape is therefore the area integral, a two-dimensional sum, over the contributions of all area elements inside the boundary

$$I_\mathrm{p} = \frac{1}{m}\int \mathrm{d}M = \int_\mathrm{A} r^2 \mathrm{d}A \,. \tag{2.33}$$

2.3.5 Transverse Waves on Bars

Figure 2.7 shows a small section of a bar, thick enough to provide a shape restoring force due to its inner bending stiffness. This makes the main difference from a perfectly flexible string, which has no shape restoring force other than the tension forces acting at the two ends.

The case of a string without stiffness is an ideal situation just like a bar without any external forces. At least gravity and bearing or clamping forces will act on any real bar. Although it is enlightening to deal with ideal cases, practically we will always have to work with combinations and with violations of simplifying assumptions.

Let us think of the bar as if it were a bundle of thin independent wires with quadratic cross-sections and 0 stiffness glued together at least at both ends. If we try to bend the bar then the outer wires will be stretched while the inner wires will be compressed. It is obvious that there is also a central layer of wires, which will be bent but do not change their lengths.

Regardless of the actual cross-sectional shape of the bar, the length change and therefore the strain of the wires only depends on their distance r from the central axis. So we can also think of the bar as a composite with many thin layers without stiffness characterized by their distance from a central layer containing the axes of all cross-sections.

By bending, each slice will be deformed in a way that the left and right boundaries are no longer parallel. This angle $\mathrm{d}\theta$ is related to the curvature of the center line. The slope of the center line is $\partial y/\partial x$ and if there is a curvature $\partial^2 y/\partial x^2$ then the difference in slope between the left and right boundary is $\partial^2 y/\partial x^2 \mathrm{d}x$. If the angle is small then $\tan\theta \approx \theta$, so $\tan(\theta + \mathrm{d}\theta) - \tan\theta \approx \mathrm{d}\theta$.

If we multiply $\mathrm{d}\theta$ with r we get an arc length that is the length change of the layer at distance r from the axis. The relative length change or strain $r\mathrm{d}\theta/\mathrm{d}x$ of a single filament at distance r to the center layer is proportional to the related force. Note that the deformation is small so we can approximate the actual arc length

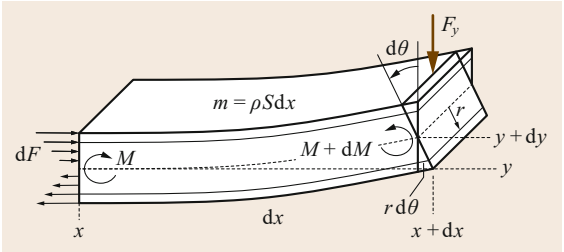


Fig. 2.7 Bending of a bar

of the central layer ds by dx. With Young's modulus E as the proportionality factor the force contribution is then $\mathrm{d}F = E r \mathrm{d}\theta/\mathrm{d}x$. The contribution to the total bending moment $\mathrm{d}M = r\mathrm{d}F = Er^2\mathrm{d}\theta/\mathrm{d}x$.

Now we need to sum up over all contributions to obtain the total moment acting on the cross-section S. Compression forces above and expansion forces below the bending axis have opposite orientation but they act in the same sense of rotation, so if r is always a positive distance, length change, strain, force and moment will always be positive and sum up correctly. The total moment at a cross-section S is therefore $M = \int_\mathrm{A} \mathrm{d}M\mathrm{d}A = E\mathrm{d}\theta/\mathrm{d}x \int_\mathrm{A} r^2\mathrm{d}A$.

Following the derivation of (2.33), it should be straight forward to derive an axial moment of inertia I_a for the cross-sectional area with respect to a bending axis

$$I_\mathrm{a} = \frac{1}{m}\int \mathrm{d}M = \int_\mathrm{A} r^2 \mathrm{d}A \,. \tag{2.34}$$

Again the variable r designates the radius of rotation, but this time it is the shortest distance to the axis because the axis is in the plane of the rotating shape.

For a rectangular bar with thickness h and width w we obtain $I_\mathrm{a} = wh^3/12$. For a circular rod with radius r we obtain $I_\mathrm{a} = r^4\pi/4$ and for a tube with inner radius r_i and outer radius r_o we obtain $I_\mathrm{a} = (r_\mathrm{i}^2 + r_\mathrm{o}^2)(r_\mathrm{o}^2 - r_\mathrm{i}^2)\pi/4$.

With this axial moment of inertia I_a and the substitution $\mathrm{d}\theta \to \partial^2 y/\partial x^2\,\mathrm{d}x$ we obtain $M = I_\mathrm{a} E \partial^2 y/\partial x^2$.

This moment depends on x and may therefore not be the same at the left and right side of a slice. Any ΔM must be compensated by an external moment $\mathrm{d}xF_y = \partial M/\partial x\,\mathrm{d}x$, which is supplied by the inertia of the segment. The vertical force component that is responsible for any vertical displacement of the bar has the same magnitude but opposite sign. F_y depends again on x so we can write

$$\frac{\partial F}{\partial x}\mathrm{d}x = \frac{\partial M}{\partial x} = I_\mathrm{a} E \frac{\partial^3 y}{\partial x^3} \,. \tag{2.35}$$

With the equation of motion

$$m\frac{\partial^2 y}{\partial t^2} = -\frac{\partial F}{\partial x}\mathrm{d}x \,, \tag{2.36}$$

and $m = \rho S \mathrm{d}x$ we obtain

$$\frac{\partial^2 y}{\partial t^2} = -\frac{E}{\rho}\frac{I_\mathrm{a}}{S}\frac{\partial^4 y}{\partial x^4} \,. \tag{2.37}$$

This is not a second-order wave equation like those that we obtained in the other cases. As will be shown later, the propagation speed is no longer constant but frequency dependent. This is a significant complication as it leads to dispersion. Composite waves become distorted as they travel along the medium. Unfortunately it is often required to include such a nonzero stiffness term in wave equations of real strings.

2.4 Solution for 1-D-Waves

Solving a differential equation is generally not a straight forward approach. For many types of differential equations templates for possible solutions are known. Such a template is a quite general expression with a certain number of unknown constants or unknown functions. This means that only the shape of a solution is known, but the unknowns are undetermined. It is easy to verify that these templates provide solutions to differential equations, if those stay correct after the substitution.

The unknown constants and functions must be determined by using whatever is known about the actual problem. Waves propagate equally in all one-, two- and three-dimensional spaces. Where and how the boundary conditions are embedded make the difference when looking at different cases.

A vibrating string has only a few boundary conditions, namely where and how the string is terminated. It makes a difference whether the ends are clamped or movable and how much tension is applied to the string. Other systems might have more complicated boundary conditions. In concert halls, for instance, it is not easy at all to define the boundary of the acoustic space.

It will be shown here how the wave equations can be solved analytically for the presented one-dimensional cases.

2.4.1 General Solution

The most general solution of the one-dimensional wave equation

$$\frac{\partial^2 F}{\partial t^2} = c^2 \frac{\partial^2 F}{\partial x^2} \,, \tag{2.38}$$

has the shape

$$F(x,t) = f(u) + g(v) \,, \tag{2.39}$$

with $u = ct - x$ and $v = ct + x$. This is the superposition of a forward traveling wave of any shape f and a back-

ward traveling wave of any shape g, both traveling with speed c in opposite directions.

Before we proceed we will introduce a very convenient way to write partial derivatives saving space and making them easier to read, for instance

$$\begin{aligned}
\frac{\partial F}{\partial t} &\to F_t \,, \\
\frac{\partial f}{\partial u} &\to f_u \,, \\
\frac{\partial g}{\partial v} &\to g_v \,, \\
\frac{\partial^2 F}{\partial t^2} &\to F_{tt} \,, \\
\frac{\partial^2 F}{\partial u, \partial v} &\to F_{uv} \\
&\dots
\end{aligned} \tag{2.40}$$

Using this notation it is straight forward to prove the correctness of the ansatz above

$$\begin{aligned}
F_t &= f_u u_t + g_v v_t = c(f_u + g_v) \\
F_{tt} &= f_{uu} u_{tt} + g_{vv} v_{tt} = c^2 (f_{uu} + g_{vv}) \\
F_x &= f_u u_x + g_v v_x = f_u(-1) + g_v(1) \\
F_{xx} &= f_{uu} u_{xx} + g_{vv} v_{xx} = f_{uu} + g_{vv} \\
F_{tt} &= c^2 F_{xx} \quad \text{(q.e.d.)} \,.
\end{aligned}$$

At a boundary the function F is not free to take on any value. Instead, the sum of the forward traveling wave f and backward traveling wave g must satisfy the physical constraints there. For a string with length L and fixed ends at $x = 0$ and $x = L$ this means that the reflected wave g must have the same magnitude and the opposite sign of the forward wave f at both ends of the string to satisfy the boundary condition $F(0, t) = F(L, t) = 0$.

In order to make things as simple as possible, we let $f = A\cos(k(ct - x))$ and $g = B\cos(k(ct - x))$ be harmonic functions. This is not a painful restriction because the superposition principle allows us to analyze any periodic function of arbitrary shape as a weighted sum of harmonic components. We define the wave number $k = \omega/c = 2\pi/\lambda$ as the number of wavelengths λ per $2\pi m$. This is analogous to the angular frequency $\omega = 2\pi/T$, which is the number of periods T per 2π s.

As already discussed in Sect. 2.1.1 we will interpret harmonic functions as real parts of complex exponentials and write

$$F(x, t) = \mathrm{Re}\left(A\,\mathrm{e}^{\mathrm{i}(\omega t - kx)} + B\,\mathrm{e}^{\mathrm{i}(\omega t + kx)}\right) . \tag{2.41}$$

Transforming the sum in the exponent into a product of exponentials allows us to factor out the time-dependent terms

$$F(x, t) = \mathrm{Re}\left(\left(A\,\mathrm{e}^{-\mathrm{i}kx} + B\,\mathrm{e}^{\mathrm{i}kx}\right)\mathrm{e}^{\mathrm{i}\omega t}\right) . \tag{2.42}$$

A and B are referred to as the complex amplitudes of the forward and backward traveling waves, respectively. They have to be determined using the available information about the boundary conditions.

2.4.2 Propagation

To be more specific, let us now return to the case of acoustic plane wave propagation in cylindrical ducts. Other one-dimensional wave equations derived above are treated likewise.

Acoustic plane waves are usually characterized by the sound pressure p, which varies over time and along the length of the tube. It does not vary across planes perpendicular to the direction of propagation. The so-called plane wave fronts allow us to treat the tube as a one-dimensional problem.

The same is true for the acoustic flow u, which is defined as the fluid velocity v times the cross-sectional area S. It is also called acoustic volume flow. It has the unit m^3/s and it is related to the oscillating fluid flow. It must be continuous over cross-sectional area changes while the fluid velocity will jump at discontinuities of the tube diameter.

Pressure p and acoustic volume flow u are related according to (2.30) and both satisfy wave equations according to (2.38) with $F \to p$ and $F \to u$. Valid solutions for the pressure p and for the flow u can be written as

$$\begin{aligned}
p(x, t) &= \left(A\,\mathrm{e}^{-\mathrm{i}kx} + B\,\mathrm{e}^{\mathrm{i}kx}\right)\mathrm{e}^{\mathrm{i}\omega t} \,, \\
u(x, t) &= \frac{S}{\rho c}\left(A\,\mathrm{e}^{-\mathrm{i}kx} - B\,\mathrm{e}^{\mathrm{i}kx}\right)\mathrm{e}^{\mathrm{i}\omega t} \,.
\end{aligned} \tag{2.43}$$

The negative sign of the backward traveling term for $u(x, t)$ is due to the opposite velocity arrow for forward and backward traveling waves. This has already been discussed for (2.30).

If we adopt this solution to known boundary conditions we can learn more. A cylindrical tube of length L, which is open at both ends (like a flute) enforces minimum pressure at both ends. Likewise, a string with length L that is fixed on both ends must have 0 displacement there: $p(0) = 0$ and $p(L) = 0$ and $y(0) = 0$ and $y(L) = 0$, respectively. This condition holds for any point in time.

Thus we obtain

$$\begin{aligned}
0 &= A\,\mathrm{e}^{-\mathrm{i}k0} + B\,\mathrm{e}^{\mathrm{i}k0} \,, \\
0 &= A\,\mathrm{e}^{-\mathrm{i}kL} + B\,\mathrm{e}^{\mathrm{i}kL} \,,
\end{aligned} \tag{2.44}$$

which leads to $A = -B$ and $e^{-ikL} = e^{ikL}$. The latter condition equates a clockwise and an anti-clockwise rotating arrow. It can only be met for $kL = n\pi$. With $k = \omega/c = 2\pi f/c$ we obtain so-called normal modes at angular frequencies $f_n = nc/2L$. On stringed instruments those are called *flageolet notes* and any linear combination of them can occur in steady state.

To derive the acoustical properties p and u at any point x of the duct assuming they are known at another point x_1, a distance d away, requires substitution of $x = x_1 - d$ into (2.43). With $Z_c = \rho c/S$ for the characteristic impedance we obtain

$$p = \cos(kd)p_1 + i\sin(kd)Z_c u_1 \,, \tag{2.45}$$

and

$$u = i\sin(kd){Z_c}^{-1}p_1 + \cos(kd)u_1 \,. \tag{2.46}$$

We can define the acoustical impedance Z and again derive its projection along the duct

$$Z = \frac{p}{u} = \frac{\cos(kd)Z_1 + i\sin(kd)Z_c}{i\sin(kd)Z_c^{-1}Z_1 + \cos(kd)} \,. \tag{2.47}$$

These results are extremely useful as p, u and Z at the end of a wind instrument are at least approximately known. If it is closed, which is the case in certain stopped pipes, u is nearly 0 and Z very large. If it is more or less open then p is very small and Z approximates 0. It is possible to determine this impedance at the end of any arbitrary duct either by analytical means or by measurements.

Once this so-called termination impedance Z_{trm} is known, it can be back propagated to the mouthpiece where it becomes the load impedance of the sound generator. The sound generator is a pressure sensitive valve either represented by the player's vibrating lips or by an oscillating reed, which modulate the air flow periodically. In flutes or certain organ pipes it is an unstable air jet (air reed), which oscillates around the edge of the labium due to its interaction with the sound pressure field in its vicinity.

It should be mentioned here that a discontinuity in the tube, meaning a sudden change in cross-sectional area from $S = S_L$, left of the discontinuity, to $S = S_R$, right of it, cannot cause any change in p, u and therefore in Z. This follows from the continuity equations for mass and momentum. What changes is the characteristic impedance Z_c.

This allows one to decompose a complete instrument with varying cross-section into a sequence of short cylindrical slices with different lengths and characteristic impedances Z_c. The known termination impedance can then be propagated back to the mouthpiece.

The impedance there is commonly called input impedance and is numerically evaluated for all frequencies of interest in the whole playing range. The input impedance magnitude spectrum shows all resonances of a wind instrument and is therefore related to its intonation and to other quality measures.

2.4.3 Input Impedance

The input impedance Z_{in} is the complex ratio of the acoustic pressure p_{in} and flow u_{in} at the input plane. Its magnitude $|Z_{in}|$ is the amplitude ratio $|p_{in}|/|u_{in}|$ and its phase angle corresponds to the phase difference between p_{in} and u_{in}. A positive phase of Z_{in} means p_{in} is ahead of u_{in}.

In wind instruments it is usually a time varying flow that excites the air column. Depending on the input impedance of the instrument this flow signal will create a certain sound pressure signal at the entry plane of that resonator tube.

A small input impedance magnitude means that the input pressure building up in response to a flow stimulus is weak. At resonance frequencies there is a high input impedance magnitude. A rather tiny flow will already be sufficient to produce an enormous sound pressure in the mouthpiece.

Why is the input impedance so sensitive to frequency? Let us now start to stimulate a very long cylindrical tube. Initially, there is no backward traveling wave. We stimulate a forward traveling pressure wave by injecting an oscillating acoustic flow. The pressure amplitude in response to the flow stimulus will not be frequency dependent.

According to (2.30) and with $u = vS$ the pressure amplitude p will be $\rho c/S$ multiplied by the flow amplitude u. The pressure wave will propagate at phase speed c to the open end. An open end enforces 0 pressure. This condition requires for the backward traveling wave to be phase shifted by an angle of π at the reflection plane. Eventually the backward traveling wave reaches the front end. The physical pressure is now the sum of the forward traveling wave and the arriving backward traveling wave.

Whether we obtain constructive or destructive interference will depend on the phase relationship between the two traveling waves. Since wavelength is a function of frequency, we will obtain an almost perfect annihilation only for a set of frequencies where twice the tube length is an exact integer multiple of one wavelength $\lambda = c/f$. Note that one should not forget the sign inversion at the open end! If twice the tube length is an integer multiple of one wavelength plus half a wavelength, then we will obtain perfect amplification.

Maxima in the input impedance magnitude spectrum therefore mark frequencies at which perfect amplification takes place during regeneration. Minima occur at frequencies in between where forward and backward waves annihilate each other.

As the tube length matches an odd integer multiple of a quarter wavelength at resonance, 0 transitions of forward and backward traveling waves coincide and do not move at all. The physical pressure profile exhibits places with 0 amplitude, the pressure nodes and places with maximum pressure fluctuations, the pressure antinodes.

As theses places do not move, we call these wave patterns *standing waves*. They exclusively occur at resonance frequencies. Inserting a probe microphone into our tube would reveal nodes and antinodes as stable regions of minimum and maximum sound pressure level.

In mechanical systems such as strings we usually speak of the *driving point impedance*. It is the ratio of a driving force F and a resulting velocity v. Keeping in mind the correspondences between force F and sound pressure p and between velocity v and acoustic flow u, we can translate whatever was said above.

The resonances of the cylindrical air column are directly related to the vibration modes of a taut string – guitarists refer to them as *flageolet tones*. There is one difference when strings are considered. Strings are fixed at both ends, while most wind instruments are played with one closed and one open end. The open end of a trumpet fixes the sound pressure p to 0 but lets the fluid velocity v and respectively the acoustic flow u *open*. A fixed end or *node* for the pressure corresponds to a free end or *antinode* for the flow and vice versa.

Taking this into account we can make a comparison: Trumpets are stimulated by a small flow injected near a flow node and it therefore requires an input impedance maximum to get a strong pressure signal there. Flutes are stimulated by a small sound pressure near an open embouchure hole, a pressure node. They require an impedance minimum to stimulate a significant acoustic flow there.

Mechanical systems such as strings are stimulated by a small force. A driving point impedance minimum will maximize the velocity at the driving point. The smaller the impedance the less force is required to stimulate a certain velocity. Mechanical systems have their resonances at driving point impedance minima also called *mobility* maxima.

Magnitude and phase of the impedance curve at resonance frequencies are connected to many musical features, e.g., intonation, response, efficiency and sound timbre. Measured or calculated input or driving point impedance spectra can therefore be used to assess or predict quality related characteristics of musical instruments.

Bore reconstruction starting from acoustical input impedance measurements is another important application where accurate modeling of input impedance is essential. A practical method based on optimization has been developed by the author [2.11].

2.4.4 Radiation Impedance

As already stated above, prior knowledge of the boundary conditions at the reflecting far end leads to the termination impedance, which is required to determine the input impedance.

We will elaborate on the radiation impedance of an open tube because it is one of the few cases where some analytic expression can be found. It is much more difficult to derive an expression for the bridge impedance of violins or pianos with their irregular three-dimensional resonators. However, this bridge impedance is the termination impedance of the strings and is required to accurately calculate the dynamic behavior of a bowed, plucked or struck string, resulting in properties like sound timbre or sustained tones.

Fortunately it is not very difficult to measure driving point impedances at the ends of the vibrating part of piano strings or at a guitar or violin bridge. Measured data can therefore replace missing analytical expressions.

The open end of a wind instrument is a velocity antinode and in a first-order approximation wave fronts are often assumed to be plane. This allows us to treat the fluid layer at the open mouth as if it was a vibrating piston radiating into the three-dimensional space.

To keep things simple an infinite baffle is often introduced with a circular disk of it being the vibrating piston. This allows one to neglect backward radiation into the half space behind the radiation plane and other related challenges.

An expression for a piston radiator terminated in an infinite baffle was given by *Rayleigh* in [2.12].

Piston Without Baffle

A first-order approximation for a piston radiator without baffle was proposed by *Levine* [2.13] and evaluates numerically to

$$\frac{p_\mathrm{T}(\mathrm{i}\omega)}{u_\mathrm{T}(\mathrm{i}\omega)} = Z_\mathrm{T} = Z_{\mathrm{c,T}}\left(\left(\frac{\omega R_\mathrm{T}}{2c}\right)^2 + \mathrm{i}\left(\frac{0.61\omega R_\mathrm{T}}{c}\right)\right) \tag{2.48}$$

with $R_\mathrm{T} = \sqrt{S_\mathrm{T}/\pi}$, $Z_{\mathrm{c,T}} = \rho c/S_\mathrm{T}$ and S_T being the cross-sectional area of the open end.

Because of the one-dimensional assumptions higher-order vibrational modes are not taken into account here. The validity of the plane wave model is limited by the first cut-off frequency ω_{cutoff}, which is usually most restrictive at the open mouth of the instrument. The cut-off frequency can be calculated from $R_{\text{T}}\omega_{\text{cutoff}}/c \approx 1.84$.

Caussé proposes an approximation for the radiation impedance of a piston radiator without baffle valid up to $\omega < 3.5c/R_{\text{T}}$ in [2.14].

Pulsating Part of Sphere

The approach of *Helie* and *Rodet* [2.15] is based on spherical wave propagation, which matches quite realistically the conditions at the end of the bell of a typical brass wind instrument. The exit wave front is treated as a pulsating portion of a sphere radiating into an anechoic three-dimensional environment with no baffle.

In order to use this result in a plane wave context the surface of the pulsating spherical cap has to be flattened and scaled in order to match the cross-sectional area of the exit wave front of the plane wave model.

Unfortunately, even the numeric evaluation of that model is quite demanding as it is represented by a slowly converging sum of terms involving spherical Hankel functions h_n as well as Legendre polynomials P_n of high order N. Helie and Rodet calculated the sum with $N = 200$ terms. With θ being the flare angle and $\nu = fr/c = f/(\pi f_{\text{cutoff}})$ the normalized frequency, a normalized radiation impedance $Z^*_{\text{rad}} = Z_{\text{rad}}/Z_{\text{c}}$ (Z_{c} being the characteristic specific impedance ρc) is given by

$$Z^*_{\text{rad}} = -\frac{2\text{i}}{1-\cos(\theta)} \sum_{n=0}^{\text{N}} \frac{\gamma(n, 2\pi\nu)\mu(n,\theta)^2}{2n+1} , \tag{2.49}$$

with

$$\mu(n,\theta) = \frac{1}{2}(P_{n-1}(\cos(\theta)) - P_{n+1}(\cos(\theta))) , \tag{2.50}$$

and

$$\gamma(n,\omega) = \frac{\omega h^{(2)}_n(\omega)}{n h^{(2)}_n(\omega) - \omega h^{(2)}_{n+1}(\omega)} . \tag{2.51}$$

The expression for the radiation impedance Z_{rad} for real frequencies f, radius r, speed of sound c, and density ρ becomes

$$Z_{\text{rad}} = -\frac{2\text{i}\rho c}{1-\cos(\theta)} \sum_{n=0}^{\text{N}} \frac{\gamma\left(n, \frac{2\pi fr}{c}\right)\mu(n,\theta)^2}{2n+1} . \tag{2.52}$$

2.4.5 Reflection and Transmission

Any variation of the characteristic impedance Z_{c} has the effect that waves get partially reflected and transmitted at each discontinuity. It might be of interest, how much is reflected and how much is transmitted at a discontinuity given by a ratio $S_{\text{L}}/S_{\text{R}}$ of the area S of the bore cross-section.

By means of (2.43) we can write two equations matching the left and right-sided pressures and volume velocities. For this purpose we substitute $S = S_{\text{L}}$, $A = A_{\text{L}}$, $B = B_{\text{L}}$ on the left sides and $S = S_{\text{R}}$, $A = A_{\text{R}}$, $B = B_{\text{R}}$ on the right sides. Simplified we obtain $A_{\text{L}} + B_{\text{L}} = A_{\text{R}} + B_{\text{R}}$ and $(A_{\text{L}} - B_{\text{L}})S_{\text{L}} = (A_{\text{R}} - B_{\text{R}})S_{\text{R}}$. The reflection coefficient $A_{\text{L}}/B_{\text{L}}$ for a forward traveling wave ($B_{\text{R}} = 0$) can now be calculated to

$$\frac{A_{\text{L}}}{B_{\text{L}}} = \frac{\frac{S_{\text{L}}}{S_{\text{R}}} - 1}{\frac{S_{\text{L}}}{S_{\text{R}}} + 1} . \tag{2.53}$$

In the same way, a transmission coefficient $A_{\text{R}}/A_{\text{L}}$ can be determined as

$$\frac{A_{\text{R}}}{A_{\text{L}}} = \frac{\frac{2S_{\text{L}}}{S_{\text{R}}}}{\frac{S_{\text{L}}}{S_{\text{R}}} + 1} . \tag{2.54}$$

2.4.6 Wall Losses

The solution $A\text{e}^{\text{i}(\omega t - kx)} + B\text{e}^{\text{i}(\omega t + kx)}$ describes lossless propagation. Amplitudes do not decay over time and both forward and backward traveling waves do not diminish as they are traveling.

In reality waves suffer from significant friction losses in a thin shear layer near the wall where a velocity gradient exists due to the finite viscosity of the air. Thermal energy is lost to the environment, which reduces the amplitude of the waves while they are traveling.

There are also losses due to volume viscosity but these are only significant when sound propagates over very large distances. A third source of loss is due to heat transfer between over-pressure and under-pressure zones, but this only concerns very low frequencies and can usually be ignored.

Exponential damping of forward and backward propagating waves can be described by using a complex wave number $k = \alpha - \text{i}\beta$. With this expression the spatial term of the solution $\text{e}^{\pm\text{i}kx} = \text{e}^{\pm\text{i}\alpha x}\text{e}^{\pm\beta x}$ becomes an arrow in the complex plane with a spatially rotating phase angle αx and an amplitude, which changes exponentially with βx. The forward propagating wave $\text{e}^{-\text{i}kx}$ decays exponentially with growing x while the backward propagating wave $\text{e}^{\text{i}kx}$ decays exponentially in the negative x-direction.

Unfortunately, with this wave number k being complex (and not purely imaginary) it is no longer possible to separate real and imaginary parts according to Euler's law. Equations (2.45)–(2.47) have therefore to be replaced. With $\cosh = \mathrm{e}^{x} + \mathrm{e}^{-x}/2$ and $\sinh = \mathrm{e}^{x} - \mathrm{e}^{-x}/2$ and $k \to \Gamma$ we obtain instead

$$p = \cosh(\Gamma d)p_1 + \sinh(\Gamma d)Z_\mathrm{c}u_1 \,, \tag{2.55}$$

$$u = \frac{1}{Z_\mathrm{c}}\sinh(\Gamma d)p_1 + \cosh(\Gamma d)u_1 \,, \tag{2.56}$$

$$Z = \frac{\cosh(\Gamma d)Z_1 + \sinh(\Gamma d)Z_\mathrm{c}}{\frac{\sinh(\Gamma d)Z_1}{Z_\mathrm{c}} + \cosh(\Gamma d)} \,. \tag{2.57}$$

We will now switch to matrix notation. Equations (2.55) and (2.56) can be understood as a vector $(p, u)^\top$ being the product of a 2 by 2 matrix $\mathbf{A}_i$ representing the projection laws with a vector $(p_1, u_1)^\top$ representing the termination conditions.

The projection equations for the i-th section of a certain sequence of cylindrical slices can be written in matrix notation as

$$\begin{pmatrix} p_i(\mathrm{i}\omega) \\ u_i(\mathrm{i}\omega) \end{pmatrix} = \mathbf{A}_i(\mathrm{i}\omega) \begin{pmatrix} p_{i+1}(\mathrm{i}\omega) \\ u_{i+1}(\mathrm{i}\omega) \end{pmatrix} \tag{2.58}$$

with

$$\mathbf{A}_i(\mathrm{i}\omega) = \begin{pmatrix} a_{i,11}(\mathrm{i}\omega) & a_{i,12}(\mathrm{i}\omega) \\ a_{i,21}(\mathrm{i}\omega) & a_{i,22}(\mathrm{i}\omega) \end{pmatrix} \tag{2.59}$$

or according to (2.55) and (2.56) more specifically

$$\mathbf{A}_i(\mathrm{i}\omega) = \begin{pmatrix} \cosh(\Gamma d) & \sinh(\Gamma d)Z_\mathrm{c} \\ \frac{1}{Z_\mathrm{c}}\sinh(\Gamma d) & \cosh(\Gamma d) \end{pmatrix} . \tag{2.60}$$

Multiple projections along several segments can be expressed by the product matrix of all elements. This can easily be cross-checked by eliminating p_{i+1} and u_{i+1} from the system

$$\begin{pmatrix} p_i \\ u_i \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} p_{i+1} \\ u_{i+1} \end{pmatrix} \tag{2.61}$$

$$\begin{pmatrix} p_{i+1} \\ u_{i+1} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} p_{i+2} \\ u_{i+2} \end{pmatrix} . \tag{2.62}$$

The whole instrument can therefore be represented by a single system matrix

$$\mathbf{A}(\mathrm{i}\omega) = \prod_{i=1}^{L} \mathbf{A}_i(\mathrm{i}\omega) \,. \tag{2.63}$$

The model for cylindrical elements has been generalized for conical segments and spherical wave propagation. It was originally published by *Mapes-Riordan* [2.16] but its notation is adapted here. It is based on work published by *Keefe* [2.17–19], *Caussé* et al. [2.14] and *Benade* [2.20].

The projection matrix of a single conical element has been derived by Mapes-Riordan as

$$\begin{aligned} a_{i,11} &= \frac{x_{i+1}}{x_i}\left(C - \frac{D}{\Gamma x_{i+1}}\right) \\ a_{i,12} &= \frac{x_i}{x_{i+1}} Z_\mathrm{c} D \\ a_{i,21} &= \frac{D}{Z_\mathrm{c}}\left(\frac{x_{i+1}}{x_i} - \frac{1}{(\Gamma x_i)^2}\right) + \frac{C}{Z_\mathrm{c}}\frac{\Gamma L}{(\Gamma x_i)^2} \\ a_{i,22} &= \frac{x_i}{x_{i+1}}\left(C + \frac{D}{\Gamma x_i}\right) \end{aligned} \tag{2.64}$$

with

$$\begin{aligned} C &= \cosh(\Gamma L) \,, \\ D &= \sinh(\Gamma L) \,, \\ \Gamma &= k\left(1.045 r_\mathrm{v}^{-1} + \mathrm{i}\left(1 + 1.045 r_\mathrm{v}^{-1}\right)\right) \,, \\ k &= \frac{\omega}{c} \,, \\ r_\mathrm{v} &= \sqrt{\frac{\rho\omega S_\mathrm{m}}{\eta\pi}} \,, \\ Z_\mathrm{c} &= Z_{\mathrm{c,i}}\left(\left(1 + 0.369 r_\mathrm{v}^{-1}\right) - \mathrm{i}0.369 r_\mathrm{v}^{-1}\right) \,, \\ Z_{\mathrm{c,i}} &= \frac{\rho c}{S_\mathrm{i}} \,, \end{aligned}$$

with ρ being the equilibrium gas density, ω the radian frequency, η the shear viscosity coefficient, c the sound velocity, S_m the cross-sectional area at the center and S_i the spherical area at the input end of the conical element, x_i the radius of the input spherical sector, x_{i+1} the radius of the output spherical sector, and L the distance between the two spheres.

The limit for $x_i \to \infty$ and $x_{i+1} \to \infty$ comes very close to the expression already derived for a cylindrical tube.

2.5 Stiffness

Equation (2.37) for the stiff bar is a fourth-order partial differential equation, which does not allow us to construct a general solution presuming constant phase velocity. We can still assume linearity, which means that a harmonic stimulus with a single frequency can only excite a harmonic motion of the same frequency.

With a displacement $y(x,t) = \mathrm{Re}(F(x)\mathrm{e}^{\mathrm{i}\omega t})$ we get $y_{tt} = -\omega^2 y$ and with (2.37) in short form $y_{tt} = -y_{xxxx} EI_\mathrm{a}/\rho S$ using the abbreviation $v^4 = \omega^2 EI_\mathrm{a}/\rho S$ we obtain

$$F_{xxxx} = \frac{\rho}{E}\frac{S}{I_\mathrm{a}}\omega^2 F = \frac{\omega^4}{v^4}F \qquad (2.65)$$

where v corresponds to the phase velocity c which is not constant but proportional to $\sqrt{\omega}$ and to $\sqrt{c_\mathrm{L}} = \sqrt{E/\rho}$, c_L being the phase velocity of a longitudinal wave in that bar.

With a spatial distribution $F(x) = A\,\mathrm{e}^{\gamma x}$ and $F_{xxxx} = \gamma^4 F$ we get

$$\gamma^4 = \frac{\omega^4}{v^4} = k^4 \,. \qquad (2.66)$$

This gives us two solutions $\gamma^2 = \pm k^2$ and from that we get the four solutions $\gamma = \pm k$ and $\gamma = \pm \mathrm{i}k$. In the complete solution for $y(x,t)$ we have to add all possibilities and we get

$$\begin{aligned} F(x) &= A\,\mathrm{e}^{kx} + B\,\mathrm{e}^{-kx} + C\,\mathrm{e}^{\mathrm{i}kx} + D\,\mathrm{e}^{-\mathrm{i}kx} \,, \\ y(x,t) &= \mathrm{Re}\left(F(x)\mathrm{e}^{\mathrm{i}\omega t}\right) \,. \end{aligned} \qquad (2.67)$$

The unspecified four constants have to be determined from the boundary conditions. If the bar is clamped then the displacement and its first derivative there, the slope, is forced to be 0. If it is only supported then the displacement and the second derivative must be 0. A supported end does not fix the slope, and therefore cannot transmit any torque. A free end can move freely and it can have any slope, but there cannot be any torque nor shearing force, so the third and fourth derivative must be 0.

The wave equation for transverse waves in a taut string can also be augmented by the fourth-order stiffness term. Remember that (2.20) is a conservation law. Any acceleration of the mass of a string segment must be due to forces. Initially only the restoring force of the string tension has been taken into account. Now we add the additional restoring force due to stiffness

$$\rho S y_{tt} = T y_{xx} - EI_\mathrm{a} y_{xxxx} \,. \qquad (2.68)$$

According to *Morse* in [2.21] it can be shown that a stiff string has eigenfrequencies that are no longer integer multiples of the fundamental. Frequency ratios of higher partials are stretched and their pitch becomes sharper with decreasing wavelength. *Fletcher* and *Rossing* present Morse's solution for a stiff string clamped at both ends in [2.10] as

$$\omega_n = n\pi\frac{c}{L}\sqrt{1 + n^2\pi^2\alpha^2}(1 + 2\alpha + 4\alpha^2) \qquad (2.69)$$

with $\alpha = \sqrt{\epsilon}/L$ and $\epsilon = EI_\mathrm{a}/T$.

Valette derives the eigenfrequencies of a supported (hinged) stiff string in [2.22]

$$\omega_n = n\pi\frac{c}{L}\sqrt{1 + n^2\pi^2\alpha^2} \,. \qquad (2.70)$$

A good review of such results for several different boundary conditions can be found in [2.23].

References

2.1 L. Euler: *Introductio in Analysin Infinitorum* (M.M.Bousquet, Lausannae 1748)

2.2 C.F. Gauss: Theoria residuorum biquadraticorum, commentatio secunda, Göttingische gelehrte Anzeigen **1**, 169–178 (1831)

2.3 A. Hirschberg, J. Gilbert, R. Msallam, A.P.J. Wijnands: Shock waves in trombones, J. Acoust. Soc. Am. (JASA) **99**(3), 1754–1758 (1996)

2.4 C. Huygens: *Traité de la lumiere* (Pieter van der Aa, Leiden 1690)

2.5 D.G. Crighton: Propagation of finite amplitude waves in fluids. In: *Handbook of Acoustics*, ed. by M.J. Crocker (Wiley, New York 1998) pp. 187–202

2.6 H.A. Conklin Jr.: Generation of partials due to nonlinear mixing in a stringed instrument, J. Acoust. Soc. Am. (JASA) **105**(1), 536–545 (1999)

2.7 B. Bank, L. Sujbert: Generation of longitudinal vibrations in piano strings: From physics to sound synthesis, J. Acoust. Soc. Am. (JASA) **117**(4), 2268–2278 (2005)

2.8 E. Bavu, J. Smith, J. Wolfe: Torsional waves in a bowed string, Acta Acustica united with Acustica **91**(2), 241–246 (2005)

2.9 F. Pinard, B. Laine, H. Vach: Musical quality assessment of clarinet reeds using optical holography, J. Acoust. Soc. Am. (JASA) **113**(3), 1736–1742 (2003)

2.10 N.H. Fletcher, T.D. Rossing: *The Physics of Musical Instruments*, 2nd edn. (Springer, New York 1990)

2.11 W. Kausel: Bore reconstruction of tubular ducts from acoustic input impedance curve, IEEE Trans. Instrum. Meas. **53**(4), 1097–1105 (2004)

2.12 J.W.S. Rayleigh: *The Theory of Sound*, Vol. 2, 2nd edn. (Repr. Dover, New York 1894)

2.13 H. Levine, J. Schwinger: On the radiation of sound from an unflanged circular pipe, Phys. Rev. **73**, 383–406 (1948)

2.14 R. Caussé, J. Kergomard, X. Lurton: Input impedance of brass musical instruments – Comparison between experiment and numerical models, J. Acoust. Soc. Am. (JASA) **75**, 241–254 (1984)

2.15 T. Hélie, X. Rodet: Radiation of a pulsating portion of a sphere: Application to horn radiation, Acustica **89**(4), 565–577 (2003)

2.16 D. Mapes-Riordan: Horn modeling with conical and cylindrical transmission-line elements, J. Audio Eng. Soc. **41**(6), 471–483 (1993)

2.17 D.H. Keefe: *Woodwind Tone Hole Acoustics and the Spectrum Transformation Function*, Ph.D. Thesis (Physics Dept., Case Western Reserve University, Cleveland 1981)

2.18 D.H. Keefe: Woodwind air column models, J. Acoust. Soc. Am. (JASA) **88**(1), 35–51 (1990)

2.19 D.H. Keefe: Acoustical wave propagation in cylindrical ducts: Transmission line parameter approximations for isothermal and nonisothermal boundary conditions, J. Acoust. Soc. Am. (JASA) **75**(1), 58–62 (1984)

2.20 A.H. Benade: Equivalent circuits for conical waveguides, J. Acoust. Soc. Am. (JASA) **83**, 1764–1769 (1988)

2.21 P.M. Morse, U. Ingard: *Theoretical Acoustics* (Princeton Univ. Press, Princeton 1987) pp. 1–927

2.22 C. Valette: The mechanics of vibrating strings. In: *Mechanics of Musical Instruments*, Courses and Lectures/International Centre for Mechanical Sciences, Vol. 355, ed. by A. Hirschberg, J. Kergomard, G. Weinreich (Springer, Wien 1995) pp. 115–184

2.23 I. Testa, G. Evangelista, S. Cavaliere: Physically inspired models for the synthesis of stiff strings with dispersive waveguides, EURASIP J. Adv. Signal Process. **7**, 964–977 (2004)

3. Waves in Two and Three Dimensions

Wilfried Kausel

This chapter deals with the generalization of the wave equation to describe wave propagation on two-dimensional surfaces and sound waves in a three-dimensional space. Again linearity is postulated, which is only justified if amplitudes are sufficiently small. It will be shown how wave equations can be derived for rectangular and circular membranes, plates and disks and how analytic results can be obtained for a three-dimensional case with relatively simple boundary conditions. This chapter will also review techniques for the calculation of resonance frequencies and for the prediction of associated modal shapes.

3.1 Waves on a Surface

Here we will generalize the one-dimensional wave equation to two dimensions and simultaneously further develop the mathematics that are needed to do that. This way we will be able to approximately describe vibrations of piano sound boards, timpani membranes and of several other percussion instruments. Finally we will cover the most general case of sound wave propagation in the free three-dimensional space.

3.1.1 Rectangular Membrane

Recalling the derivation of transverse string vibrations, we will find many similarities in the two-dimensional case of Fig. 3.1. The string tension T, a force stretching each string element, now becomes the surface tension T_S, a force per unit length, from which two pairs of forces, $\pm T_S\,\mathrm{d}x$ and $\pm T_S\,\mathrm{d}y$ can be derived that stretch each surface element tangentially in x and in y directions. The resulting z component will drive the segment mass vertically.

This uncompensated z component of the tension is caused by a curvature in the x direction ($\partial^2 z/\partial x^2 \mathrm{d}x$) and by a curvature in the y direction ($\partial^2 z/\partial y^2 \mathrm{d}y$). If there is zero curvature then there will be no resulting z component because tension vectors at opposite sides of the segment will have the same direction but opposite orientation and therefore cancel out each other.

The equation of motion can now be written as

$$m\frac{\partial^2 z}{\partial t^2} = \left(T_S\,\mathrm{d}y\frac{\partial^2 z}{\partial x^2}\mathrm{d}x + T_S\,\mathrm{d}x\frac{\partial^2 z}{\partial y^2}\mathrm{d}y\right). \tag{3.1}$$

With the mass $m = \rho_A\,\mathrm{d}x\mathrm{d}y$ of the differential segment we obtain a two-dimensional wave equation for the displacement z of a stretched membrane in Cartesian

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_3

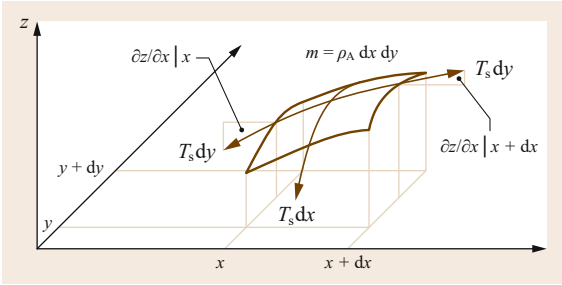


Fig. 3.1 Waves on a surface

coordinates

$$\frac{\partial^2 z}{\partial t^2} = \frac{T_S}{\rho_A}\left(\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2}\right). \tag{3.2}$$

The phase propagation speed in x and y directions is $c = \sqrt{T_S/\rho_A}$. It is again constant for a homogeneous and isotropic membrane with constant area density ρ_A and tension.

Vector Analysis

It is now required to introduce some operators commonly used to make reading and writing of differential equations much easier, because they will frequently appear in the following sections.

We already mentioned the velocity gradient, meaning a gradual increase or decrease of the fluid velocity along a spatial dimension. It is simply the rate of change; the spatial derivative with respect to the space variable.

In the three-dimensional space the change rate per unit length can be different depending on whether we go in the x, y or z direction. If we imagine the xy-plane to be a hot wall, then the temperature change rate in the z direction (distance to the wall) will be perceptible while no change will be noticed in the x and y directions – at least if the wall is very big.

The change rate in three-dimensional space therefore becomes a vector with the components $\partial f/\partial x$, $\partial f/\partial y$ and $\partial f/\partial z$. Such a vector can be assigned to each point in space and it always points in the direction of the steepest ascent of the corresponding scalar quantity like temperature, pressure or velocity magnitude. Its length is the change rate per unit length in the indicated direction.

This vector is called the *gradient* of a scalar field $f(x, y, z)$ and we introduce the differential operator $(\partial/\partial x, \partial/\partial y, \partial/\partial z)^\top$, abbreviated by ∇ or *grad* to write gradients in two or more dimensions. A scalar field is a function that assigns a scalar value, like temperature or pressure, to all points in space.

Formally the ∇ operator is a vector and it is therefore possible to square it before it is applied to the field variable

$$\nabla^2 = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix}^2 = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right). \tag{3.3}$$

The operator ∇^2, often written as Δ, is called the Laplace operator. Using it, the wave equation (3.2) can be rewritten as

$$\frac{\partial^2 z}{\partial t^2} = \frac{T_S}{\rho_A}\Delta z\,. \tag{3.4}$$

Applying the ∇ operator to a scalar field gives us a vector field; that means each point in space is a vector assigned to indicating the direction and strength of the steepest increase of the underlying scalar field. When we apply the ∇ operator a second time, and this is exactly what we are doing when we use the Laplace operator Δ, then we apply it to a vector field now in order to obtain a new scalar field.

First let us verify that: $(\boldsymbol{\nabla}\cdot\boldsymbol{\nabla})z = \boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}z)$. This is obviously correct. In both cases we obtain a scalar result and a scalar multiplication can be done before or after the dot (inner or scalar) product.

Applying the ∇ operator to a vector field is called divergence (written as *div*). The Laplace operator therefore calculates the divergence of the gradient of a scalar field. Divergence is also called source or sink strength of the field.

If we confine each point of the vector field in an infinitesimally small cubic cage summing up and normalizing the vector components orthogonal to the faces of the cube, counting outward pointing contributions as positive and inward pointing contributions as negative, then we will obtain information about whether this volume or space element acts as a source or as a sink for the physical quantity represented by the vectors.

In the explanation of (2.27) a density change inside a small fluid element was related to a different velocity of the left and right fluid boundary. The velocity field is a vector field because velocity in general has a direction and a magnitude. In that one-dimensional case only two directions are possible, so we only need to take two opposite faces of our measurement cage into account.

If the velocity vector leaving the cage on the right side is greater then the velocity vector entering on the left side then we have a positive divergence, which was required to lower the density. We will find this equation generalized for three dimensions later as (3.13). Instead of $\partial v/\partial x$ it will contain the ∇ or *div* operator applied to the vector field $\boldsymbol{v}$.

Note that in this subsection we are working on a problem in two spatial coordinates only. All derivatives with respect to a third or higher spatial coordinate are zero. Therefore it is not necessary to define separate differential operators for the two-dimensional, three-dimensional and even more-dimensional space.

3.1.2 Circular Membrane

In order to take advantage of the symmetry of a circular membrane, it is advantageous to use a polar coordinate system with an origin in the center of the membrane. Each point of the membrane is characterized by a coordinate pair (r, θ). r determines its distance from the center point and θ specifies an angle in the range $[-\pi, \pi]$ from the r axis, an arbitrarily chosen ray extending from the center point.

The correspondence between the coordinates is given by

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \\ r &= \sqrt{x^2+y^2} \\ \tan\theta &= \frac{y}{x}\,. \end{aligned} \tag{3.5}$$

Using these correspondences to substitute x and y we can transform a function $f(x, y)$ into an equivalent function $g(r, \theta)$. In many cases the function g can be found directly, so f might not be known.

Although it would be possible to reconstruct f using the inverse transformation, we want to avoid that, because it usually leads to lengthy expressions involving $\tan^{-1}$ (arctan or inverse tangent) and ugly square root functions. Therefore we first need to transform the required differential operators into the polar coordinate system.

To make equations shorter and easier to read we will introduce a new way to write partial derivatives, according to

$$\begin{aligned} \frac{\partial f}{\partial x} &\to f_\mathrm{x}\,, \\ \frac{\partial g}{\partial r} &\to g_\mathrm{r}\,, \\ \frac{\partial^2 f}{\partial y^2} &\to f_\mathrm{yy}\,, \\ \frac{\partial^2 g}{\partial r,\partial\theta} &\to f_{\mathrm{r}\theta}\cdots\,. \end{aligned} \tag{3.6}$$

To calculate the first derivatives of any function u of two variables using the correspondences from (3.5) the chain rule of differentiation has to be applied

$$\begin{aligned} u_\mathrm{r} &= u_\mathrm{x}x_\mathrm{r} + u_\mathrm{y}y_\mathrm{r} \\ &= u_\mathrm{x}\cos\theta + u_\mathrm{y}\sin\theta \\ u_\theta &= u_\mathrm{x}x_\theta + u_\mathrm{y}y_\theta \\ &= -u_\mathrm{x}r\sin\theta + u_\mathrm{y}r\cos\theta\,. \end{aligned} \tag{3.7}$$

Multiplying one of these two equations for u_r and u_θ with $\cos\theta$ and the other one with $\sin\theta$ we can obtain u_x and u_y in terms of r and θ. This gives us the gradient of the scalar field u in polar coordinates

$$\begin{aligned} u_\mathrm{x} &= u_\mathrm{r}\cos\theta - \frac{u_\theta\sin\theta}{r} \\ u_\mathrm{y} &= u_\mathrm{r}\sin\theta + \frac{u_\theta\cos\theta}{r}\,. \end{aligned} \tag{3.8}$$

The same derivation for the second derivatives is lengthy. Therefore only the starting point and the result is given

$$\begin{aligned} u_\mathrm{rr} &= \frac{\partial}{\partial r}\left(u_\mathrm{x}\cos\theta + u_\mathrm{y}\sin\theta\right) = \cdots \\ u_{\theta\theta} &= r\frac{\partial}{\partial\theta}\left(-u_\mathrm{x}\sin\theta + u_\mathrm{y}\cos\theta\right) = \cdots \\ u_\mathrm{rr} + \frac{u_{\theta\theta}}{r^2} &= \cdots \\ \cdots &= u_\mathrm{xx} + u_\mathrm{yy} - \frac{1}{r}u_\mathrm{r}\,. \end{aligned} \tag{3.9}$$

Now we have the Laplace operator

$$\begin{aligned} \Delta &= \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \\ &= \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial\theta^2} \end{aligned} \tag{3.10}$$

and the wave equation in polar coordinates

$$\frac{\partial^2 z}{\partial t^2} = c^2\left(\frac{\partial^2 z}{\partial r^2} + \frac{1}{r}\frac{\partial z}{\partial r} + \frac{1}{r^2}\frac{\partial^2 z}{\partial\theta^2}\right). \tag{3.11}$$

3.1.3 Rectangular Plate

Looking at the correspondence between the wave equations of a vibrating string and of a rectangular membrane, we might generalize (2.37) intuitively by replacing $\partial^4 y/\partial x^4$ by $\nabla^4 z$ and by substituting $h^2/12$ for I_a/S. The latter can been derived for a rectangular plate of thickness h by means of (2.34). This approach leads to a nearly correct equation. The correct

version is [3.1–3]

$$\frac{\partial^2 z}{\partial t^2} = -\frac{E}{\rho(1-\nu^2)}\frac{h^2}{12}\nabla^4 z\,. \tag{3.12}$$

The only small difference is the term involving the Poisson's ratio ν. When we derived the equation for the bar, we assumed a bundle of wires. We did not take into account the fact that wires that are stretched are getting thinner and wires that are compressed are getting thicker. In a bar this does not matter too much. The change in thickness is not constrained if the bar is much longer than its width and height.

In a plate with a significant width the distortion of the cross-sectional area due to length differences of top and bottom layers, expressed by the Poisson's ratio ($\nu \approx 0.3$), has to overcome higher resistance, which increases the effective stiffness.

For the same reason even compressional (longitudinal) waves propagate faster in an infinite plate than in strings or stiff bars. The effective Young's modulus is scaled by the same factor

$$c'^2 = \frac{E}{\rho(1-\nu^2)}\,. \tag{3.13}$$

3.1.4 Circular Disk

For circular disks we need to express $\nabla^4 = \Delta^2$ in (3.12) in polar coordinates. With (3.10) we obtain

$$\begin{aligned}\nabla^4 &= \left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}\right)^2 \\ &= \frac{\partial^4}{\partial r^4} + \frac{1}{r^2}\frac{\partial^2}{\partial r^2} + \frac{1}{r^4}\frac{\partial^4}{\partial\theta^4} \\ &\quad + \frac{2}{r}\frac{\partial^3}{\partial r^3} + \frac{1}{r^2}\frac{\partial^4}{\partial r^2\partial\theta^2} + \frac{2}{r^3}\frac{\partial^3}{\partial r\partial\theta^2}\,.\end{aligned} \tag{3.14}$$

This allows us to specify boundary conditions and solutions in terms of polar coordinates. We will do this in another section below.

3.2 Solution for Waves on a Surface

In this section it will be shown how the two-dimensional wave equation of a rectangular or circular membrane can be solved to get information about the structural vibrations of timpani membranes or thin rectangular plates like piano soundboards.

3.2.1 Rectangular Membrane

Looking at the ansatz (2.42) for the solution of a one-dimensional wave equation, we observe that it is a product of easily differentiable (exponential) terms, each solely dependent on a single variable. This solution technique is called *separation of variables* and can also be applied to the more-dimensional case.

We therefore assume a solution of the form

$$z(x,y,t) = F(x)G(y)\mathrm{e}^{\mathrm{i}\omega t} \tag{3.15}$$

for the two-dimensional wave equation (3.2). With

$$\begin{aligned} z_{\mathrm{xx}} &= F_{\mathrm{xx}}G\mathrm{e}^{\mathrm{i}\omega t} \\ z_{\mathrm{yy}} &= G_{\mathrm{yy}}F\mathrm{e}^{\mathrm{i}\omega t} \\ z_{\mathrm{tt}} &= -\omega^2\mathrm{e}^{\mathrm{i}\omega t}FG \end{aligned} \tag{3.16}$$

the wave equation $z_{\mathrm{tt}} = c^2(z_{\mathrm{xx}} + z_{\mathrm{yy}})$ becomes

$$-\omega^2\mathrm{e}^{\mathrm{i}\omega t}FG = c^2\left(F_{\mathrm{xx}}G\mathrm{e}^{\mathrm{i}\omega t} + G_{\mathrm{yy}}F\mathrm{e}^{\mathrm{i}\omega t}\right) \tag{3.17}$$

and therefore

$$-\frac{\omega^2}{c^2} - \frac{G_{\mathrm{yy}}}{G(y)} = \frac{F_{\mathrm{xx}}}{F(x)}\,. \tag{3.18}$$

It can be seen that the left side of (3.18) only varies with y while the right side only varies with x. For any given value of x it should be possible to vary y in some way to make the equation wrong. The equation can only be correct for all possible pairs of x and y if both sides actually do not vary at all and are independent of both x and y and therefore constant.

Calling this constant $-k^2$ we have to solve

$$\begin{aligned} -k^2 &= \frac{F_{\mathrm{xx}}}{F(x)} \\ -k^2 &= -\frac{\omega^2}{c^2} - \frac{G_{\mathrm{yy}}}{G(y)}\,. \end{aligned} \tag{3.19}$$

These equations are satisfied by

$$\begin{aligned} F(x) &= \mathrm{Re}\left(A\mathrm{e}^{\mathrm{i}kx}\right) \\ G(y) &= \mathrm{Re}\left(B\mathrm{e}^{\mathrm{i}\sqrt{\frac{\omega^2}{c^2}-k^2}\,y}\right)\,. \end{aligned} \tag{3.20}$$

Considering a rectangular membrane fixed at $x = 0$ and $x = L_{\mathrm{x}}$ as well as at $y = 0$ and $y = L_{\mathrm{y}}$ we obtain the fol-

lowing system

$$\begin{aligned}
0 &= \mathrm{Re}\left(A\mathrm{e}^{\mathrm{i}k0}\right)\mathrm{Re}\left(B\mathrm{e}^{\mathrm{i}\sqrt{\frac{\omega^2}{c^2}-k^2}y}\right)\\
0 &= \mathrm{Re}\left(A\mathrm{e}^{\mathrm{i}kL_\mathrm{x}}\right)\mathrm{Re}\left(B\mathrm{e}^{\mathrm{i}\sqrt{\frac{\omega^2}{c^2}-k^2}y}\right)\\
0 &= \mathrm{Re}\left(A\mathrm{e}^{\mathrm{i}kx}\right)\mathrm{Re}\left(B\mathrm{e}^{\mathrm{i}\sqrt{\frac{\omega^2}{c^2}-k^2}0}\right)\\
0 &= \mathrm{Re}\left(A\mathrm{e}^{\mathrm{i}kx}\right)\mathrm{Re}\left(B\mathrm{e}^{\mathrm{i}\sqrt{\frac{\omega^2}{c^2}-k^2}L_y}\right). \qquad (3.21)
\end{aligned}$$

From this we obtain the conditions ($m, n \in \mathbb{N}$)

$$\begin{aligned}
\mathrm{Re}(A) &= 0\\
kL_\mathrm{x} &= m\pi\\
\mathrm{Re}(B) &= 0\\
\sqrt{\frac{\omega^2}{c^2}-k^2}L_\mathrm{y} &= n\pi\,. \qquad (3.22)
\end{aligned}$$

The complex amplitudes A and B have to be purely imaginary, which rotates the complex exponentials by $\pi/2$. Equations 3.20 become

$$F(x) = |A|(\mathrm{i}\cos(kx) - \sin(kx)) \qquad (3.23)$$

$$\begin{aligned}
G(y) = |B|\Bigg(&\mathrm{i}\cos\left(y\sqrt{\frac{\omega^2}{c^2}-k^2}\right)\\
&-\sin\left(y\sqrt{\frac{\omega^2}{c^2}-k^2}\right)\Bigg). \qquad (3.24)
\end{aligned}$$

The z-displacement of $m\times n$ normal modes of a rectangular membrane can therefore be written as

$$\begin{aligned}
z_{m,n}(x,y,t) = C\sin\left(\frac{m\pi}{L_\mathrm{x}}x\right)\sin\left(\frac{n\pi}{L_\mathrm{y}}y\right)\\
\times\sin(\omega t+\phi)\,. \qquad (3.25)
\end{aligned}$$

The integer numbers $m+1$ and $n+1$ are the numbers of nodal lines along the x- y-dimension respectively. Solving (3.22) for ω gives their corresponding angular frequencies as

$$\omega_{m,n} = c\,\pi\sqrt{\left(\frac{m}{L_\mathrm{x}}\right)^2+\left(\frac{n}{L_\mathrm{y}}\right)^2}. \qquad (3.26)$$

The physical displacement of the membrane must be any linear combination of these normal modes e.g., the sum of any number of normal modes with individual amplitude and individual phase.

Especially in the case where L_x and L_y are equal or have a common multiple, normal mode frequencies (eigenfrequencies) of different modes (e.g., $\omega_{n,m} = \omega_{m,n}$) will coincide.

If the membrane is stimulated at one of these frequencies all of the related normal modes will occur simultaneously with arbitrary phase differences between them. This can create unexpected patterns with rotating nodal lines and even shapes that only rotate but not oscillate, as shown in Fig. 3.2. They are called *degenerate modes*.

3.2.2 Circular Membrane

Now we apply the same method to the wave equation in polar coordinates given by (3.11). Separating the variables r, θ and t gives us a solution of the form

$$z(r,\theta,t) = F(r)G(\theta)\mathrm{e}^{\mathrm{i}\omega t}\,. \qquad (3.27)$$

With

$$\begin{aligned}
z_{\mathrm{rr}} &= F_{\mathrm{rr}}G\mathrm{e}^{\mathrm{i}\omega t}\\
z_\mathrm{r} &= F_\mathrm{r}G\mathrm{e}^{\mathrm{i}\omega t}\\
z_{\theta\theta} &= G_{\theta\theta}F\mathrm{e}^{\mathrm{i}\omega t}\\
z_{\mathrm{tt}} &= -\omega^2\mathrm{e}^{\mathrm{i}\omega t}FG \qquad (3.28)
\end{aligned}$$

the wave equation $z_{\mathrm{tt}} = c^2(z_{\mathrm{rr}} + z_\mathrm{r}/r + z_{\theta\theta}/r^2)$ can be reordered with the intention to let both sides depend on a single variable only. This technique, called *separation of variables*, has already been explained in Sect. 3.2.1. We obtain

$$-\frac{G_{\theta\theta}}{G} = \frac{r^2F_{\mathrm{rr}}}{F}+\frac{r^2\omega^2}{c^2}+\frac{rF_\mathrm{r}}{F}\,. \qquad (3.29)$$

This equation can only be correct for all possible pairs (r,θ) if both sides are actually constant. Giving this constant the name m^2 we obtain two equations

$$\begin{aligned}
m^2 &= -\frac{G_{\theta\theta}}{G}\\
m^2 &= \frac{r^2F_{\mathrm{rr}}}{F}+\frac{r^2\omega^2}{c^2}+\frac{rF_\mathrm{r}}{F}\,. \qquad (3.30)
\end{aligned}$$

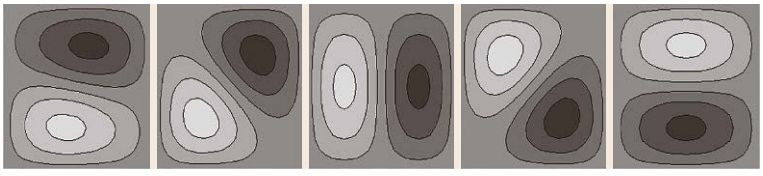


Fig. 3.2 Square membrane, degenerate mode

The first one we already know. It has the solution $G(\theta) = Ae^{\pm im\theta}$. θ is not restricted to the range $[0, 2\pi]$ and we want $G(\theta)$ to be single valued; that means $G(\theta) = G(\theta + 2\pi)$. Therefore the constant m must be an integer number. It can take the values $0, 1, 2, \ldots$ and it is related to the number of nodal diameters.

The second equation can be rearranged as

$$F_{rr} + \frac{F_r}{r} + F\left(\frac{\omega^2}{c^2} - \frac{m^2}{r^2}\right) = 0\,. \tag{3.31}$$

This is Bessel's equation and the general solutions are known as the so-called *Bessel functions* $J_m(kr)$ with $k = \omega/c$ and integer order m.

A circular membrane is fixed by the circular tensioning hoop where zero displacement is enforced. This requires $J_m(kr) = 0$ at $r = R$, R being the radius of the membrane. But Bessel functions have many zeros, so the boundary condition is satisfied for the whole ordered sequence of zeros of $J_m(kr)$.

If we take the smallest zero, the only nodal circle is the tensioning hoop. If we let the n-th zero be the one required at $r = R$, then we will get $n-1$ additional nodal circles at radii corresponding to all smaller zeros. The number n therefore designates the number of nodal circles and it is the sequential index of the corresponding zero of $J_m(kr)$.

The resonance frequency of mode (m, n) can be calculated from $\gamma_{m,n}$ being the n-th zero of the Bessel function of order m according to $f_{m,n} = c\gamma_{m,n}/2\pi R$. Figure 3.3 shows the first 15 normal modes sorted by relative frequency. It can be seen that the radial modes with a single nodal circle $(1, 1), (2, 1), (3, 1), \cdots$ are very close to a harmonic series. They are giving the timpani its audible pitch.

3.2.3 Rectangular Plate

Following the example of the rectangular membrane we assume a general solution according to (3.15) with partial derivatives as shown in (3.16). The equation of motion of a plate (3.12) is of fourth order so we also need fourth-order derivatives

$$\begin{aligned} z_{xxxx} &= F_{xxxx} G e^{i\omega t} \\ z_{yyyy} &= G_{yyyy} F e^{i\omega t} \\ z_{xxyy} &= F_{xx} G_{yy} e^{i\omega t}\,. \end{aligned} \tag{3.32}$$

Substituting these derivatives into the fourth-order differential equation for the plate yields

$$F_{xxxx} G + G_{yyyy} F + 2F_{xx} G_{yy} = \frac{\omega^2}{\Lambda} FG \tag{3.33}$$

with $\Lambda = h^2 E/(12\rho(1-\nu^2))$.

It can be seen that it is not possible to separate the variables like in (3.18) because of the mixed term z_{xxyy}. Assuming sinusoidal (or strictly speaking cosinusiodal) spatial profiles in both axial directions we can again use the ansatz $F(x) = \mathrm{Re}(Ae^{ik_x x})$ and $G(y) = \mathrm{Re}(Be^{ik_y y})$. This gives us a relation between the wave numbers k_x, k_y and the frequency ω

$$k_x^4 + k_y^4 + 2k_x^2 k_y^2 = \frac{\omega^2}{\Lambda} \tag{3.34}$$

or

$$k_x^2 + k_y^2 = k^2 = \pm\frac{\sqrt{12}\omega}{h}\sqrt{\frac{\rho(1-\nu^2)}{E}} \tag{3.35}$$

with k being the wave number in the actual direction of wave propagation. It can be seen that there are imaginary solutions for the wave number k. An imaginary

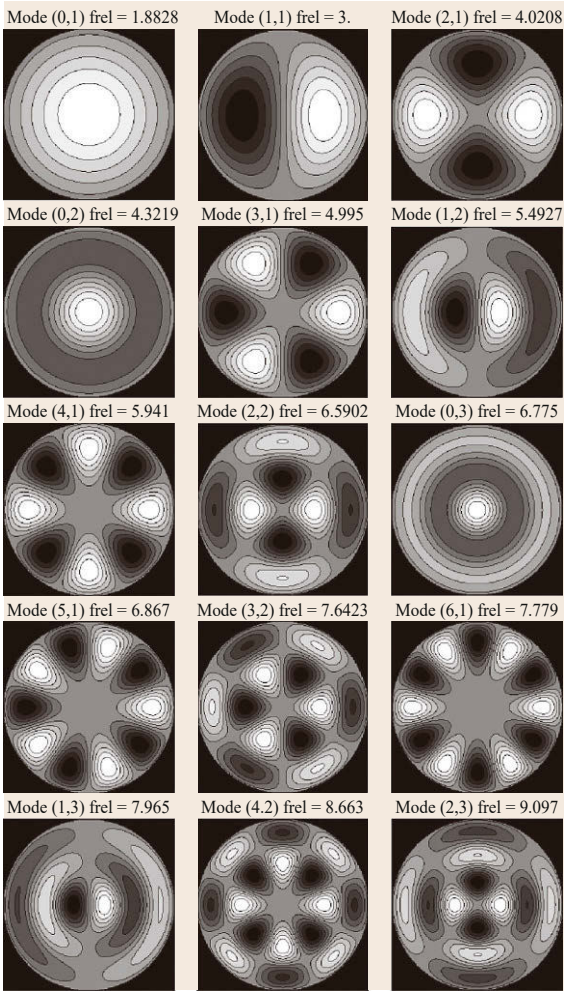


Fig. 3.3 Circular membrane normal modes

wave number causes a real but negative coefficient of the spatial variables x and y in the exponential terms for the profiles F and G. This means an exponential decay of the wave amplitude. It is said the wave is *evanescent* and does not propagate. It decays by about 55 dB per wavelength [3.4].

The phase speed c_{ph} of waves is given by ω/k. In this case it is

$$c_{\mathrm{ph}} = \sqrt{\frac{h\omega}{\sqrt{12}}\sqrt{\frac{E}{\rho(1-\nu^2)}}}\,. \tag{3.36}$$

Unlike in strings and membranes without stiffness it is not constant here but frequency dependent. This is why (3.35) is named the *dispersion relation*. The relation between frequency f and wave number k is

$$f = hk^2\frac{1}{2\pi\sqrt{12}}\sqrt{\frac{E}{\rho(1-\nu^2)}}\,. \tag{3.37}$$

It can be used to calculate resonance frequencies if the wave number k is fixed by boundary conditions.

Hagedorn calculated mode shapes and frequencies for a simply supported rectangular plate [3.2] by splitting the fourth-order differential equation into two second-order ones, one of them being identical to the wave equation of the rectangular membrane. He obtained the same modal shapes as for the rectangular membrane but at different modal frequencies

$$\omega_{m,n} = \pi^2\sqrt{\Lambda}\left(\left(\frac{m}{L_{\mathrm{x}}}\right)^2 + \left(\frac{n}{L_{\mathrm{y}}}\right)^2\right). \tag{3.38}$$

For all other boundary conditions rectangular plates have to be treated numerically.

3.2.4 Circular Disk

Let us see how the forth-order differential equation can be split into a product of second-order equations

$$(\nabla^4 - \gamma^4)Z = (\nabla^2 + \gamma^2)(\nabla^2 - \gamma^2)Z = 0\,. \tag{3.39}$$

Comparing (3.39) with (3.12) we can determine that for $z(x,y,t) = Z(x,y)\mathrm{e}^{\mathrm{i}\omega t}$ the factor γ must satisfy $\gamma^4 = \omega^2/\Lambda$. The abbreviation Λ has already been defined with (3.33).

Therefore a solution for the spatial profile $Z(x,y)$ must satisfy either $(\nabla^2 + \gamma^2)Z = 0$ or $(\nabla^2 - \gamma^2)Z = 0$. A differential equation of the first type has already been solved for the rectangular membrane as well as for the circular membrane with $\gamma = \omega/c$. Recapitulating the derivation in polar coordinates of Sect. 3.2.2 we obtain two solutions according to

$$Z_1(r,\theta) = A\mathrm{e}^{\pm\mathrm{i}m\theta}J_m(\gamma r)\,. \tag{3.40}$$

In order to solve the second equation we have to substitute γ by the imaginary factor $\mathrm{i}\gamma$. The solutions of Bessel's equation are now called hyperbolic Bessel functions $I_{\mathrm{m}}(\gamma r) = \mathrm{i}^{-m}J_{\mathrm{m}}(\mathrm{i}\gamma r)$. This gives us more solutions for Z according to

$$Z_2(r,\theta) = A\mathrm{e}^{\pm\mathrm{i}m\theta}I_m(\gamma r)\,. \tag{3.41}$$

Possible spatial profiles are therefore the linear combinations

$$Z(r,\theta) = \mathrm{e}^{\pm\mathrm{i}m\theta}(AJ_m(\gamma r) + BI_m(\gamma r))\,. \tag{3.42}$$

If the plate is clamped at a distance R from the center, boundary conditions

$$\begin{aligned} Z(R,\theta) &= 0 \\ \frac{\partial}{\partial r}Z(r,\theta) &= 0|r = R \end{aligned} \tag{3.43}$$

are enforced. This requires

$$\begin{aligned} AJ_m(\gamma R) + BI_m(\gamma R) &= 0 \\ \frac{\partial}{\partial r}(AJ_m(\gamma R) + BI_m(\gamma R)) &= 0|r = R\,. \end{aligned} \tag{3.44}$$

The first condition is satisfied by

$$B = -A\frac{J_m(\gamma R)}{I_m(\gamma R)}\,. \tag{3.45}$$

Now the second condition becomes

$$I_m(\gamma R)\frac{\partial J_m(\gamma r)}{\partial r} = J_m(\gamma R)\frac{\partial I_m(\gamma r)}{\partial r}|r = R\,. \tag{3.46}$$

The solutions for $\beta_{mn} = R\gamma_{mn}/\pi$, with m the number of nodal diameters and n the number of nodal circles, have been given by *Morse* in [3.3] as

$$\begin{aligned} \beta_{01} &= 1.015 & \beta_{11} &= 1.468 & \beta_{21} &= 1.879 \\ \beta_{02} &= 2.007 & \beta_{12} &= 2.483 & \beta_{22} &= 2.992 \\ \beta_{03} &= 3.000 & \beta_{13} &= 3.490 & \beta_{23} &= 4.000\,. \end{aligned}$$

With these solutions it is possible to determine the corresponding frequencies according to

$$f_{mn} = \frac{\pi}{2R^2}\sqrt{\frac{h^2E}{12\rho(1-\nu^2)}}\beta_{mn}^2\,. \tag{3.47}$$

3.3 Sound Waves in Space

It is straightforward to generalize (3.4) for sound propagation in three-dimensional space as

$$\frac{\partial^2 p}{\partial t^2} = c^2 \Delta p \,, \tag{3.48}$$

with Δ being the Laplace operator in three dimensions now. But first, this does not give us any relationship between the speed of sound c and basic physical quantities, and second, now might be an opportunity to demonstrate how this equation can also be systematically derived from the first principles of physics.

3.3.1 Wave Equation in Three Dimensions

The fundamental laws of fluid dynamics are the Navier–Stokes conservation laws. The first deals with conservation of mass. No mass can appear from nothingness or disappear into nothingness. The second deals with conservation of momentum mv. Actually it says that any change of a momentum $\mathrm{d}(mv)/\mathrm{d}t$ must be caused by a force. The third is about conservation of energy, but it is enough to consider the first two here.

What remains of the conservation laws for small perturbations $\hat{\rho}$, $\hat{p}$ and $\hat{\boldsymbol{v}}$ (note, that the fluid velocity v is a vector with components in all three spatial dimensions now) of a stagnant ($v_0 = 0$, $v = \hat{v}$) uniform fluid characterized by $p = \hat{p} + p_0$ and $\rho = \hat{\rho} + \rho_0$ can now be given as

$$\begin{aligned} \frac{\partial \hat{\rho}}{\partial t} + \rho_0 \nabla \cdot \hat{\boldsymbol{v}} &= 0 \\ \rho_0 \frac{\partial \hat{\boldsymbol{v}}}{\partial t} + \nabla \hat{p} &= 0 \,. \end{aligned} \tag{3.49}$$

Building the time derivative of the mass conservation law (first equation of (3.49)) and the divergence of the momentum equation (the second one)

$$\begin{aligned} \frac{\partial^2 \hat{\rho}}{\partial t^2} + \rho_0 \frac{\partial}{\partial t}(\nabla \cdot \hat{\boldsymbol{v}}) &= 0 \\ \rho_0 \frac{\partial}{\partial t}(\nabla \cdot \hat{\boldsymbol{v}}) + \Delta \hat{p} &= 0 \end{aligned} \tag{3.50}$$

we can eliminate all v-related terms by subtracting the second equation from the first one. With the proportionality between ρ and p assuming adiabatic state transitions we obtain the wave equation

$$\frac{\partial^2 \hat{p}}{\partial t^2} = c^2 \nabla^2 \hat{p} \equiv c^2 \Delta \hat{p} \,, \tag{3.51}$$

with $c = \sqrt{(c_\mathrm{p} - c_\mathrm{v})T}$. Here $\hat{p}$ is usually replaced by p, if there is no danger of mistaking linearized sound pressure and atmospheric pressure any longer.

This wave equation does not cover the cases where significant air flow, high sound pressures or other nonlinearities are present. Especially in the field of musical acoustics there are essential questions that cannot be answered based on such simplifications.

Sound propagation in brass wind instruments can easily be driven into a nonlinear region for which the wave equation is no longer valid. Wave steepening and shock wave effects are due to the high sound pressure levels generated inside the instrument. To model flow-driven instruments like flue pipes or flutes the Navier–Stokes equations in their general form have to be solved.

Modeling wave propagation does not cover all the nonlinear effects occurring in the primary oscillators of wind instruments either. Waves are propagating on the brass player's lips, the singer's vocal folds and in the air columns of all wind instruments including the singing voice, but the coupling between the vibrating reeds of clarinets, oboes or tubas to the resonator, the air column of the instrument, is highly nonlinear.

3.3.2 Rectangular Coordinates

Calculating the resonances of a simple rectangular room is a good exercise here and will have significance for room acoustics. We separate the variables for the pressure p just like in Sect. 3.2.1 on the rectangular membrane according to

$$p(x, y, z, t) = F(x)G(y)H(z)\mathrm{e}^{\mathrm{i}\omega t} \,. \tag{3.52}$$

Now we avoid complex exponentials for the space-dependent terms F, G and H because we learned in Sect. 3.2.1 that we did not gain much in using them in that place. Here we will work with the real part of the complex exponentials only. We try with

$$\begin{aligned} F(x) &= A\cos(k_1 x) \\ G(y) &= B\cos(k_2 y) \\ H(z) &= C\cos(k_3 z) \end{aligned} \tag{3.53}$$

leading to

$$\begin{aligned} p_{\mathrm{xx}} &= -k_1^2\, p(x, y, z, t) \\ p_{\mathrm{yy}} &= -k_2^2\, p(x, y, z, t) \\ p_{\mathrm{zz}} &= -k_3^2\, p(x, y, z, t) \\ p_{\mathrm{tt}} &= -\omega^2\, p(x, y, z, t) \,. \end{aligned} \tag{3.54}$$

With the three-dimensional wave equation $p_{\mathrm{tt}} = c^2(p_{\mathrm{xx}} + p_{\mathrm{yy}} + p_{\mathrm{zz}})$ we obtain a relation for the angular frequency ω of a mode when the wave numbers k_i in

the three axial directions are known

$$\omega = c\sqrt{k_1^2 + k_2^2 + k_3^2}\,. \tag{3.55}$$

We know that rigid walls represent a free boundary for the sound pressure but a fixed boundary for the velocity. While zero velocity is always enforced by rigid walls, this is not true for the sound pressure. At resonance the sound pressure will exhibit a local maximum at the wall. We can also say that rigid walls enforce a vibrational antinode of the sound pressure.

A local maximum always coincides with a zero derivative, which means a horizontal tangent. And this is how we are going to specify our boundary conditions.

The first derivative of the pressure with respect to a spatial coordinate must always be zero at the left boundary $x = 0$, $y = 0$ or $z = 0$ and at the right boundary $x = L_x$, $y = L_y$ or $z = L_z$ of a room with dimensions L_x, L_y and L_z.

As $\sin(0)$ is zero, the left boundary conditions are always met – we do not have to care about them. The remaining boundary conditions are therefore

$$\begin{aligned} 0 = p_x &\equiv -Ak_1 \sin(k_1 x) GH e^{i\omega t}|_{x=L_x} \\ 0 = p_y &\equiv -Bk_2 \sin(k_2 y) FH e^{i\omega t}|_{y=L_y} \\ 0 = p_z &\equiv -Ck_3 \sin(k_3 z) FG e^{i\omega t}|_{z=L_z}\,. \end{aligned} \tag{3.56}$$

These conditions are always and everywhere only met for

$$\begin{aligned} k_1 L_x &= n_1 \pi \\ k_2 L_y &= n_2 \pi \\ k_3 L_z &= n_3 \pi\,. \end{aligned} \tag{3.57}$$

This gives us the wave numbers of all possible normal modes of that room and with (3.55) the corresponding resonance frequencies

$$\begin{aligned} k_1 &= \frac{n_1 \pi}{L_x} \\ k_2 &= \frac{n_2 \pi}{L_y} \\ k_3 &= \frac{n_3 \pi}{L_z} \\ \omega &= c\pi \sqrt{\left(\frac{n_1}{L_x}\right)^2 + \left(\frac{n_2}{L_y}\right)^2 + \left(\frac{n_3}{L_z}\right)^2} \end{aligned} \tag{3.58}$$

with $n_i = 0, 1, 2, 3, \cdots$.

If room dimensions have common integer multiples – a room with $L_x = L_y = L_z$ being the most extreme example – then different resonances will coincide and become very strong but sparse. Rooms are said to be good for music performances if resonances are evenly spread and don't cluster.

However, resonant instrument bodies are far away from having a rectangular shape. Therefore their normal modes can only be determined numerically or experimentally.

3.3.3 Spherical Coordinates

If there is a sound source whose dimensions are small compared to the wavelengths to be studied, a point source with spherical wave propagation can be assumed. According to the symmetry of the problem a spherical coordinate system will preferably be chosen according to Fig. 3.4.

The Cartesian coordinates can be expressed in terms of radial distance r, polar angle θ and axial or azimuthal angle ψ according to

$$\begin{aligned} x &= r\cos\theta \sin\psi \\ y &= r\sin\theta \sin\psi \\ z &= r\cos\psi\,. \end{aligned} \tag{3.59}$$

The transformation of the three-dimensional Laplace operator $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$ into spherical coordinates would take up too much space here. It can be derived following the two-dimensional example from (3.7)–(3.10), but using the correspondences of (3.59).

Writing p for $\hat{p}$ the wave equation (3.51) now becomes

$$p_{tt} = \frac{c^2}{r^2}\left((r^2 p_r)_r + \frac{1}{\sin\theta}(p_\theta \sin\theta)_\theta + \frac{1}{\sin^2\theta} p_{\psi\psi} \right). \tag{3.60}$$

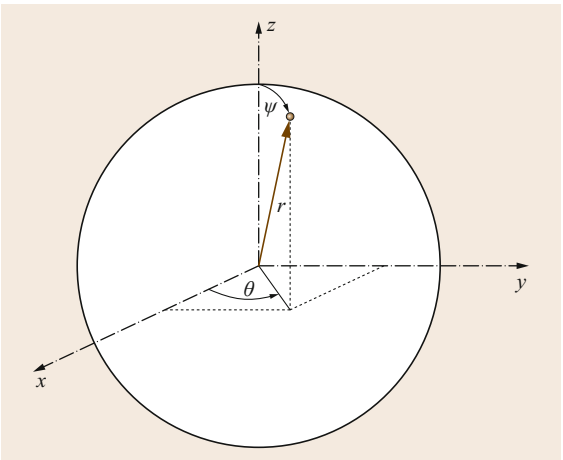


Fig. 3.4 Spherical coordinates

If the problem has a perfect spherical symmetry, like a pulsating sphere of any size in an undisturbed three-dimensional space, then all the derivatives with respect to the angles θ and ψ will become zero. The remaining wave equation is

$$p_{\mathrm{tt}} = \frac{c^2}{r^2}\left(\left(r^2 p_{\mathrm{r}}\right)_{\mathrm{r}}\right) . \tag{3.61}$$

Substituting $p \to p^* r$ and therefore $p_{\mathrm{r}} \to p^*_{\mathrm{r}}/r - p^*/r^2$ we finally obtain

$$p^*_{\mathrm{tt}} = c^2 p^*_{\mathrm{rr}} , \tag{3.62}$$

which is a one-dimensional wave equation for the spatial coordinate r and a pressure p^*, which is related to the original sound pressure p in that its relative amplitude is proportional to the distance r from the center.

The general solution for the sound pressure p is

$$p = \left(\frac{A}{r}\mathrm{e}^{-\mathrm{i}kr} + \frac{B}{r}\mathrm{e}^{\mathrm{i}kr}\right)\mathrm{e}^{\mathrm{i}\omega t} \tag{3.63}$$

Following the derivation of (3.30) with spatial variable r instead of x and with a different spatial derivative $p_{\mathrm{r}} = -\mathrm{i}kp - p/r$ we can obtain the impedance

$$Z = \frac{p}{v} = \rho c \frac{\mathrm{i}kr}{\mathrm{i}kr + 1} = -\rho c \frac{(kr)^2 + \mathrm{i}kr}{1 - (kr)^2} . \tag{3.64}$$

For $kr >> 1$ the impedance Z approaches the plane wave impedance ρc according to (2.30). On the other side if $kr << 1$ then the impedance approaches $-\rho c\mathrm{i}kr$ thus becoming purely imaginary with a phase difference between p and v of $-\pi/2$.

The sound intensity I of a single forward traveling wave ($B = 0$) with frequency ω at some distance r from the center is defined as the product of the amplitudes $\hat{p}/\sqrt{2}$ and $\hat{v}/\sqrt{2}$. As the velocity $v = p/Z$ the intensity can also be calculated according to $I = \hat{p}^2/(2Z) = \hat{p}^2 Y/2$, with Y being the reciprocal of the complex impedance Z, named *admittance*.

With the complex admittance

$$Y = \frac{1}{\rho c} - \frac{\mathrm{i}}{\rho c k r} \tag{3.65}$$

the real part of the sound intensity I becomes

$$\mathrm{Re}(I) = \hat{p}^2 \frac{\mathrm{Re}(Y)}{2} = \frac{|A|^2}{2\rho c r^2} . \tag{3.66}$$

It decreases with the square of the distance r because the total effective power $\mathrm{Re}(P)$ contained in a wavefront is spread over the surface $S = 4r^2\pi$ of a steadily growing sphere.

The imaginary part of the sound intensity I

$$\mathrm{Im}(I) = \hat{p}^2 \frac{\mathrm{Im}(Y)}{2} = -\frac{|A|^2}{2\rho c k r^3} \tag{3.67}$$

decreases with the third power of the distance r. This means that the total reactive power of a wavefront $\mathrm{Im}(P) = \mathrm{Im}(I)4r^2\pi$ is no longer constant but decreases with r. Close to the center it can become very high and even infinity for $r = 0$. This is not a violation of the law of conservation of energy because reactive power corresponds to a periodic transfer of energy between the pressure and the velocity wave and it is not dissipated.

It has to be noted that all the results obtained for spherical wave propagation in three-dimensional space are equally valid for one-dimensional wave propagation in a conical duct of any length or size. This is understandable because there is only a dependency on the distance from the center of the sphere, i. e., from the apex of the cone.

It doesn't matter whether we study spherical wavefronts as a whole or only a part of these spheres within a certain solid angle inside the boundaries of a cone. All variables are constant across the surface of any such sphere. Conical boundaries do therefore not change anything if the apex of the cone is identical to the center of the spheres. This is the reason why (2.64) is equally valid for cylindrical and conical ducts.

3.3.4 Cavities with Vents

It is well known that vented cavities like bottles exhibit acoustic resonances. These resonances can be excited by blowing air across the bottle neck or by popping it with the palm at the mouth of the bottle neck.

Helmholtz used these well-defined resonators, which later on were given his name, to detect frequency components in composite sounds [3.5]. His original drawing is shown in Fig. 3.5. Inserting the short nipple marked with the letter *b* into the ear, he could detect even faint air resonances stimulated by sound present at the open mouth labeled with the letter *a*.

This special kind of resonator is of essential importance in musical acoustics. All hollow bodies of stringed instruments, like violins, guitars, lutes and others, which are not completely closed but allow some acoustic flow through some kind of f-holes, c-holes or sound holes, act in this way.

Their cavity resonance is tuned to support the lowest register of these instruments, at frequencies that could otherwise never be radiated by normal plate resonances because the related wavelengths are usually much greater than the size of the plates.

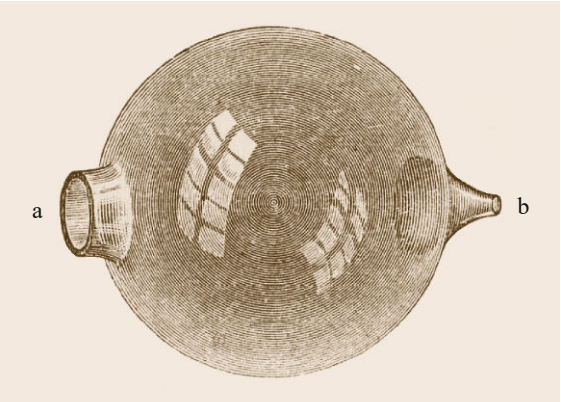


Fig. 3.5 Helmholtz resonators (after [3.5])

Although it appears as if Helmholtz resonators were three-dimensional cavities with wave propagation inside mainly closed boundaries, it is not necessary to treat them in this way. Usually their main resonance frequency is low enough so that the cavity is small compared to the wavelength at that frequency.

In that case the actual shape of the cavity no longer matters and solely its volume determines the acoustical characteristics, together with length and cross-sectional area of the vent.

The resonator can be treated as a single mass-spring system. The mass is provided by the air enclosed in the vent. It is given by $m = \rho SL$, with density of air ρ, cross-sectional area S and length L of the *bottle neck*.

In case of a very short *neck*, or no neck at all, L is an equivalent length taking a so-called *end-correction* into account. This end-correction of an open cylindrical tube lengthens the tube at each open end by a certain distance d, which was theoretically determined by *Levine* and *Schwinger* in [3.6] as $d = 0.61R$ (R being the radius of the cylindrical neck) and has later been confirmed experimentally.

In the case of sound holes we have to apply the end-correction twice (inside and outside) and get $L = T + 0.61D$ with thickness T of the top plate and diameter D of the sound hole. The open area S of the sound hole becomes $S = D^2\pi/4$.

The spring is provided by the elastic air enclosed in the cavity. A spring constant $E = F/z = \mathrm{d}F/\mathrm{d}z$ is the proportionality factor between force F and displacement z or the differential quotient if F is a function of z. The force is due to the air pressure p inside the cavity and given by $F = pS$.

For ideal gases we have $pV = nRT$. If we compress the air in our cavity (initial volume V_0, pressure p_0) by a piston of air in the neck then the reduced volume $V_0 - \mathrm{d}V$ will lead to an increased pressure $p_0 + \mathrm{d}p$.

Assuming that neither the temperature nor the amount of substance in the now reduced volume has significantly changed we can equate the products

$$\begin{aligned} p_0V_0 &= (p_0 + \mathrm{d}p)(V_0 - \mathrm{d}V) \\ &= p_0V_0 + V_0\,\mathrm{d}p - p_0\,\mathrm{d}V - \mathrm{d}p\,\mathrm{d}V\,. \end{aligned} \tag{3.68}$$

Neglecting the second-order term, substituting $\mathrm{d}V = S\,\mathrm{d}z$ and solving for $\mathrm{d}p$ we obtain

$$\mathrm{d}p = \frac{p_0}{V_0} S\,\mathrm{d}z\,. \tag{3.69}$$

For adiabatic conditions p_0 must be replaced by Kp_0 which equals to $c^2\rho$ according to (2.28).

With $\mathrm{d}F = \mathrm{d}pS$ we get the spring constant E according to

$$E = \frac{\mathrm{d}F}{\mathrm{d}z} = \frac{p_0}{V_0}S^2 = \frac{c^2\rho_0}{V_0}S^2 \tag{3.70}$$

using $Kp = c^2\rho$ derived from (2.28).

This leads to the Helmholtz resonance

$$\omega_0 = \sqrt{\frac{E}{m}} = c\sqrt{\frac{S}{V_0L}}\,. \tag{3.71}$$

Higher air modes cannot be explained by a single mass-spring model but can definitely be observed in stringed instrument bodies at frequencies where the shapes of cavity and sound holes as well as wave propagation inside the body can no longer be ignored. A full three-dimensional treatment of standing waves inside the body would be required to analyze these resonances. For such problems there are no analytical solutions and numerical methods have to be applied.

3.3.5 Solution for Long Ducts

It will now be shown how the three-dimensional wave equation can be solved for the boundary conditions given by long acoustic ducts with cylindrical or rectangular cross-section. The solution will not be restricted to cases where wavefronts can be treated as plane waves. Especially in horns with flaring bells the plane wave assumption is not at all realistic and one-dimensional approximations deviate significantly from the reality in which we encounter internal wavefronts, which are neither plane nor spherical.

In order to make use of an available axial symmetry, the wave equation is formulated in cylindrical coordinates. Just like any arbitrary displacement profile of a vibrating string at a given time point can be represented by a linear combination of all possible normal

modes (spatial superposition principle), any wavefront inside the duct, at a certain axial distance from the open end, and at a certain point in time, can also be described as a weighted sum of all possible normal modes satisfying the local boundary conditions.

If the propagation characteristics of all normal modes are known, the actual physical wavefront and its propagation in space as well as its periodic fluctuation in time can be described as a superposition of basic shapes. This method is referred to as modal decomposition.

The first normal mode is the plane wave. If we truncate after the first mode we obtain the already-derived plane-wave approximation. Higher modes are described by Bessel functions of the first kind of order zero, if the duct is straight (nodal lines must be circles). In tubes with toroidal bends [3.7] wavefronts will also contain normal modes described by higher-order Bessel functions with diameters as nodal lines.

Here we will follow the presentation of *Kemp* [3.8] for the axisymmetric case. In a similar way modal decomposition can also be applied to get a three-dimensional solution within a duct with rectangular cross-section. This is the case in certain wooden organ pipes. For this derivation please refer to [3.8].

3.3.6 Modal Decomposition

In cylindrical coordinates (r, θ, z) the Laplacian operator Δ becomes

$$\Delta = \Delta_\perp + \frac{\partial^2}{\partial z^2} = \left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2}\right) + \left(\frac{\partial^2}{\partial z^2}\right). \quad (3.72)$$

If we express pressure p and axial velocity v_z using infinite series according to [3.9] we get the solutions

$$p(r, \theta, z, t) = \sum_i P_i(z)\Psi_i(r, \theta)\exp(\mathrm{i}\omega t) \quad (3.73)$$

and

$$v_z(r, \theta, z, t) = \frac{1}{S(z)}\sum_i U_i(z)\Psi_i(r, \theta)\exp(\mathrm{i}\omega t)\,, \quad (3.74)$$

with pressure profile P_i and velocity profile U_i of the i-th mode along the z-axis and the profile Ψ_i orthogonal to it. The complex-valued profiles contain the spatial distribution of oscillation amplitudes and their initial phases at time $t = 0$. $S(z)$ describes the geometry of the acoustical duct by specifying the cross-sectional area along the z-axis. Instrument makers are more acquainted with the so-called bore profile defining bore diameters instead of cross-sectional areas. But as ducts do not necessarily need to be perfectly cylindrical, it is better to base computations on the cross-sectional area.

We know that modal profiles along the z-axis

$$P_i(z) = \exp(\mathrm{i}k_i z), \quad U_i(z) = \exp(\mathrm{i}k_i z)\,, \quad (3.75)$$

are sinusoidal with modal wave numbers k_i. The profiles Ψ_i across the longitudinal axis are the so-called classical eigenfunctions obeying

$$\Delta_\perp \Psi_i = -{\alpha_i}^2 \Psi_i \quad (3.76)$$

with boundary condition $\partial\Psi_i/\partial r = 0$ for $r = R$, R being the radius of the cylindrical duct. The eigenvalues in the axisymmetric case are given by $\alpha_i(z) = \gamma_i/R$ with γ_i being the successive zeros of the Bessel function of order one.

Substituting p from (3.73) into the wave (3.51), resolving the differential operator Δ from (3.72), dividing by p and substituting P_i and $\Delta_\perp\Psi_i$ from (3.75) and (3.76) we can get rid of all partial derivatives and obtain the relation

$$k_i^2 = k^2 - \alpha_i^2\,, \quad (3.77)$$

with the free space wave number $k = \omega/c$ and and modal wave lengths $\lambda_i = 2\pi/k_i$ being a function of modal wave numbers k_i.

This leads us to the interpretation that the propagation of the different modes along a cylindrical piece of duct does not differ from the plane wave case, described by (2.42) and (2.30), yet each mode having got its own wave number k_i and its own characteristic impedance $Z_{c,i} = \pm k\rho c/(k_i S)$. If the wave number k_i becomes imaginary (for $k < \alpha_i$), exponential damping is observed.

Equations (2.45) and (2.46), the modal projection equations from plane 1 to plane 0, a distance d away, can therefore be rewritten for the multimodal case as

$$P_{0,i} = \cos(k_i d)P_{1,i} + \mathrm{i}\sin(k_i d)Z_{c,i}U_{1,i} \quad (3.78)$$

and

$$U_{0,i} = i\sin(k_i d)Z_{c,i}^{-1}P_{1,i} + \cos(k_i \mathrm{d})U_{1,i}\,. \quad (3.79)$$

Defining column vectors $\boldsymbol{P}$ and $\boldsymbol{U}$ containing the pressure amplitudes P_i with respect to the volume flow amplitudes U_i of all modes at any cross-sectional plane

of the duct, we can introduce a multimodal impedance matrix $\mathbf{Z}$, with

$$\boldsymbol{P} = \mathbf{Z}\boldsymbol{U}\,, \tag{3.80}$$

which describes the relationship between pressure and volume flow amplitudes of different modes at any given position inside the duct. If the matrix $\mathbf{Z}$ is diagonal then the different modes are uncoupled and no modal conversion takes place. Otherwise energy is transferred from one mode to another, an effect which is of particular interest at discontinuities of the duct.

The projection equations (3.78) and (3.79) can be rewritten in matrix notation as

$$\boldsymbol{P}_{(0)} = \mathbf{D_1}\boldsymbol{P}_{(1)} + \mathbf{D_2}\mathbf{Z}_\mathrm{c}\boldsymbol{U}_{(1)} \tag{3.81}$$

and

$$\boldsymbol{U}_{(0)} = \mathbf{D_2}\mathbf{Z}_\mathrm{c}^{-1}\boldsymbol{P}_{(1)} + \mathbf{D_1}\boldsymbol{U}_{(1)} \tag{3.82}$$

with diagonal matrices $\mathbf{D_1}(i,i) = \cos(k_i d)$, $\mathbf{D_2}(i,i) = \mathrm{i}\sin(k_i d)$ and $\mathbf{Z}_\mathrm{c}(i,i) = k\rho c/(k_i S)$.

If these equations are substituted into (3.80) we can (for the derivation refer to [3.8] obtain the projection equation for impedance matrices

$$\mathbf{Z}_{(0)} = (\mathbf{Z}_{(1)} + \mathrm{i}\mathbf{D}\mathbf{Z}_\mathrm{c})(\mathrm{i}\mathbf{D}\mathbf{Z}_\mathrm{c}^{-1}\mathbf{Z}_{(1)} + \mathbf{I})^{-1} \tag{3.83}$$

with diagonal matrix $\mathbf{D}(i,i) = \tan(k_i d)$.

Assuming axisymmetric pressure distributions only, the solution of the eigenproblem in (3.76) is

$$\Psi_i = \frac{J_0\left(\frac{\gamma_i r}{R}\right)}{J_0(\gamma_i)} \tag{3.84}$$

with J_0 being the Bessel function of the first kind of order zero.

Bruneau et al. studied a multimodal wave propagation model including viscous losses [3.10]. Starting with a lossy boundary condition he obtained complex modal wave numbers

$$k_i = \pm\sqrt{k^2 - \left(\frac{\gamma_i}{R}\right)^2 + \left(\frac{2k}{R}\right)(\mathrm{Im}\,(\epsilon_i) - \mathrm{iRe}\,(\epsilon_i))} \tag{3.85}$$

with

$$\epsilon_i = \left(\frac{1-(\gamma_i)^2}{(k^2R^2)}\right)\epsilon_\mathrm{v} + \epsilon_\mathrm{t} \tag{3.86}$$

with $\epsilon_\mathrm{v} = (1+\mathrm{i})2.03\times10^{-5}\sqrt{f}$ and $\epsilon_\mathrm{t} = (1+\mathrm{i})0.95\times10^{-5}\sqrt{f}$ (simplified, refer to [3.8, 10].

3.3.7 Modal Conversion

While different modes in a uniform duct propagate independently of each other, mode conversion takes place where the duct's cross-sectional area changes. The i-th mode after the discontinuity will be composed from a weighted sum of all modes before the discontinuity. The pressure and volume velocity mode amplitude vectors $\boldsymbol{P}$ and $\boldsymbol{U}$ at one side of the discontinuity (cross-section S_0) are related to the corresponding vectors at the other side (cross-section S_1) by

$$\boldsymbol{P_0} = \mathbf{F}\boldsymbol{P_1}, \quad \boldsymbol{U_1} = \mathbf{F}^\top\boldsymbol{U_0}, \quad S_0 < S_1 \tag{3.87}$$

and

$$\boldsymbol{P_1} = \mathbf{V}\boldsymbol{P_0}, \quad \boldsymbol{U_0} = \mathbf{V}^\top\boldsymbol{U_1}, \quad S_0 > S_1\,. \tag{3.88}$$

The matrices $\mathbf{F}$ and $\mathbf{V}$ for cylindrical and rectangular ducts have been derived by *Kemp* in [3.8]. The results for the cylindrical case are repeated here as

$$F_{n,m}(\beta) = \frac{2\beta\gamma_m J_1(\beta\gamma_m)}{(\beta^2{\gamma_m}^2 - {\gamma_n}^2)J_0(\gamma_m)}\,, \tag{3.89}$$

with $\beta = R_1/R_2$ and $F(0,0) = 1$ and $V_{n,m}(\beta) = F_{n,m}(1/\beta)$.

From the equations (3.87) and (3.88) the projection of the impedance matrix across a discontinuity can be calculated as

$$\mathbf{Z}_0 = \mathbf{F}\mathbf{Z}_1\mathbf{F}^\top, \quad S_0 < S_1\,, \tag{3.90}$$

and

$$\mathbf{Z}_0 = \mathbf{V}^{-1}\mathbf{Z}_1(\mathbf{V}^\top)^{-1}, \quad S_0 > S_1\,. \tag{3.91}$$

3.3.8 Multimodal Radiation

In [3.11], *Zorumski* published a numerically applicable multimodal solution for the radiation impedance of a cylindrical pipe terminated in an infinite flange.

Multimodal wave propagation in axisymmetric or rectangular ducts has superbly been reviewed and extended by *Kemp* in [3.8]. Important studies of multimodal wave propagation have been published by *Pagneux* [3.9] and *Amir* [3.12].

The multimodal result for the radiation impedance of a circular opening in an infinite baffle is repeated here

$$Z_{n,m} = \frac{\rho c}{S}\int_0^{\frac{\pi}{2}} \sin\phi D_n(\sin\phi)D_m(\sin\phi)\,\mathrm{d}\phi + \frac{\mathrm{i}\rho c}{S}\int_0^{\infty} \cosh\xi D_n(\cosh\xi)D_m(\cosh\xi)\,\mathrm{d}\xi\,, \tag{3.92}$$

where

$$D_i(\tau) = \frac{-\sqrt{2}\tau J_1(\tau kR)}{\left(\frac{\gamma_i}{(kR)}\right)^2 - \tau^2} \,. \tag{3.93}$$

In a study published by *Hélie* and *Rodet* in [3.13] the multimodal radiation impedance was calculated for a pulsating portion of a sphere without any solid wall or object other than the rest of the sphere, which remains motionless.

The resulting expression for the impedance Z as a function of frequency f, angle of the axis of symmetry θ and mode number n is

$$Z_n(f,\theta) = -\mathrm{i}Z_c\gamma_n\left(\frac{2\pi f r_0}{c_0}\right)\mu_n(\theta_0)\mathrm{P}_n(\cos\theta)\,, \tag{3.94}$$

where

$$\mu_n(\theta_0) = \frac{\mathrm{P}_{n-1}(\theta_0) - \mathrm{P}_{n+1}(\theta_0)}{2} \tag{3.95}$$

with r_0 being the radius of the sphere, $Z_c = \rho_0 c_0$ being the characteristic specific impedance, ρ_0 and c_0 the mass density and speed of sound, θ_0 being the maximum opening angle θ at the edge of the pulsating portion of the sphere, P_n being the Legendre polynomials and $\gamma_n(z) = h_n(z)/h'(z)$ with h_n representing the outgoing spherical Hankel functions.

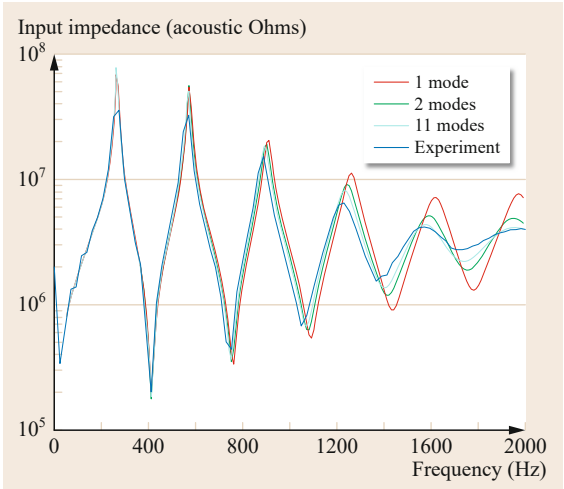


Fig. 3.6 Trumpet section input impedance as calculated by *Kemp* (after [3.8])

Taking multimodal propagation into account can improve modeling of musical instruments significantly. Even below the cutoff frequency, modal conversion makes quite a notable difference as shown in Fig. 3.6. Comparisons with numerical finite element method (FEM) simulations made by *Amir*, *Pagneux* and *Kergomard* [3.9, 12] show good agreement.

References

3.1 N.H. Fletcher, T.D. Rossing: *The Physics of Musical Instruments*, 2nd edn. (Springer, New York 1990)

3.2 P. Hagedorn: Mechanical oscillations. In: *Mechanics of Musical Instruments*, Courses and Lectures/International Centre for Mechanical Sciences, Vol. 355, ed. by A. Hirschberg, J. Kergomard, G. Weinreich (Springer, Wien 1995) pp. 7–78

3.3 P.M. Morse, U. Ingard: *Theoretical Acoustics* (Princeton Univ. Press, New Jersey 1987)

3.4 M. Heckl: Vibrations of one- and two-dimensional continuous systems. In: *Handbook of Acoustics*, ed. by M.J. Crocker (Wiley, New York 1998) pp. 579–596

3.5 H. von Helmholtz: *Die Lehre von den Tonempfindungen als Physiologische Grundlage Für die Theorie der Musik* (F. Vieweg, Braunschweig 1863)

3.6 H. Levine, J. Schwinger: On the radiation of sound from an unflanged circular pipe, Phys. Rev. **73**, 383–406 (1948)

3.7 C.P.A. Braden: *Bore Optimisation and Impedance Modelling of Brass Musical Instruments*, Ph.D. Thesis (Univ. Edinburgh, Edinburgh 2006)

3.8 J.A. Kemp: *Theoretical and Experimental Study of Wave Propagation in Brass Musical Instruments*, Ph.D. Thesis (Univ. Edinburgh, Edinburgh 2002)

3.9 V. Pagneux, N. Amir, J. Kergomard: A study of wave propagation in varying cross-section waveguides by modal decomposition. Part I. Theory and validation, J. Acoust. Soc. Am. (JASA) **100**(4), 2034–2048 (1996)

3.10 A.M. Bruneau, M. Bruneau, P. Herzog, J. Kergomard: Boundary layer attenuation of higher order modes in waveguides, J. Sound Vib. **119**(1), 15–27 (1987)

3.11 W.E. Zorumski: Generalized radiation impedances and reflection coefficients of circular and annular ducts, J. Acoust. Soc. Am. (JASA) **54**(6), 1667–1673 (1973)

3.12 N. Amir, V. Pagneux, J. Kergomard: A study of wave propagation in varying cross-section waveguides by modal decomposition. Part II. Results, J. Acoust. Soc. Am. (JASA) **101**(5), 2504–2517 (1997)

3.13 T. Hélie, X. Rodet: Radiation of a pulsating portion of a sphere: Application to horn radiation, Acustica **89**(4), 565–577 (2003)

4. Construction of Wooden Musical Instruments

Chris Waltham, Shigeru Yoshikawa

This work aims to provide an overview of why and how wood is used in musical instruments, primarily strings, woodwind and percussion. The introduction is a description of the desirable properties of a musical instrument and how these relate to the physical properties of wood. A summary is given of the most important woods mentioned in this chapter, including common and Latin names. Section 4.2 discusses the physical properties of woods most relevant to musical instruments and how they relate to their biological taxonomy and also to organology. Sections 4.3 and 4.4 are devoted respectively to woods that make up the acoustically radiant parts of instruments (tonewoods), and those whose function is to transmit vibrations from one part to another, or are simply structural (framewoods). Section 4.5 deals with how the wood is selected, prepared, assembled into an instrument, and finished.

4.1 Scope

This chapter describes the construction of wooden instruments from the point of view of a physical scientist or mechanical engineer, emphasizing the role played by the material properties of wood. While the authors have tried to include in the discussion as geographically wide a variety of instruments as possible, their experience and backgrounds have naturally led them to concentrate on East Asian and Western instruments. The rich variety of string instruments from the rest of Asia, Africa and elsewhere, will have to await a future study.

4.1.1 General Physical Properties of Musical Instruments

Power Output

If we take a comfortable sound pressure level to be 60 dB and ask what isotropic acoustic power source will provide this at a distance of 10 m, the result is of the order 1 mW. This is typical of the normal range of a trained human voice [4.1, p. 622]. For brief periods, music may reasonably rise to 90 dB at 10 m, but this seldom happens with a single instrument.

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_4

Section 4.3.3 shows how a typical musical *box* i. e., a guitar, can radiate 60 dB at 10 m over most of its frequency range with an application of a reasonable force of 1 N rms at the bridge. Admittance measurements show that the displacement of the bridge is of the order of 1 μm at 440 Hz, which is structurally quite acceptable for a 2.5 mm spruce plate.

Frequency and Temporal Response

The human ear can nominally respond to frequencies between 20 Hz and 20 kHz. However, in practice our sensitivity is very low at either end of the spectrum and sounds of musical importance tend to lie between about 100 Hz and 5 kHz. Many instruments have the fundamentals of lowest notes below 100 Hz, but we tend to hear only the overtones and *hear* the fundamental only by inference.

The upper limit for musical *speed*, even in electronic music where there are no restrictions on the ability of the player, is about 20 notes per second. Thirty-second notes played at $\text{♩} = 120$ implies a rate of 16 Hz, so each note lasts 60 ms. Hence the ability to define the pitch of these notes is $\Delta f \approx 16$ Hz, which is $\pm$ half a semitone at middle C (262 Hz). In other words, faster notes would have no well-defined pitch.

The *speed limit* puts a limit on the maximum Q of the completed instrument; $Q = \pi f_0 \tau$, where τ is the decay time. For τ of 0.06 s and middle C (262 Hz), $Q \lesssim 50$. For the wooden boxes of string instruments, a Q of 20 or 30 is typical and the need for resonances to be close enough together not to leave big gaps in the radiation gives a similar restriction (Sect. 4.3.3).

4.1.2 Why Wood?

In spite of all recent advances in materials science, wood remains the construction material of choice for the majority of all musical instruments worldwide. Table 4.1 lists all the material properties discussed in this chapter, and gives values for three woods common in musical instruments (a softwood, a hardwood and a monocot) and compares them to a plastic and a metal (acrylic and aluminum). Some distinguishing features of woods and metals can immediately be seen, and the reasons for and consequences of these numerical differences will be examined.

Let us briefly consider why wood is selected as the material for string instruments. Wood was certainly one of the few options available to builders in preindustrial times, but despite extensive research into alternatives in the last hundred years or so, it shows no sign of being displaced. In any case, material scientists are looking to improve secondary characteristics like durability and uniformity while reproducing the primary vibrational properties of, typically, Sitka spruce [4.4]. A look at the Table 4.1 gives some clues. The first is weight; the mass of a given plate scales as ρ/c [4.5], so a 70 g spruce violin top would have to be replaced by a 440 g acrylic top or a 400 g aluminum one, either of which would make for a very heavy instrument. The second parameter that stands out is the quality factor; acrylic is much less resonant than most woods, and aluminum much more resonant. The basic considerations outlined in Sect. 4.1 indicate the desirability of the intermediate quality factors of wood, and this is explored later in Sect. 4.3.3.

In Table 4.1 Sitka spruce (used for the top plate of the violin and guitar) and amboyna wood (used for the body and neck of the shamisen, a Japanese three-stringed instrument) are shown as representatives of tonewoods and framewoods respectively. Bamboo, particularly Japanese madake, is traditionally used for the body of the shakuhachi, nohkan (noh flute), shoh, hichiriki, and other Asian woodwind and percussion instruments.

4.1.3 Summary of Woods

The nomenclature of woods is often confusing. Table 4.2 lists all the woods discussed in this chapter, along with binomial and other common names, and examples of musical uses.

Table 4.1 Vibroacoustic properties of wood, plastic and metal bars (after [4.2, 3]). For orthotropic wood, E refers to E_L, G refers to G_{LR}, and c to c_L i. e., properties in the direction of the grain

Property	Sitka spruce	Amboyna	Bamboo	Acrylic	Aluminum
Density ρ (kg/m^3)	470	870	700	1200	2700
Young's modulus E (GPa)	12	20	15	5.3	71
Shear modulus G (GPa)	1.1	1.6	1.3	1.9	27
Elastic modulus ratio (E/G)	11	12.5	11.5	2.8	2.6
Quality factor (Q)	131	155	140	17	980
Sound speed (in bar) $c = \sqrt{E/\rho}$ (m/s)	5100	4800	4600	2100	5130
Wave resistance ρc (MPa s/m)	2.4	4.2	3.2	2.5	13.8
Vibration parameter c/ρ (m^4/kg/s)	11	5.5	6.6	1.75	1.9
Transmission parameter cQ (10^5 m/s)	6.7	7.4	6.4	0.36	50.3
Acoustic conversion efficiency cQ/ρ	1420	855	920	30	1860

Table 4.2 Summary of woods discussed and their nomenclature. H = hardwood, S = softwood, M = monocot

Common name	Type	Binomial name	Other names	Native	Musical use
Amboyna	H	*Pterocarpus indicus*	Padauk, rosewood	Southeast Asia	Shamisen body, marimba bars
Bamboo (Japanese)	M	*Phyllostachys bambusoides*	Madake	Japan	Shakuhachi etc.
Boxwood	H	*Buxus sempervirens*	Tsuge		Biwa plectrum, marimba mallet
Cypress (Japanese)	S	*Chamaecyparis obtusa*	Hinoke	Central Japan	Stage floor, Noh
Ebony	H	*Diospyros*	Kokutan	West Africa, South Asia	Clarinet
Grenadilla	H	*Dalbergia melanoxylum*	African blackwood	Africa, dry	Clarinet, oboe
Maple (Japan)	H	*Acer palmatum*	Kaede	Japan	
Maple (Norway)	H	*Acer platanoides*		Temperate zone	Violin family back, ribs
Maple (Sycamore)	H	*Acer pseudoplatanus*		Central Europe, Southwest Asia	
Magnolia (Japan)	H	*Magnolia obovata*	Hohnoki	Japan	Biwa fret
Mulberry	H	*Morus alba*	Kuwa	East Asia	Biwa body
Padauk	H	*Pterocarpus soyauxii*		Africa	Marimba bars
Paulownia	H	*Paulownia tomentosa*	Tung, kiri	China	Eastern soundboards
Pernambuco	H	*Guilandia ecbinata*		Brazil	Violin bow
Rosewood (Brazil)	H	*Dalbergia nigra*	Jacaranda	Brazil	Guitar back, ribs, marimba bars
Spruce (Norway)	S	*Picea abies*		Northern temperate, high	Western soundboards
Spruce (Sitka)	S	*Picea sitchensis*		Canada pacific coast	Western soundboards
Zelkova (Japan)	H	*Zelkova serrata*	Keyaki	E.Asia	Wa-daiko

4.2 Physical Properties of Wood

4.2.1 Stiffness, Density, Damping, and Orthotropy

Wood is an orthotropic material; its elastic and strength properties are dependent on the direction of the force with respect to the grain. The natural coordinate system is radial (R), longitudinal (L) and tangential (T). The elastic properties are defined by 12 constants, nine of which are independent: three moduli of elasticity (E_i), three shear moduli ($G_{ij}, i \neq j$), and six Poisson ratios μ_{ij} (related by $\mu_{ij}/E_i = \mu_{ji}/E_j$ for $i \neq j$) [4.6, Chap. 5-2]. Elastic moduli have been shown to be constant enough in the audio frequency range to be characterized by single numbers [4.7]; however they do vary from sample to sample and also with moisture content (and therefore age).

Wood for Western instruments is invariably quarter-sawn (in contrast to some Asian instruments, see Sect. 4.5.2) i. e., the face of the instrument is the radial plane. This gives the largest bending strength of any orientation, as E_T is usually the smallest of the elastic moduli.

Tabulations of Q for wooden bars in the L-direction exist in a venerable work by *Barducci* and *Pasqualini* [4.8] and more recently by *Bucur* [4.9]. Details of how to make a full measurement in all orientations are given by *McIntyre* and *Woodhouse* [4.10], including how to average orthonormal values for complex mode shapes involving movement in both directions. Norway Spruce is given as an example, in which case, Q in the L-direction was 140, while that in the R-direction was in the 40s.

Although often quoted to three significant figures, Q varies with sample and also with frequency. The loss factor is often assumed to be independent of frequency [4.5], but in a 1950 paper, *Fukada* [4.7] showed that woods from conifers often displayed a maximum Q at around 1 kHz, dropping off by tens of percent below 200 Hz or above 3 kHz, while some hardwoods showed a minimum value of Q in the few kHz range. The quality factor Q measures the dissipation of vibrational energy by the internal friction of material. Since cellulosic microfibril is highly crystalline in tone woods, they have high Q in general.

We will now turn to a discussion of the classification of wood by cellular structure i. e., softwood, hardwood, monocot. As it happens, these cellular distinctions mostly, but not always, coincide with the distinction between tone and frame woods. Notable exceptions are paulownia, a *soft* hardwood, widely used in Asia for string instrument soundboards, and yew, a *hard* softwood, once used for harp frames. Also included here is a brief discussion of plywoods, whose use is not confined to cheap student instruments.

4.2.2 Classification Wood by Cellular Structure

Softwoods

Softwoods come from a group of seed-producing trees called gymnosperms (which are most often, for the purposes of our discussion, conifers). They are the most primitive of woods and have the simplest cellular structure. The vast majority of cells, the tracheids, run in the longitudinal direction, with the occasional resin canal running in the radial direction [4.11, p. 14]. This structure gives softwoods a markedly anisotropic elasticity, with E_L/E_R typically 10 and E_L/E_T typically 25 [4.6, Chap. 5-2]. The quality factor in the radial direction is often very much smaller than that in the longitudinal.

Hardwoods

Hardwoods come from a group of flowering trees called angiosperms. Their longitudinal cells come in two main forms, the larger pores and smaller fibers, and radial ray cells are much more abundant than in softwoods [4.11, p. 14]. This structure gives hardwoods a less anisotropic elasticity than softwoods, with E_L/E_R typically 7 and E_L/E_T typically 20 [4.6, Chap. 5-2].

Composite Woods

The most common composite wood used in musical instruments is plywood. This layered composite consists of sheets of wood with different grain orientations and it therefore has more isotropic properties than single pieces of wood. Plywood is used for the soundboards of small harps (where strength is an issue due to the tension of the strings), some of high quality, and the backs of cheaper guzhengs.

The veneer on spruce harp soundboards may be considered as a form of *plywood* as the veneer grain runs longitudinally i. e., perpendicular to the transverse soundboard grain. This orientation is to prevent cracking of the glue joints in the multipiece soundboard. The effect is to reduce the longitudinal-radial anisotropy of the spruce by a factor of three [4.12] i. e., E_L/E_R from 12 to 4. In addition the quantity of glue used in the bonding needs to be minimized, as a typical $100\,\mathrm{g/m^2}$ layer ($Q \approx 5$) will decrease the overall Q of the soundboard by 25% [4.13].

Monocots

Monocots also come from flowering trees, and the only variety significant in musical instrument making is bamboo, which is widely used in Asia for the tubular body of woodwinds and for the vibrating plates of percussions. The bamboo culm is divided quasiperiodically by nodes, one of which forms a solid cross wall and is called the diaphragm. The internodes have a culm wall surrounding a large hollow cavity, called a lacuna. The culm consists of about 50% parenchyma, 40% fibers, and 10% conducting cells (vessels and sieve tubes) on average [4.14]. The polylamellated wall structure of the culm fibers (vascular bundle sheaths) results in an extremely high tensile strength. Bamboo culms have no cambium, and no radial cell elements (such as rays in trees) exist in the internodes. Bamboo is therefore quite different from other woods (although both natural materials have orthotropic properties) and also from isotropic materials such as metals and man-made polymers [4.14]. Similarly to wood, the longitudinal axis of bamboo is parallel to the fibers along the culm, the radial axis is perpendicular to the culm circumference, and the tangential axis is along the culm circumference.

4.2.3 Classification of Acoustic Woods

The most acoustically important properties of wood appear to be the elastic constant E_L, the density ρ, the anisotropy E_L/E_R and some measure of the damping, e.g., Q. Several ways of combining these quantities have been proposed to select acoustically useful woods, and to distinguish between woods best suited to soundboards (*tonewoods*), ribs and backs (*frame-woods*), bows, woodwinds etc.

Consider the material properties listed in Table 4.1. The longitudinal sound speed is almost the same in dry woods, bamboo, and even metals, which implies a strong correlation between E_L and ρ. *Wegst* plots E_L versus the loss coefficient and finds no clear correlation between these two quantities [4.15, Fig. 5]. However, a group of woods with high E_L (20–40 GPa) includes those favored for violin bows, and if the quality factor is high enough, woods for xylophone bars. Soundbox woods appear at lower E_L (centered around 12 GPa), with those for soundboards having low loss and those for ribs/backs higher loss. Woods favored for wind instruments are not distinguished from soundboard woods on this plot.

Several authors starting with *Schelleng* [4.5] have discussed the importance of the ratio c/ρ, variously called the *vibration parameter* [4.2], or since the vibration of a wood plate radiates sound, the *radiation ratio* [4.16, 17] and *sound radiation coefficient* [4.15]. Schelleng derived c/ρ by supposing that both the stiffness and the inertia of two plates should be the same if their vibrational properties are to be the same. The ratio c/ρ for Sitka spruce is six times higher than that for aluminum, and twice that for amboyna, which is typical of the ratio between tone and framewoods [4.15].

The low density of woods and bamboo compared to metals gives a low wave resistance ρc and a high vibration parameter. A lower wave resistance means a better coupling to the surrounding air. The low wave resistance of Sitka spruce (and its European counterparts) 2.4 MPa s/m makes it ubiquitous in the soundboards of Western string instruments. Paulownia has an even lower value [4.2] (1.4 MPa s/m) and is common in Asian instruments. The higher ρc of amboyna wood causes higher reflection of sound within the instrument body, suggesting it as a good frame material in string instruments.

The acoustic conversion efficiency (ACE) proposed by *Yankovskii* [4.25] has also been used to characterize acoustic materials [4.22, 26]. This ACE is the ratio of acoustic energy radiated from a beam to the vibration energy of the beam and is proportional to cQ/ρ. However, cQ and ρ are fairly independent of each other for a wide variety of woods, so one of the current authors (Shigeru Yoshikawa) proposed using cQ (see (4.7) in the Appendix for physical meaning) as a measure for distinguishing tone and frame woods [4.2].

In Fig. 4.1 cQ is plotted against ρ/c, the reciprocal of the vibration parameter, which may be called the *antivibration parameter*. A clear separation is seen between tonewoods (paulownia, Sitka spruce, and Norway spruce) and framewoods (Brazilian/Rio rosewood, amboyna wood, Japanese and Norway maple). Also included in Fig. 4.1 are a selection of other woods: mulberry (Japanese kuwa, Morus alba) for the best quality Japanese lutes (Satsuma biwa) [4.2, 20]; ebony [4.18] and grenadilla [4.21], which are used for the best quality clarinet bodies; pernambuco [4.15, 16, 19] for violin bows; zelkova [4.20, 27] for the shells of Japanese big drums (wa-daiko); bamboo for shakuhachi bodies [4.15, 23, 24]; and Japanese cypress [4.27] for the stage floor for Noh plays (the percussive sound of the player's stamping feet is an important element in these plays). These woods have different physical properties that cannot be explained well by the quality criteria for string instrument woods, though Japanese cypress and zelkova can be excellent substitutes for Sitka spruce and Japanese maple respectively. Mulberry, ebony and grenadilla are very heavy and hard, and show very low vibration transmission. Pernambuco is also heavy, but shows very high vibration transmission. Bamboo shows characteristics in between tonewoods and framewoods, and has been proposed for several applications to musical instruments [4.15].

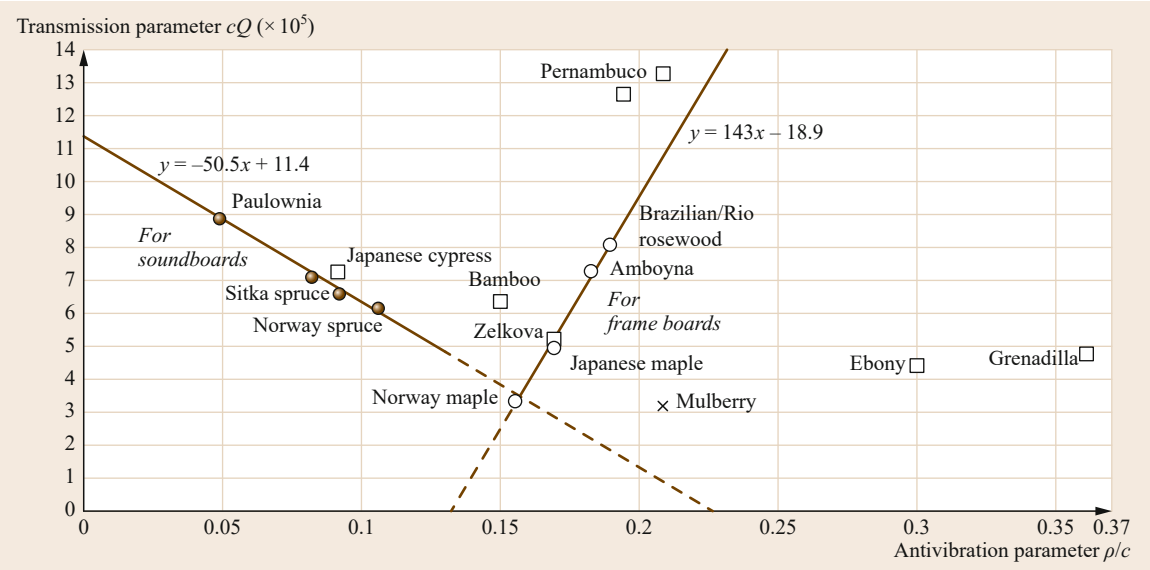


Fig. 4.1 Classification diagram of traditional woods for string instruments and for other instruments. •: tonewoods for the soundboard; ○: framewoods; ×: traditional wood for the Satsuma biwa; □: traditional woods for other instruments (after [4.2, 15, 18–24])

4.3 Tonewoods

The purpose of tonewoods is to radiate sound. In this section we will examine acoustic radiation from vibrating wooden bars, plates and boxes.

4.3.1 Bars (Idiophones)

The most straightforward uses of tonewoods are in the idiophone family. Bars of wood, not naturally harmonic, are shaped to have an overtone structure with a definite pitch [4.28]. The bars serve as their own radiator, sometimes with the assistance of sympathetically resonating tubes, as in the marimba.

Wegst's classification scheme [4.15] shows woods favored for idiophones have higher E, higher ρ and consequently higher Q than woods used for soundboards. The reason for this is partly because the input energy comes from a single impulse and is not sustained by a string, and partly because the wood has to be hard to avoid damage from the mallet. As the sound comes from weakly radiating bending modes the bars have to be struck sharply. Typical woods for the bars are rosewoods (genus *dalbergia*) and padauk, while the very hard boxwood makes for good mallets.

Bork [4.29] gives the following expression for T_{60}, the time it takes a sound to decay to a level of 60 dB below its peak

$$T_{60} = \frac{2.2}{\eta f} \,. \tag{4.1}$$

A typical loss value for a marimba bar $\eta \approx 0.005$ so a well-suspended middle-C bar will have $T_{60} \approx 2$ s.

4.3.2 Plates

Few musical instruments have simple plates as their acoustic radiators. A notable exception is the piano which has a two-dimensional soundboard that is large enough to avoid acoustic short-circuiting between the two sides. However, the behavior of plates is an important step to understanding soundboxes.

In Fig. 4.2, tap tones from sample plates of Sitka spruce, maple and aluminum are analyzed in the frequency domain in 1/3 octave bands [4.3]. The ordinate P_m indicates the maximum pressure level in each 1/3 octave band from 25 Hz to 20 kHz. Several peaks and troughs of P_m depend on the sample size. Although Sitka spruce and maple show the strongest response around 1 kHz and similar response below 1 kHz, the response of Sitka spruce is much weaker than that of maple above about 3 kHz. This weaker radiation of Sitka spruce in higher frequencies seems to be preferred as tonewood because the auditory emphasis of lower frequencies is desirable to the pitch sensation in Western music. In contrast to Sitka spruce and maple, aluminum shows much weaker response below 1 kHz and much stronger response above 3 kHz. This characteristic of metal is generally unwelcome in Asian and Western music.

A large difference of radiation characteristics between wood and metal at higher frequencies (above around 2 kHz in Fig. 4.2) is attributed to the relative strength of shear and bending deformations in flexural vibrations. As the frequency of the flexural vibration increases, the shear deformation component in the flexural deformation increases in wood, but the bending deformation component is still dominant in isotropic materials like metals. The shear effect, which is indicated by higher values of elastic modulus ratio E/G ($= E_L/G_{LR}$ in wood) in Table 4.1, causes an appreciable increase of loss factor η in higher frequencies [4.3]. Although E/G of the Sitka spruce example shown in Table 4.1 is lower than that of the amboyna wood example, Sitka spruce can often have a much higher value (> 15) than that of amboyna wood (which can be < 8) [4.3]. The stronger shear effect of Sitka spruce than in maple and aluminum seems to be relevant to sound radiation from the soundboard because its low-pass filter effect with a cutoff frequency of about 2 kHz tends to lend the radiated sound a desired softness [4.30].

Although it is difficult to determine the frequency spectrum of plate radiation, the frequency f_B (in Hz)

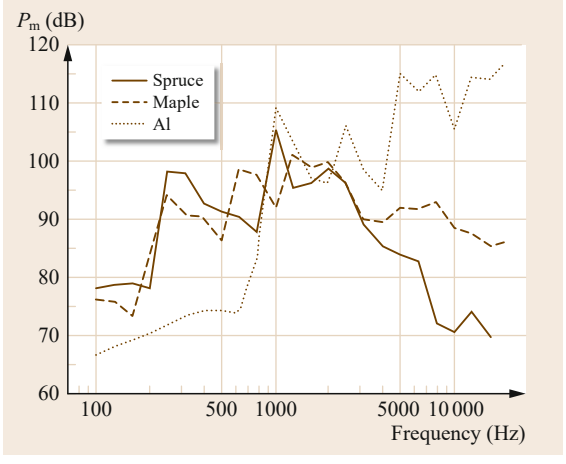


Fig. 4.2 Acoustic radiation characteristics of tap tones from the sample plates ($105 \times 105 \times 2$ mm Sitka spruce and maple) and bar ($105 \times 16 \times 2$ mm aluminum). Peak sound pressure levels are shown in 1/3-octave bands (after [4.3, Fig. 12])

of a bending wave in a thin uniform plate is given by [4.31]

$$f_B = 0.0459\, h c_L k_n^2 \tag{4.2}$$

$$c_L = \sqrt{\frac{E_L}{\rho(1-\nu^2)}}\,, \tag{4.3}$$

where h is the plate thickness, c_L is the longitudinal wave velocity in an infinite plate, ν is Poisson's ratio, and k_n is the wave number corresponding to the normal modes of vibrations that depend on the boundary conditions of the plate. Therefore, h (as well as ρ and E) is important to determine the vibrational properties of woods, particularly of tonewoods. The thinness of the top plate in the violin, cello, and guitar largely contributes to decrease their resonance frequencies [4.20]. Also, the mechanical input impedance Z_m of a thin infinite plate (the ratio of force to the normal velocity of the plate at the driving point) depends on the plate thickness as well as the characteristic impedance ρc. For bending waves in an infinite plate, Z_m is surprisingly a real, frequency-independent constant [4.32]

$$Z_m = \frac{4}{\sqrt{3(1-\nu^2)}}\rho c h^2\,. \tag{4.4}$$

Plate thickness is important to meet two conflicting requirements. Sufficient vibratory energy must be transmitted from the string to the soundboard to produce an audible tone, but at the same time, the energy should not be transmitted so rapidly that the string vibration dies down quickly producing an unmusical *thud* [4.33]. In order to achieve high sound quality, the mechanical impedances of the strings and the soundboard must be controlled very carefully in string instruments [4.15]. In the violin, a bridge stands on the top plate and forms a termination of the strings. Since the bridge is usually made of maple, its impedance is considerably higher than the top plate impedance. However, a careful design of bridge geometry (pattern, shape, and thickness) realizes a desirable impedance range of the bridge that joins the strings and the top plate to create excellent sound quality of the violin. At the same time, very subtle thickness adjustments of the top and back plates are essential to create the tonal excellence as easily inferred from (4.4).

4.3.3 Boxes (String Instruments)

Although vibrating strings are necessary to define the fundamental frequency and overtone series of an instrument, the acoustic power output of a vibrating string itself is far too low to be of any practical use. Hence string instruments need a radiator, a soundbox, which usually takes the form of a thin wooden shell with a hole or holes.

Any solid structure has resonant frequencies of its own, and a string instruments radiation is a product of the string modes and the modes of the soundbox. These modes should be spaced close enough together to allow all possible string frequencies to radiate. There are particular constraints of the quality factors of the box; the damping should be low enough that the box can resonate, but no smaller than that which would give a 3 dB bandwidth of about one semitone (the interval of which is about 6% in frequency; $Q \approx 1/0.06 = 17$). Tonewoods have natural Qs well in excess of 20, even in the radial direction (Sect. 4.2), leaving room for other wall-surface losses and, of course, radiation loss to depress the overall value. It is important to note that often the only Q-factor quoted in the literature is for bending waves traveling in the longitudinal direction of the grain, whereas the losses of two-dimensional (2-D) plates are often dominated by the Q-factor in the radial direction, which is typically a factor of four lower.

Boxes provide fairly regularly spaced modes. For a plate of area S, thickness h and sound speeds c_x and c_y in each of the orthonormal directions, the mean frequency interval Δf is as follows [4.34, p. 587]

$$\Delta f \approx \frac{\sqrt{c_x c_y h}}{1.5S}\,. \tag{4.5}$$

This equation yields values of 73 and 108 Hz for violin top and back plate dimensions respectively. For real violin plates, the bass end of the frequency spectrum is *blocked* by the curvature [4.34], which makes for a larger spacing. The bracing of a guitar top plate has the same function.

Firth [4.35] notes that for the clarsach, a small Scottish harp, the modes of the bare soundboard are about 50 Hz apart. This spacing grows to about 100 Hz with the addition of the stiffening string bar (which goes a long way to removing the effect of the anisotropy of the soundboard wood), and then to almost 200 Hz for the completed soundbox. Comparable values were found for a modern concert harp soundbox [4.36].

Mechanical Efficiency

The basic scaling laws of plate vibrations [4.5] and their radiation efficiency [4.37, p. 156] mean that for a given fundamental plate frequency the power radiated by a plate for a given input force *falls* with plate size. In his study of the guitar, Davis points out that the size of instruments is limited by how thin the plates can be made [4.38], and in the case of spruce this is about 2.5 mm.

It has been noted [4.38] that most string instruments have their two lowest resonances tuned to the fundamental frequency of the lowest string and approximately a factor of two (one octave) above that. For instruments as diverse as the guitar, violin, oud and ruan, these two resonances result from the coupling of the lowest soundboard resonance with the Helmholtz resonance of the air in the soundbox. (Although this is not true of the harp, whose lowest strings are tuned much below the principal resonances of the soundboard [4.36]).

The guitar has a particularly simple low-frequency resonant structure, which makes it a useful instrument to study as it is the low frequencies that most characterize the *voice* of an instrument. The two lowest resonances lie at around 100 and 200 Hz [4.39]. The *air* resonance at 100 Hz is almost the lowest found in Western instruments (except for the double bass, whose *air* resonance is at 60 Hz [4.40]). The structure of these modes can be understood in a simple way: for the lower mode the top plate and the air in the soundhole move in antiphase, while for the higher mode the top plate and the air move in phase (Fig. 4.3). Both modes are significant radiators and they contribute much to the character of the guitar sound. The octave gap between them is large, however, and between these modes, the soundboard admittance drops 50 dB. This would cause a detrimental gap in the radiation between the two resonances if that were the end of the story. However, there is a phase flip between the two modes that prevents total cancellation in regions between them, caused by the fact that at low frequency, if one pushes down on the top plate, air will come out of the soundhole [4.41]. The result is only a 20 dB drop in radiativity (sound pressure at a specified distance divided by input force), one which the ear can accommodate without apparent difficulty.

The other advantage to having a soundhole is that it puts the lowest radiating mode in a frequency range that would otherwise only be accessible with a very thin soundboard [4.38].

The quality factors for the two guitar modes are around 20, and the similarity is noteworthy considering that the balance of losses in the tonehole and viscoelastic losses in the wood are different for the two modes. Models and measurements show that the radiation efficiency (acoustic power radiated divided by mechanical power input at the bridge) at low frequencies is about 30% [4.39]. *Schelleng* [4.42] gives a very similar value for the violin. *Daltrop* et al. [4.36] find Q for the lowest primary modes of a concert harp to be about 20.

At higher frequencies the acoustic efficiency drops off as higher-order modes in the guitar plates cause acoustic short-circuiting that occurs when the wavelength in the plates is much smaller than the wavelength of the radiated sound. However this is not as critical as the low-frequency efficiency. For one thing, human ears work much better at high frequencies than at low. Secondly the density of soundbox resonances tends to be fairly constant in Hz across a wide frequency range – typically a few tens of Hz for a guitar. In contrast the frequency space between musical notes spreads out at high frequencies, so the chance of all the partials of a note *falling between the cracks* of soundbox radiativity is much reduced at higher frequencies.

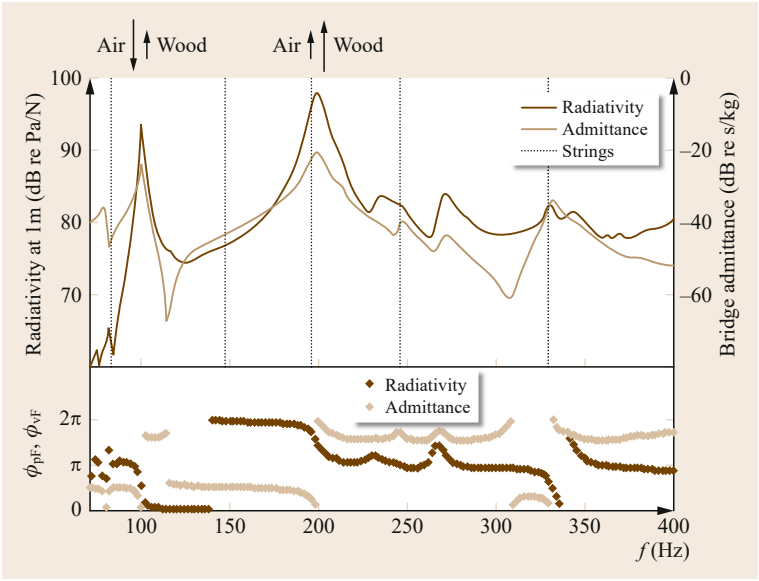


Fig. 4.3 Guitar radiativity and bridge admittance. The principal *air-mode* occurs at 100 Hz; the principal *wood-mode* occurs at 200 Hz. Higher-order plate and cavity modes above 200 Hz have relatively low radiativity compared to the bridge admittance

4.3.4 Examples of Tonewoods

Spruce (Softwood, Tonewood)

There are many different species of spruce; those of most interest musically are Norway spruce from northern Europe and high-altitude central Europe, and Sitka spruce (Fig. 4.4) from the Pacific coast of southern Alaska down to the tip of northern California. Old growth from slow-growing areas is favored for its tight, straight and even grain. Spruce selected for instruments has low density and high Q, and its large anisotropy (E_L/E_R in excess of 10) is particularly important in the low frequency breathing modes (A0, C2 and C3) of the violin family (see note on mode nomenclature in [4.43]). In fact the anisotropy is enhanced by the bass bar, which adds longitudinal stiffness while allowing transverse flexibility. In the case of modern harp soundboards, the anisotropy is less important than the other qualities, and is actually reduced by the addition of veneer and harmonic bars [4.12, 44].

Maple (Hardwood, Tonewood and Framewood)

Maple comes in many different varieties, all of the genus *Acer*. North American maple is often broadly described as *hard* and *soft*. Extreme examples listed in the Forest Products Laboratory Wood Handbook [4.6, Chap. 5-5] are the hard sugar maple (*Acer saccharum*) and the soft silver maple (whose binomial name is confusingly *Acer saccharinum*). Sugar maple has similar properties to the maples shown in Fig 4.1 and silver maple has 75% of the density and 56% of the E_L of sugar maple. Hard varieties are preferred for musical instruments, and their mechanical properties sit at the intersection between tonewoods and framewoods.

As maple sits at the intersection between tone- and framewoods, it is possible to make an entire instrument out of it e.g., a gothic harp [4.45]. However, maple is most renowned for its use in members of the violin family. The material of choice for violin backs and ribs is curly maple, variously known as tiger, flame and most obviously, fiddleback maple (Fig. 4.4). The names do not refer to a distinct species but to the rippled formation of the grain, which makes for singularly beautiful patterns when varnished, patterns that alter with light direction. *Ghemeziu* and *Beldie* [4.9] note that curly maple *Acer pseudoplatanus* has a lower density and higher E_L/E_R ratio than straight-grain *Acer pseudoplatanus*. However the reasons for its choice are aesthetic and not mechanical. It is certainly not chosen for ease of working; the constantly varying grain direction makes planing and carving problematic, as a less-than-perfectly sharp blade will scoop out chunks of wood and ruin the surface.

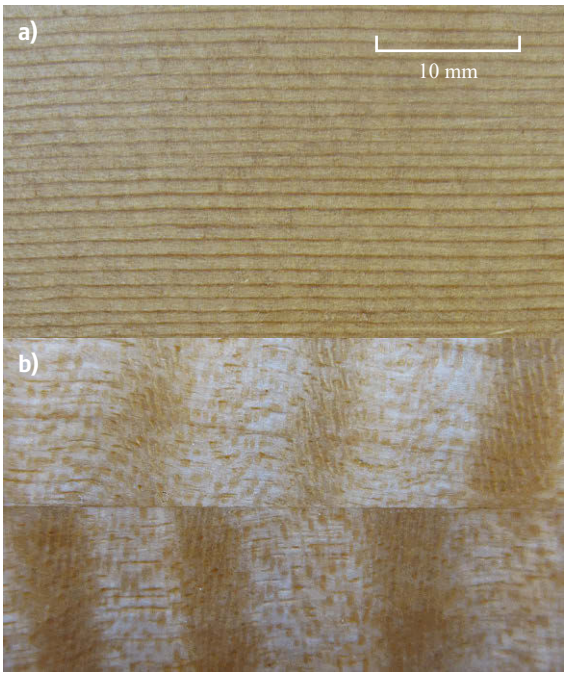


Fig. 4.4a,b Violin tonewoods. **(a)** Straight-grained Sitka spruce for top plate. **(b)** Centerline of matched two-piece curly maple back plate; note the *wavy horizontal grain*, the *vertical rays* and the alternating *light-dark pattern* along and across the *centerline*

Paulownia (Hardwood, Tonewood)

Paulownia (Japanese kiri) is a hardwood species, but it is rather lighter (around $250\,\mathrm{kg/m^3}$) than most softwoods. Its anisotropy is very high ($E_L/E_R \approx 48$, the value of $E_R = 0.134\,\mathrm{GPa}$ was measured at Tsukuba University [4.46]). However, the elastic modulus ratio $E_L/G_{LR} \approx 10{-}12$, less than that for Sitka spruce ($E_L/G_{LR} \approx 20$). A very high value of E_L/E_R may be explained as follows: In general, the E_L of a honeycomb material is proportional to its density while E_R is proportional to ρ^3 [4.47]. Consequently the E_L/E_R ratio should be proportional to ρ^{-2}. Since E_L/E_R of spruce ($\rho \approx 450\,\mathrm{kg/m^3}$) is usually 10–20, the very high E_L/E_R of paulownia becomes understandable when its low density is considered.

Paulownia is widely used in Asian string instruments e.g., the Japanese long zithers, koto and soh and the Chinese guzheng and yueqin (Fig. 4.5). In general, thinner cell wall, that is, lighter wood gives higher dimensional stability. This is because the dimensional change of wood is dominated by the swelling and shrinkage of the wood cell wall (thicker cell wall tends to cause the dimensional change). Therefore, the

Fig. 4.5 Paulownia soundboard of a yueqin

very fine cell wall of paulownia prevents the distortion and warp caused by humidity change. This makes precise and delicate workmanship possible for musical instruments as well as for furniture. In addition, paulownia consists of consecutive small and individual spaces (called the apotracheal parenchyma cells or the tylosis structure), which can yield the smallest density and the dehumidification.

Paulownia can be given a very smooth surface, which makes it suitable for the sliding of bridges (called *ji* in Japanese) on the top plates of kotos, guzhengs and the Korean gayageum when a chord change is required during the performance [4.48]. The koto strings are plucked with small plectra worn on three fingers (the thumb, index and middle) of the right hand. Because this plucking is not so strong and the koto body is very large and long, the material must support vibration well to maintain the sound. Both the highly resonant nature and high anisotropy of paulownia seem to be relevant requirements. In addition, it is observed that harmonics higher than 10 kHz usually appear in koto sounds [4.48]. This may be because small and hard plectrums made of ivory are used for the plucking instead of the player's fingernails. This type of high-frequency emphasis in the koto yields a sharp impression at the attack transient. This tonal characteristic may be supported by low E_L/G_{LR} value (about 10) of paulownia in comparison with high E_L/G_{LR} value (about 20) of Sitka spruce, which produces the high-cut filter effect above 3 kHz (Fig. 4.2). It should be also noted that paulownia is better than spruce in both vibration (radiation) and transmission properties and that paulownia is a material quite the opposite to mulberry used for the Satsuma biwa (Fig. 4.1).

Paulownia was traditionally used to make wooden chests for clothes and safe boxes for valuables. Such chests are now themselves very valuable.

4.4 Framewoods

The way framewoods are used in woodwind, string instruments and membranophones is quite different, so these cases will be considered separately.

4.4.1 Woodwind Instruments

Since the sound of woodwind instruments is produced by the resonance of the air column enclosed by the instrument body, the body wall material is not primarily important from the acoustical viewpoint. Strength, lightness and dimensional stability are obviously necessary however, and makers and players of woodwind instruments have strong preferences for certain woods. For example, bamboo has been used for longitudinal end-blown flutes, Japanese shakuhachi and Chinese xiao (Fig. 4.6), and grenadilla has been used for the clarinet and the oboe.

Bamboo (Monocot, Framewood)

As shown in Table 4.1 and Fig. 4.1, physical and acoustical properties of bamboo are between those of tonewoods and framewoods. However, the properties of bamboo significantly vary from the inside to the outside of the culm wall. For example, the density ranges from about 600 kg/m^3 at the inner regions that contain a small amount of fibers (about 15%) to about 1000–1200 kg/m^3 at the outer surface that contains large amount of fibers (about 60%) [4.15, 23]. The variation of E_L and tensile strength with position in culm wall is almost correlated with that of density. The value of E varies from about 5 GPa at the inside to about 17 GPa at the outside [4.23]. Also, the innermost and outermost surfaces of the culm wall are formed by heavily thickened and lignified parenchyma cells (called the pitch ring) and by the cortex with

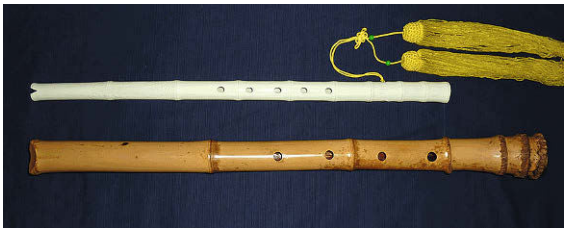


Fig. 4.6 Photos of a Chinese xiao (made of ivory-like material imitating bamboo nodes, 58.6 cm long, E♭4) and a Japanese shakuhachi (made of bamboo called madake, 69.5 cm long, B♭3, manufactured by Johzan Iso (b. 1929))

Fig. 4.7 Violin peg turned out of ebony. Ebony can be machined with a sharp edge, as can be seen in the detailing

a wax-coated layer (called the epidermis) respectively [4.15].

Bamboo is readily available in Asia, and its fully dense and fibrous outermost surface is probably suited for woodwind body and playability. In addition, the thick root end of bamboo, which is used to make the shakuhachi, strengthens the wall rigidity. On the other hand, slender dark-stained bamboo pipes are usually used for free-reed mouth organs (Chinese sheng, Korean saeng, Japanese sho, and Indonesian khaen) [4.49]. The much smaller diameter of these organ bores probably supports the wall rigidity.

Grenadilla, Ebony (Hardwood, Framewood)

The lower right-hand end of Fig. 4.1 (weak vibration generation and transmission, and high ρ/c) indicates woods suited for reed instrument bodies, which are subjected to high acoustic pressures in the bore. Hardwoods such as grenadilla and ebony have been traditionally used for the clarinet and oboe bodies. Physical properties of grenadilla ($\rho = 1280\,\mathrm{kg/m^3}$, $E_\mathrm{L} = 15.9\,\mathrm{GPa}$, and $Q = 137$) [4.21] give very low longitudinal sound speed $c = 3520\,\mathrm{m/s}$. On the other hand, grenadilla shows very high shear modulus $G_\mathrm{LR} = 3.1\,\mathrm{GPa}$ (Table 4.1). Grenadilla trees often grow in very gnarled and twisted shapes, and its grain is frequently interlocked. Such interlocked grain probably yields low sound speed and high shear modulus (low elastic modulus ratio $E_\mathrm{L}/G_\mathrm{LR}$) [4.50].

Ebony has an almost invisible grain, and is easy to machine to fine tolerances and with sharp edges, a feature crucially important for the tonehole of wind instruments. The wood can be polished to a fine finish without varnish (Fig. 4.7). Because it is a hard-wearing wood, ebony also finds use for violin pegs and fingerboards.

4.4.2 String Instruments

The purpose of framewoods in string instruments is primarily to create a box around the top plate (or both plates, where both are prominent radiators as in the case of the violin family, or the guzheng).

Mulberry (Hardwood, Framewood)

The Japanese Satsuma biwa, which is almost completely made of mulberry, is a unique string instrument judging from the position of mulberry in Fig. 4.1. Mulberry lies far off the quality criteria for Western string instruments, given by the two regression lines. It lies in between Norway maple and ebony, and has a very high value of the antivibration parameter and a very low value of the transmission parameter. However, the poor vibrational properties of mulberry seem to match the playing style of the Satsuma biwa, in which the string is strongly struck with a large triangular wooden plectrum (*bachi* in Japanese) made of hard boxwood. Striking the Satsuma biwa strings yields very characteristic impact tones because the top plate, which has low resonance, is simultaneously struck by a stroke of a large plectrum. It is very important that a mechanism has been invented to compensate for the poor resonance of the mulberry body. The mechanism, called the *sawari* (meaning *gentle touch*), allows the strings to vibrate against the neck or frets, emphasizing high frequencies [4.20, 48]. The biwa frets are very wide compared to those of the guitar and are usually made of hard magnolia (Japanese hohnoki, *Magnolia obovata*) to generate the subtle sawari effects. A few variants of this sawari are seen as *jawari* (in Hindi) on the Indian sitar and tambura, and as brays on medieval and baroque harps [4.51].

Pernambuco (Hardwood, Framewood)

Pernambuco shows values of antivibration parameter a little higher than that of Brazilian rosewood, but shows values of transmission parameter much higher than that of Brazilian rosewood. This extraordinarily high cQ suggests the property necessary to the violin bow.

4.4.3 Membranophones

Wood frames (or shells) for drums generally have negligible effects on the sound from the membranes

stretched with high tension at both sides. However, we will consider one case where the frame is tonally important, that of the Japanese drum.

Zelkova (Hardwood, Framewood)

It is well known that the most suitable shell material for a Japanese traditional wooden drum (called *wa-daiko*, where *wa* means *Japan* and *daiko* or *taiko* means *drum*) is keyaki (zelkova). Although that reason is not clear, it is said that zelkova has a kind of toughness (or viscosity). In Fig. 4.1 zelkova is plotted very closely to maple. However, Japanese maple is never applied to the wa-daiko shell. A physical or acoustical parameter should be explored to represent the above toughness of wood.

According to *Ono* et al. [4.52], a sharp peak is detected in the tap tone of the shell of a keyaki wa-daiko at 360 Hz when the membranes are stretched, but that peak is shifted to 260 Hz when the membranes are removed. This result shows that the shell of a keyaki wa-daiko is elastically deformed during the attachment of membranes with high tension. Ono explains that a material with a large E_L or c exhibits high elastic deformability. However, Norway spruce and paulownia, which are not used for drum shells, have a very large c value of about 5300 m/s, while the sound speed in zelkova is only 4180 m/s [4.2, 16, 22]. Zelkova is also used for tone plates in Japanese temples – plates that are suspended from the roof and hit with a wooden hammer to call out the time of day.

4.5 Construction

String instruments tend to be constructed out of carved plates, tuned, and glued into the form of a box. Woodwind instruments are often turned on a lathe, or make use of a biological tube, like bamboo.

4.5.1 Woodwind Instruments

The shakuhachi, a Japanese end-blown flute, is formed from the root end of a stem of bamboo. The length or the lowest pitch is determined by the top node and the root end that is precisely cut; the instrument has about seven nodes (Fig. 4.6).

The original construction method of the shakuhachi had no finish applied to the inside surface (this is called a *ground-paste-free* bore), and the culm was not divided (Fig. 4.6). The diaphragms were not completely removed and a kind of protuberance was formed along the bore near the node as shown in Fig. 4.8. The remaining portion of the diaphragm subtly affects the intonation and produces a natural tone color, which cannot be heard in a modern shakuhachis that have a ground-paste applied to the inside. Since the ground-paste-free shakuhachi made by the original method was played exclusively by a group of itinerant priests called *Komusoh*, it is called the Komusoh shakuhachi. In contrast, the ground-pasted shakuhachi is referred to as a modern shakuhachi [4.53]. The finger holes on the Komusoh shakuhachi are undercut; such an undercut is not observed in the modern instrument. The acoustical effects of the undercut should be investigated in the near future.

Finishing

Usually the inner walls of the modern shakuhachi, nohkan, ryuteki etc. are finished with a protective coating of urushi lacquer. However, its application in the shakuhachi depends on the use or otherwise of a ground paste called *ji* (meaning *ground*), a kind of paste made by mixing the polishing powder into Japanese lacquer, urushi [4.54]. The paste becomes hard enough when it is dried. In order to apply this ground paste to the inner surface of bamboo culm, all the diaphragms (solid cross walls) at the nodes are completely removed and the culm is divided near the middle (between the third and fourth finger holes). The ground paste makes it easy to adjust the intonation by modifying the bore shape and to amplify the sound volume by polishing the bore surface. The ground-pasted shakuhachi was possibly first made around the 1900s. If no ground is used, the shakuhachi is usually coated once or twice with urushi. However in the very rare but famous example, the *Rodohdoh* (shown in Fig. 4.8), the urushi coating is inferred from the thickness (over 1 mm) to have been done at least 30 times and then polished. According to a skilled professional maker [4.55], such a thick coating was probably carried out to fine-tune the intonation and tonal balance after a few years of use. The ground-pasted shakuhachi is coated at least three times because the ground paste absorbs the urushi lacquer.

Western woodwind instruments are made on a lathe. The hardwoods such as ebony and grendilla used need only to be polished (e.g., with 0000 steel wool) to produce an attractive lustrous finish. The surfaces are hard enough that a protective coating like varnish is not necessary.

4.5.2 String Instruments

Once the type of wood has been decided upon, there are two more physically significant factors involved in the choice of pieces: grain and age. There are also aesthetic

considerations, like color and the peculiar case of curly maple (Sect. 4.3.4), which will not be discussed further.

Wood Selection

Grain. Studies of wood used in classic members of the violin show [4.9] that straight grain, fine texture, and low density are favored properties. The acceptable ring width grows with instrument size: 0.8–2.5 mm limits for the violin and viola; an average 3 mm for the cello; an average 5 mm for basses. The variation in ring width within a single instrument should be such that the minimum-to-maximum ratio is larger than 0.7. In contrast, piano strings have such a wide variation in frequency that their soundboard ring widths are graded 0.7–3 mm from the treble to the bass end of a single instrument. The difference between early- and late-wood densities should be as large as possible, typically 280–900 kg/m^3, and the proportion of latewood should be no greater than 25% (more latewood increases E and internal friction).

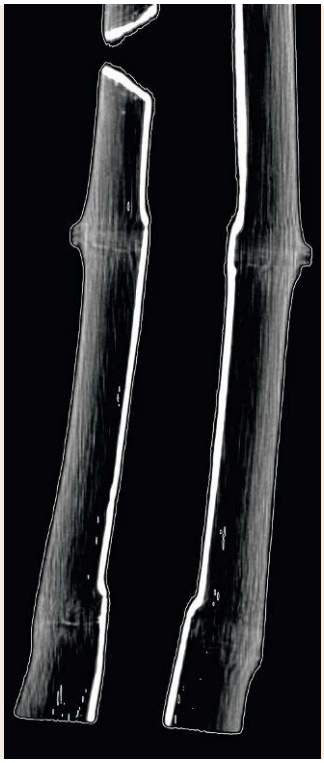


Fig. 4.8 A picture of a computed tomography-scanned old famous ground-paste-free shakuhachi (made by Shinryu Matano (1886–1936) and inscribed as *Rodohdoh*, 585 mm long, diameter 12–15 mm, in C4). Only the root side is shown. Note that the diaphragms remain a little. The bright layer indicates the urushi coating (courtesy of Hamamatsu Museum of Musical Instruments)

The Japanese long zither, koto or soh, consists of top and back plates. The top plate, about 180 cm long and about 25 cm wide, is made from a thick board of paulownia by scraping it out. The thickness of the top-plate part is about 37 mm near the middle and about 30 mm at both ends. However, the thickness at both shoulders is reduced to about 20 mm by making round grooves along the inner corners. More interestingly, the thickness of the back plate is only about 11 mm, and so acoustic radiation at high frequencies is often stronger from the back plate (with two soundholes near the ends) than from the top plate [4.56]. In contrast to the quarter-sawn wood for Western string instruments, paulownia for the koto is usually cross-grained, that is, the face of the top plate is the circumferential plane. Moreover, a high-quality koto is made of a part of the trunk close to the bark because of finer grain. The top plate is charred by a heated iron and then polished up with an steel brush and a special scrubbing brush to bring out the full beauty of the grain [4.56]. A dent or damage made on the top plate is easily restored by applying a heated iron through a wet towel. This restoring method based on the peculiarity of paulownia (although it also works with spruce) is often applied to the Chikuzen biwa whose top plate is also made of paulownia, unlike the mulberry of the Satsuma biwa [4.56].

Age. Wood is a biological material and changes continuously from the moment the tree is cut. Hence there are two significant time scales: the time between the tree cutting and the fabrication of the instrument, and the age of the instrument. For the first, dimensional stability is crucial, as green wood may be 80% water by weight and after long seasoning the moisture content may be only 10%. After several years of drying in the open air, hygroscopic equilibrium can be reached at 8–10% humidity with 60–65% environmental relative humidity. Forced drying by applied heat can, with care, speed the process up somewhat [4.57, Chaps. 6–8]. For the much longer timescales associated with old instruments, *Noguchi* et al. [4.58] studied Japanese red pine (*Pinus densiflora*) samples 8–296 yr old, and found that the sound speed increased 15% over this time span, and the loss tangent decreased 25%. The authors attributed the change to *the crystallization of cellulose, depolymerization of hemicelluloses, and cross-linking in the lignin complex*. These conclusions seem to support the widely held view that old Italian instruments are superior to modern ones. This said, recent blind comparison tests [4.59] of newly made and old Italian violins have come to the controversial conclusion that, knowing nothing of the provenance of a violin, experienced players tend to choose modern instruments.

Other string instruments are reckoned to improve with age, for example the guqin, some of which are hundreds of years old and reputed to be even better if slightly worm eaten! [4.60] (see Fig. 4.9). In contrast, harps tend not to have a long playing life, owing to the stress the strings place on the soundboard [4.44].

Western wind instruments made from hardwoods tend to be polished rather than varnished and can often be adversely affected by thermal and moisture cycling over decades. The fit and sealing of holes and keys is particularly susceptible to small dimensional changes.

Tuning Bars and Plates

The final step before gluing the parts of a soundbox together is the tuning of the plates. Tuning the wooden bars of an idiophone is a similar procedure, and is included in this section.

Bars. The first three partials of low and medium register marimba and xylophone bars are tuned to f_1, $4f_1$ and $10f_1$; higher register bars have partials tuned with smaller spacing [4.28]. *Bork* [4.29] gives detailed instructions on how to tune the partials of a bar more-or-less independently of each other. He also demonstrates how to position and tune the tubular resonators beneath the bars.

Plates. Ideally one would like to know how a soundbox is going to behave before all the constituent parts are glued together. To understand the relationship between plates and completed box from first principles is hard, and the solution lies at the edge of even early-21st-century computational physics. In the case of the violin family, the approach has been one of reverse engineering; the availability of large numbers of high quality old instruments, and the reversibility of hide glue, has made it possible to discover how to tune plates and make modern copies. Thus we have a set of aiming points for the principal modes and their frequencies. The first, second and fifth lowest modes, labeled 1, 2 and 5, (Fig. 4.10) define the plate stiffness to torsion, bending in the longitudinal direction, and bending in the transverse direction respectively. The quality of the plates remains the best indicator of the quality of the completed instrument [4.61].

Erik Jansson's practical guide to violin and guitar making describes in detail how to get from spruce and maple blocks to finished plates with the right modes [4.62]. Since this was written, the availability of software to record and analyze audio signals has made the process much more straightforward, at least for those of us who have trouble identifying the pitch of a tapped plate by ear. Free spectral analysis software can be used to find modal patterns without the need of more traditional approaches like the Chladni technique, which is hard to apply for deeply curved plates, and also very loud.

Adhesives

There are several mechanisms by which adhesives bond two wood surfaces together. The two most important are secondary forces (van der Waals, H-bonds) [4.63, p. 14] and mechanical interlocking [4.64, p. 492]. The strongest bonds are formed between two smooth clean conformal surfaces separated by a few µm of adhesive. Bonding between coarsely sanded surfaces is less strong, because of detritus on the wood surface.

Fig. 4.9 Worm holes in the soundboard (probably Paulownia) of a Ming dynasty guqin, seen through the tonehole on the underside

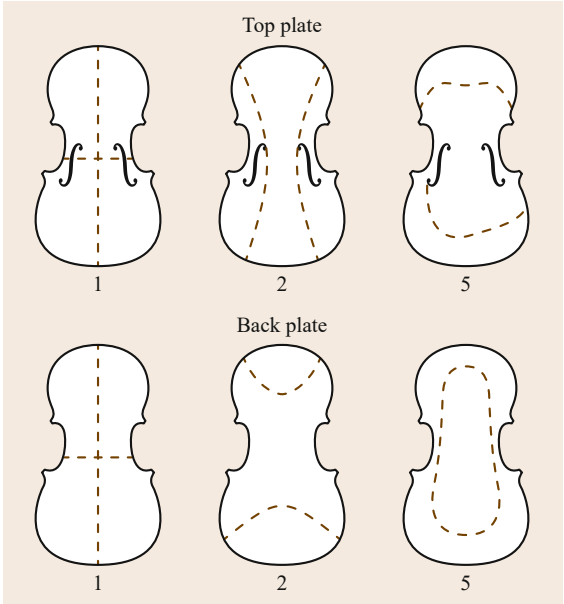


Fig. 4.10 The most important modes to get right when tuning violin plates (after [4.62])

For string instruments, one of the most important properties in an adhesive is resistance to creep. The tension in the strings causes a steady long-term stress on many wood joints. An extreme but not unusual case is that of a modern concert harp, where the combined tension of all the strings pulling on the soundboard at an angle of about 30° is of the order 10 kN.

While many modern chemical adhesives (aliphatics, polyvinyl alcohol glue, etc.) have attractive qualities, and are stronger than the wood they bond [4.64, p. 533], hide glue remains the adhesive of choice for many instrument makers.

Hide glue is made from the collagen of cattle [4.64, p. 492]. Bone glue is an alternative, but has a lower molecular weight and thus less strength. Hide glue is cheap, and comes in dry granules that can be dissolved in water at 60 °C. Care had to be taken with the temperature, as the glue degrades if overheated. When cool it forms a moderately strong bond, which gains in strength with natural drying. Hide glue is more creep resistant than common industrial glues, it does not leave marks on the wood, and joints can be undone without damage to the wood by the application of a little steam.

Hide glue properties are critical to the construction of violin family members. The front and back plate that determine the sound quality of the instruments are typically made from two pieces of wood glued along the centerline (Fig. 4.4). In order to maintain symmetry between the wood figure on either side of this line, one piece is cut down the middle, planed on one edge and folded out like opening a book. With some skill, these two edges are now correctly jointed (the technique being insensitive to small errors in the planing angle). If hide glue is applied to these two cleanly planed edges, the two halves of the plate can be bonded without clamping; the natural shrinkage of the glue being sufficient to draw the two wood surfaces together [4.65]. The great advantage of not using clamps is that there is now only minimal stress built into the glue joint, and it will survive environmental changes much better than if external clamps had been used to force the two pieces of wood together.

With hide glue it is also possible to assemble a violin family soundbox with dry glue joints, giving time to arrange the delicate clamping that is necessary to keep the two plates and ribs together while the glue sets. Glue is applied to all contact surfaces and allowed to dry until no longer sticky. Then the instrument is assembled and clamped in a calm and considered manner, and finally a jet of steam is quickly applied to the glue to soften it and complete the bonding.

Varnish

Volumes have been written about violin varnish. Let us simply quote here the view of Antoine Vidal, a friend of the great 19th century French luthier Jean-Baptiste Vuillaume (quoted in [4.66]):

> *The violin finished and not varnished has more power and mellowness in its tone; but if it remains in this virgin state, it becomes modified little by little, and, after a somewhat short time, the tone becomes poor and feeble. It must therefore be concluded that the varnish, while giving a more pleasing appearance, conserves, preserves, and that therein, above all, consists its great utility.*

Barlow and Woodhouse used electron microscopy to examine the varnish layers of old Italian and Dutch violins and celli [4.67]. The wood was first covered with a relatively thick (about 30 µm) *ground* covered by a much thinner layer of varnish. The ground layer consisted of particulate matter resembling plaster of Paris, and apparently comparable to modern Polyfilla, which, once varnished, becomes transparent. The function of this layer is certainly the sealing of the wood to prevent the colored varnish soaking in and producing a patchy finish; spruce, being more porous than maple, is more susceptible to this effect. However, the ground layers appeared to be much thicker than necessary to achieve this end, so there may have been additional reasons for its use.

Schelleng's study of the vibrational effects of varnishing [4.68] showed definitively that varnish should be hard and thin to avoid excessive damping of the plates. More recent tests show varnishing increases stiffness and mass (thus the speed of sound can go either way); in particular the modulus of elasticity and internal friction in the R direction is much increased for spruce [4.9]. In studies of harp soundboards *Gunji* et al. [4.13] found that a $400\,g/m^2$ layer of lacquer ($Q \approx 50$) will decrease the quality factor of the soundboard by 20%.

Testing

The ultimate test of a recently constructed instrument is of course the playing of it. However, many physical tests can be made on the complete instrument, of which tap tests are the most accessible to luthiers without sophisticated equipment. The principal air and wood modes of the guitar, ruan and even a concert harp [4.69] can be identified by tapping the instrument with and without light styrofoam blocks in the soundholes. Many articles have been written about identifying violin modes and describe corrective action if some are in the wrong place, e.g., [4.70].

4.6 Conclusion

In this chapter we have surveyed the relationship between wood, and how it is chosen, worked and finished, and the musical instruments from which it is made. In all cases we have worked backwards from well-established instruments, some of which reached their current state of evolution long before the age of scientific analysis. In the manufacture of wooden musical instruments, a little technical analysis is a great benefit; it will not ensure the invention or production of great instruments, but it can prevent a lot of mistakes.

4.A Appendix

The quality factor Q is a measure of the damping of a material; the higher the Q the more resonant the material is. If τ is the time it takes for a system freely vibrating at frequency f_0 to decay to $1/e$ times its original amplitude, then $Q = \pi f_0 \tau$. Alternately, in the frequency domain, if the width of a resonance at half-power is Δf, then $Q = f_0/\Delta f$. In older texts damping is referred to as the logarithmic decrement δ, and for $Q \gg 1$, $Q \approx \pi/\delta$ (or if the author is working in base-ten logarithms, $Q \approx 1.365/\delta_{10}$. Another form is the dimensionless damping ratio $\zeta = 1/(2Q)$, the loss factor η and the loss angle ψ.

$$Q = \frac{1}{\eta} = \frac{\pi}{\delta} = \frac{1}{\tan\psi} \,. \tag{4.6}$$

Another important characteristic for instrument woods is vibration transmission [4.2, 20, 48]. If the damping is relatively weak, the characteristic acoustic transmission is the reciprocal of the attenuation constant α of the longitudinal wave. The solution of the lossy wave equation gives

$$\alpha^{-1} = \frac{2Q}{k} = \frac{2cQ}{\omega} \,, \tag{4.7}$$

where k and ω are the wave number and angular frequency respectively [4.32, 48]. Since ω is not a wood property, cQ may be used instead of cQ/ω to express the transmission characteristic of the vibration excited in wood. In the text, cQ is called the transmission parameter.

References

4.1 I.R. Titze, J. Sundberg: Vocal intensity in speakers and singers, J. Acoust. Soc. Am. **91**, 2936–2976 (1992)
4.2 S. Yoshikawa: Acoustical classification of woods for string instruments, J. Acoust. Soc. Am. **122**, 568–573 (2007)
4.3 T. Ono: Transient response of wood for musical instruments and its mechanism in vibrational property, J. Acoust. Soc. Jpn. (E) **12**, 117–124 (1999)
4.4 T. Ono, S. Miyakoshi, U. Watanabe: Acoustic characteristics of unidirectionally fiber-reinforced polyurethane foam composites for musical instrument soundboards, Acoust. Sci. Technol. **23**, 135–142 (2001)
4.5 J. Schelleng: The violin as a circuit, J. Acoust. Soc. Am. **35**, 326–338 (1963)
4.6 D.E. Kretschmann: Mechanical properties of wood. In: *The Wood Handbook*, ed. by R.J. Ross (US Department of Agriculture, Madison 2010)
4.7 E. Fukada: The vibration properties of wood I, J. Phys. Soc. Japan **5**, 321–327 (1950)
4.8 I. Barducci, G. Pasqualini: Measurement of the internal friction and the elastic constants of wood, Nuovo Cimento **5**, 416–466 (1948)
4.9 V. Bucur: Wood species for musical instruments. In: *Acoustics of Wood*, ed. by V. Bucur (Springer, Berlin 2006)
4.10 M.E. McIntyre, J. Woodhouse: On measuring the elastic and damping constants of orthotropic sheet materials, Acta Metallurg. **36**, 1397–1416 (1988)
4.11 R.B. Hoadley: *Identifying Wood* (Stevens, Newtown 1990)
4.12 I.M. Firth, A.S. Bell: The acoustical effects of wood veneer, Acustica **66**, 114–116 (1988)
4.13 T. Gunji, E. Obataya, K. Aoyama: Vibrational properties of harp soundboard with respect to its multilayered structure, J. Wood Sci. **58**, 322–326 (2012)
4.14 W. Liese: *The Anatomy of Bamboo Culms* (International Network for Bamboo and Rattan, Beijing 1998)
4.15 U.G.K. Wegst: Bamboo and wood in musical instruments, Annu. Rev. Mater. Res. **38**, 323–349 (2008)
4.16 D.W. Haines: On musical instrument wood, Catgut Acoust. Soc. Newsl. **31**, 23–32 (1979)
4.17 C.Y. Barlow: Materials selection for musical instruments, Acoust. Aust. **19**, 69–78 (1997)
4.18 I. Brémaud: Personal communication (2007)
4.19 U.G.K. Wegst, S. Oberhoff, M. Weller, M.F. Ashby: Materials for violin bows, Int. J. Mater. Res. **12**, 1230–1237 (2007)
4.20 S. Yoshikawa, M. Shinoduka, T. Senda: A comparison of string instruments based on wood properties: Biwa versus cello, Acoust. Sci. Technol. **29**, 41–50 (2008)
4.21 E. Obataya: Personal communication (2013)
4.22 H. Aizawa, E. Obataya, T. Ono, M. Norimoto: Acoustic converting efficiency and anisotropic nature of wood, Wood Res. **85**, 81–83 (1998)

4.23 J.J.A. Janssen: The mechanical properties of bamboo used in construction. In: *Bamboo Research in Asia*, ed. by G. Lessard, A. Chouinard (International Development Research Centre, Ottawa 1980)
4.24 Y. Kubojima, Y. Inokuchi, Y. Suzuki, M. Tonosaki: Shear modulus of several kinds of japanese bamboo obtained by flexural vibration test, J. Wood Sci. **56**, 64–70 (2010)
4.25 B.A. Yankovskii: Dissimilarity of the acoustic parameters of unseasoned and aged wood, Sov. Phys. Acoust. **13**, 125–127 (1967)
4.26 E. Obataya, T. Ono, M. Norimoto: Vibrational properties of wood along the grain, J. Mater. Sci. **35**, 2993–3001 (2000)
4.27 H. Aizawa: *Frequency Dependence of Vibration Properties of Wood in the Longitudinal Direction* (Kyoto Univ, Kyoto 1998)
4.28 T.D. Rossing, J. Yoo, A. Morrison: Acoustics of percussion instruments: An update, Acoust. Sci. Technol. **25**, 406–412 (2004)
4.29 I. Bork: Practical tuning of xylophone bars and resonators, Appl. Acoust. **46**, 103–127 (1995)
4.30 H.F. Meinel: Regarding the sound quality of violins and a scientific basis for violin construction, J. Acoust. Soc. Am. **29**, 817–822 (1957)
4.31 N.H. Fletcher, T.D. Rossing: *The Physics of Musical Instruments* (Springer, New York 1998)
4.32 E. Meyer, E.G. Neumann: *Physical and Applied Acoustics* (Academic, New York 1972) pp. 14–22
4.33 A.H. Benade: *Fundamentals of Musical Acoustics* (Oxford Univ. Press, New York 1976)
4.34 C. Gough: Musical acoustics. In: *Handbook of Acoustics*, ed. by T. Rossing (Springer, New York 2007)
4.35 I.M. Firth: Acoustics of the harp, Acustica **37**, 148–154 (1977)
4.36 S. Daltrop, C.E. Waltham, A. Kotlicki: Vibro-acoustic characteristics of an aoyama amphion concert harp, J. Acoust. Soc. Am. **128**, 466–473 (2010)
4.37 F. Fahy, P. Gardonio: *Sound and Structural Vibration* (Academic, Amsterdam 2007)
4.38 E.B. Davis: Designing soundboards with flexural disk models, Proc. Meet. Acoust. **12**, 035003 (2012)
4.39 O. Christensen, B.B. Vistinen: Simple model for low frequency guitar function, J. Acoust. Soc. Am. **68**, 758–766 (1980)
4.40 A. Askenfelt: Double bass. In: *The Science of String Instruments*, ed. by T.D. Rossing (Springer, New York 2010)
4.41 G. Weinreich: What science knows about violins, and what it does not know, Am. J. Phys. **61**, 1067–1077 (1993)
4.42 J. Schelleng: Erratum: The violin as a circuit, J. Acoust. Soc. Am. **35**, 1291 (1963)
4.43 J. Curtin, T.D. Rossing: Violin. In: *The Science of String Instruments*, ed. by T.D. Rossing (Springer, New York 2010)
4.44 C.E. Waltham: Harp. In: *The Science of String Instruments*, ed. by T.D. Rossing (Springer, New York 2010)
4.45 S. Daltrop, C.E. Waltham, A. Kotlicki, F. Gautier, B. Elie: Vibroacoustic characteristics of a gothic harp, J. Acoust. Soc. Am. **131**, 837–843 (2012)
4.46 E. Obataya, H. Akahoshi: Personal communication (2013)
4.47 L.J. Gibson, M.F. Ashby: *Cellular Solids: Structure and Properties* (Cambridge Univ. Press, New York 1988)
4.48 S. Yoshikawa: Plucked string instruments in Asia. In: *The Science of String Instruments*, ed. by T.D. Rossing (Springer, New York 2010)
4.49 A. Baines: *The Oxford Companion to Musical Instruments* (Oxford Univ. Press, New York 1992)
4.50 E. Obataya, H. Yamauchi: Applicability of laminated veneer cylinder to sustainable production of woodwind instruments. In: *Proc. 2012 IUFRO* (International Union of Forest Reserch Organisation, Copenhagen 2012)
4.51 I.M. Firth: Harps of the baroque period, J. Catgut Acoust. Soc. **1**(II), 52–61 (1989)
4.52 T. Ono, I. Takahashi, Y. Takasu, Y. Miura, U. Watanabe: Acoustic characteristics of wadaiko (traditional Japanese drum) with wood plastic shell, Acoust. Sci. Technol. **30**, 410–416 (2009)
4.53 S. Simura: *An Organology of Kokan Syakuhati (Old Pipe Syakuhati)* (Shuppan-geijutsu-sha, Tokyo 2002)
4.54 E. Obataya, Y. Ohno, M. Norimoto, B. Tomita: Effects of oriental lacquer (urushi) coating on the vibrational properties of wood used for the soundboards of musical instruments, Acoust. Sci. Technol. **22**, 27–34 (2001)
4.55 S. Yamaguchi: Personal communication (2013)
4.56 Y. Ando: *Acoustics of Musical Instruments* (Ongaku-no-tomo-sha, Tokyo 1996)
4.57 R.B. Hoadley: *Understanding Wood* (Taunton, Newtown 2000)
4.58 T. Noguchi, E. Obataya, K. Ando: Effects of aging on the vibrational properties of wood, J. Cultural Herit. **13S**, S21–S25 (2012)
4.59 C. Fritz, J. Curtin, J. Poitevineau, P. Morrel-Samuels, F.-C. Taoi: Player preference among new and old violins, PNAS **109**, 760–763 (2012)
4.60 A. Thrasher: Personal communication (2013)
4.61 J. Curtin: Tap tones and weights of old violin tops, J. Violin Soc. Am. **20**, 161–173 (2006)
4.62 E. Jansson: Acoustics for Violin and Guitar Makers, http://www.speech.kth.se/music/acviguit4/ (2013)
4.63 A. Pizzi: *Advanced Wood Adhesives Technology* (Dekker, New York 1994)
4.64 E.M. Petrie: *Handbook of Adhesives and Sealants* (McGraw Hill, New York 2007)
4.65 H.A. Strobel: *Violin Making Step by Step* (Strobel, Oregon 1994)
4.66 G. Fry: *The Varnishes of the Italian Violin Makers of the 16th and 17th Centuries and Their Influence on Tone* (Stevens, London 1904)
4.67 C.Y. Barlow, J. Woodhouse: Of old wood and varnish: Peering into the can of worms, J. Catgut Acoust. Soc. **1**, 2–9 (1989)
4.68 J. Schelleng: Acoustical effects of violin varnish, J. Acoust. Soc. Am. **44**, 1176–1183 (1968)
4.69 C.E. Waltham: The acoustics of harp soundboxes, Am. Harp J. **22**, 26–31 (2010)
4.70 C. Hutchins, D. Voskuil: Mode tuning for the violin maker, Catgut Acoust. Soc. **2**(4), 5–9 (1993)

5. Measurement Techniques

Thomas Moore

The measurements normally required to understand the physics of musical instruments, including the human voice, usually fall into one of three categories: measuring the airborne sound, measuring the deflection of the surface of an instrument, or measuring the input impedance. This chapter introduces the most common measurement techniques that provide information on these three physical parameters with an emphasis on the first two, which are the measurements most commonly desired by musical acousticians. The chapter begins with a discussion of airborne sound and how it is sensed. Specifically, several types of microphones are introduced followed by a discussion of some of the techniques that rely on sensing by microphones. A review of the techniques for measuring and visualizing deflection shapes is then presented. These techniques range from observing nodal lines using simple Chladni patterns to visualizing deflection shapes using electronic speckle pattern interferometry. The topic of impedance measurement is addressed next, with discussions of both measurements of the input impedance of wind instruments and the measurement of mechanical impedance. This review is not meant to be a complete analysis of each measurement technique. Instead, it is meant to serve as an introduction to the most commonly used techniques and provide references for the interested reader to pursue further study. The advent of new technologies continually changes the equipment that is available to the scientist, but the underlying physical principles remain relevant.

5.1 Measurement of Airborne Sound

The most ubiquitous measuring device in musical acoustics, as is the case with almost all other areas of acoustical study, is the microphone. Since musical sound is normally transmitted to the listener through the air, it is the sound in air that is normally of analytical interest. Therefore, the microphone is the most useful instrument for detecting the sound. Once detected, the signal from the microphone can be analyzed in various ways, but the accurate translation of the sound into an electrical signal is of paramount importance; this is the purpose of the microphone. Translating air pressure into an electrical signal can be accomplished through various means, and therefore there are several different types of microphones. Similarly, there are different types of sounds, and therefore there are microphones that are designed for different applications. This section introduces several applications and types of microphones, with an emphasis on the strengths and limitations of microphones in different applications involving the study and recording of musical instruments.

5.1.1 Types of Microphones: Form

Microphones can be classified by both form and function; form referring to how the microphones are constructed and function referring to how they are applied.

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_5

There are many types of microphones in common usage but condenser microphones are by far the most commonly used in musical acoustics. Other types have specific advantages that cause them to occasionally be used by musical acousticians, therefore we will briefly review dynamic microphones and ceramic microphones as well. Detailed discussions of these and several other types of microphones can be found in [5.1, 2].

Condenser microphones utilize a change of capacitance to detect external pressure (condenser is an obsolescent term for capacitor). Electret microphones are one type of condenser microphone, however, they are usually referred to as electret microphones to distinguish them from the more common type, which does not use an electret material.

Electret microphones are made by depositing an electret film on a diaphragm and placing the diaphragm in close proximity to a plate that is connected to an amplifier. The electret film holds a charge that is permanent, therefore it forms a charged parallel-plate capacitor with the plate. As the diaphragm moves, the distance between the diaphragm and plate changes. The capacitance of a parallel plate capacitor with an air gap is given by

$$C = \frac{\varepsilon_0 A}{d} , \tag{5.1}$$

where A is the area of the plates, d is the distance between them and ε_0 is the permittivity of free space. The capacitance of a charged capacitor is given by

$$C = \frac{Q}{V} , \tag{5.2}$$

where Q is the charge on the capacitor and V is the potential difference between the plates. From these two equations it is obvious that the potential difference between the two plates is linearly proportional to the distance between them provided Q is a constant. An electret has a constant charge, and therefore the motion of the diaphragm is linearly translated into a change in the potential difference between the plates. Electret microphones do not require external power, however, they often contain an integrated preamplifier that does.

The microphone more commonly referred to as a condenser microphone works on the same principle as the electret microphone, however, the charge on the plates is the result of being connected to a power source. It is sometimes referred to as an externally biased or externally polarized microphone. The charge may come from a battery or may be delivered through a preamplifier. When the power source comes from an external preamplifier it is often supplied at 48 V and is referred to as *phantom power*. Condenser microphones, whether they involve an electret or not, have a flatter frequency response over the range of human hearing than dynamic microphones and are typically more sensitive than ceramic microphones. Both dynamic and ceramic microphones are discussed in the following paragraphs.

It is becoming common to refer to electret microphones as simply condenser microphones, so it is often unclear what type of condenser microphone is being used. Since the response is similar for both types it is rarely important when analyzing the electrical signals, but the power requirements for the two can be different. Therefore, it is important to know the type of condenser microphone in use when designing an experimental apparatus.

Dynamic microphones have the same construction as a common speaker, but rather than transforming an input electrical signal to a pressure wave in the air, a dynamic microphone produces an electric signal from the motion of a diaphragm. The diaphragm is attached to a coil that is free to move in a static magnetic field and the resulting electrical signal is a consequence of Faraday's law, i. e., the induced electromotive force in a coil moving in a magnetic field is given by

$$E = -\frac{\mathrm{d}\Phi}{\mathrm{d}t} , \tag{5.3}$$

where Φ is the magnetic flux through the coil, $\mathrm{d}t$ is the differential element of time, and the SI system of units assumed. These microphones are very robust, but have limited frequency response and local electric fields from nearby equipment (e.g., transformers) can induce current in the coil leading to noise in the signal.

Ceramic microphones have a piezoelectric element attached to a diaphragm. When the diaphragm moves in response to a pressure change it induces a strain in the piezoelectric element, which produces a potential difference between the two sides of the crystal that is linearly proportional to the stress. These microphones are extremely robust and can be used for measurements of high pressures, for example inside of brass wind instruments. The frequency response of ceramic microphones is sometimes limited in the high frequencies, so it is advisable to review the response curve before using one. Like dynamic microphones, they do not require a power supply.

5.1.2 Types of Microphones: Function

The application will determine which type of microphone to choose, but it also defines the function of the microphone. The function is usually characterized by the type of acoustic field that is being measured. The three most common acoustic fields are free, diffuse (random incidence), and pressure.

A free-field microphone is intended to be used in a situation where there is a single source and there are no reflections from nearby objects. This situation usually occurs in an anechoic chamber or outdoors. Although the acoustic field is distorted by the presence of the microphone, a free-field microphone is designed to compensate for these effects. The result is that the measurement is as if the microphone were not present. When in use, the microphone should be positioned so that the diaphragm is normal to the incoming pressure variation. That is, it should be pointed directly at the source. Free-field microphones are normally used for scientific investigations of musical instruments.

A random-incidence or diffuse-field microphone is intended to be used in situations where the sound source is not localized. That is, it is not a free field. This situation occurs naturally in enclosed spaces where the reflections from walls cause the pressure variations to be incident on the microphone from multiple directions. Sometimes referred to as an omnidirectional microphone, the diffuse-field microphone corrects for both its presence in the field and the variations caused by sound being incident from different angles. Random-incidence microphones are normally used for recording the sounds of musical instruments during performance or for testing room acoustics.

A pressure-field microphone is not normally used in musical acoustics. It is designed to measure the acoustic field as it actually exists in front of the microphone. Therefore, the presence of the microphone is not taken into account. These microphones are usually mounted into structures (e.g., walls) to determine the incident pressure field.

5.1.3 Microphone Arrays and Near-Field Acoustic Holography

Microphones are often used individually to sense discrete points in a sound field, but they are also often used in combination with other microphones. Multiple microphones have been common for recording musical performances since the advent of stereophonic recording. More recently, the common use of multiple speakers in playback of a recording (i. e., surround-sound) has resulted in a considerable body of work analyzing many of the possible arrangements. Arrays of three and four microphones are common for recording music and analyzing acoustic spaces, [5.3, 4] but spherical arrays of microphones have been used recently and can provide three-dimensional information about a sound field [5.5, 6]. However, when investigating musical instruments, microphone arrays are typically used to identify the position of the radiating areas on the instrument. Precisely locating such areas can lend insight into the motion of the instrument as well as the pattern of acoustic radiation emanating from it.

Determining the origin of the radiation of a musical instrument can often be accomplished by imaging the motion of the surface. Several processes useful for this will be discussed in Sect. 5.2. However, in many instruments the origin of the sound cannot be directly imaged, such as when the sound originates from holes in the instrument. In these cases the radiation pattern can be determined by placing an array of microphones in the acoustic near field (less than a wavelength from the source) and determining the originating source velocity through a process that has become known as near-field acoustic holography (NAH). NAH is a natural outgrowth of acoustic holography and allows one to reconstruct the pressure and particle velocity at any point between the sensing array and the radiating object, hence the holographic reference.

NAH was first described by *Williams* and *Maynard*, who noted the possible application to musical acoustics as one of the motivations for developing the process [5.7]. The process is explained in detail in [5.8] and an excellent overview of acoustical holography and NAH can be found in [5.9]. The interested reader should consult these sources for a detailed explanation, but a short introduction is sufficient to understand acoustic holography and gain an appreciation for the analytical power of the process. We begin by analyzing the situation where the radiated sound field is measured in one plane and the field at a plane further from the source is calculated from this information (conventional acoustic holography). We then make the logical transition to NAH.

Conventional Acoustic Holography

If the acoustic pressure is known at any plane some distance from a radiating source, then it is possible to calculate the pressure and particle velocity at any other plane in free space. From these two quantities the acoustic intensity can be calculated. Since this is true at any plane, it is possible to construct a three-dimensional image of the sound field analytically. Although this process, known as acoustic holography, is theoretically similar to that of the more familiar optical holographic process, in practice it is much different. The differences are discussed in detail in [5.9].

Although there are more complete and rigorous ways to describe the process, it is sufficient to understand the process of acoustic holography by noting that any solution to the wave equation for a pressure in free space can be written as a sum of a complete set of plane waves. These plane waves are defined by their angular frequency ω, amplitude P and wave vector $\boldsymbol{k}$. In what follows we assume a single angular frequency is of in-

terest, keeping in mind that in any given situation there may be many frequencies that must be analyzed separately.

For simplicity we proceed from here by assuming a Cartesian coordinate system, realizing that any orthogonal coordinate system can be used if it better suits the physical system under investigation.

We may represent each plane wave in space in the normal way as an amplitude P with a spatially oscillating part that is dependent upon the direction, speed and frequency of the wave. These plane waves are represented by

$$p_{\boldsymbol{k}}(\boldsymbol{r}) = P_{\boldsymbol{k}} \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}} \,, \tag{5.4}$$

where $P_{\boldsymbol{k}}$ is the wave amplitude, $\boldsymbol{k}$ is the wave vector defined in the Cartesian coordinate system as $\boldsymbol{k} = k_x\hat{i} + k_y\hat{j} + k_z\hat{k}$ with a magnitude of ω/c, $\boldsymbol{r}$ is the spatial coordinate (x, y, z), and c is the speed of sound in the medium. Since a single frequency is assumed we have suppressed the time-dependent nature of (5.4).

On any plane of constant z the field can be represented by a summation of plane waves in the x–y plane. The amplitude of each plane wave, which is dependent upon the wave vector, can be determined by projecting the pressure wave onto the basis set of plane waves in the x–y plane. Therefore

$$P(k_x, k_y, z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y, z) \times \mathrm{e}^{-i(k_x x + k_y y)} \,\mathrm{d}x\mathrm{d}y \,, \tag{5.5}$$

and any solution to the free space wave equation can be written as

$$p(x, y, z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(k_x, k_y, z) \times \mathrm{e}^{\mathrm{i}(k_x x + k_y y)} \mathrm{d}k_x \mathrm{d}k_y \,. \tag{5.6}$$

Note that (5.5) and (5.6) form a Fourier transform pair and therefore can be written as

$$\mathcal{F}\{p(x, y, z)\} = P(k_x, k_y, z) \tag{5.7}$$

and

$$\mathcal{F}^{-1}\{P(k_x, k_y, z)\} = p(x, y, z) \,. \tag{5.8}$$

$P(k_x, k_y, z)$ is usually referred to as the angular spectrum.

Let us assume the situation shown in Fig. 5.1, where a radiating planar source is present on the plane $z = 0$. If the pressure is known on the (x', y') plane, which is

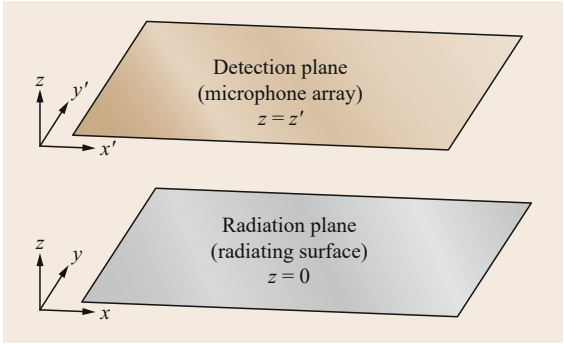


Fig. 5.1 Coordinate system for analyzing acoustic holography. The instrument under study is located at coordinates $(x, y, 0)$ and the microphones are located at coordinates (x', y', z')

parallel to the radiating surface and located some distance z' away, the pressure can be found on any other parallel plane by calculating the angular spectrum at z', propagating it to the new plane at z, and projecting the spectrum onto all possible plane waves in the x–y plane. The propagator function (i. e., Green's function) is

$$\mathrm{e}^{\mathrm{i}k_z(z-z')} \,, \tag{5.9}$$

and therefore

$$p(x, y, z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(k_x, k_y, z') \times \mathrm{e}^{\mathrm{i}(k_x(x-x') + k_y(y-y') + k_z(z-z'))} \mathrm{d}k_x \,\mathrm{d}k_y \,. \tag{5.10}$$

Or put more simply,

$$p(x, y, z) = \mathcal{F}^{-1}\left\{P\left(k_x, k_y, z'\right) \mathrm{e}^{\mathrm{i}k_z(z-z')}\right\} \,, \tag{5.11}$$

where x and y in (5.5) and (5.6) are replaced with $(x - x')$ and $(y - y')$. In this way the pressure at any point in space can be determined simply by measuring the pressure on the plane defined by $z = z'$.

It is important to remember that k_z is not independent of k_x and k_y since

$$k = \frac{\omega}{c} = \sqrt{k_x^2 + k_y^2 + k_z^2} \,,$$

and conventional acoustic holography demands that k_z be real. That is, $k^2 - k_x^2 - k_y^2 > 0$. Physically this restriction means that only freely propagating waves are considered since waves in which $k^2 - k_x^2 - k_y^2 < 0$ are evanescent. More will be said about this in the next section.

The challenge of acoustic holography is to determine the pressure field on the z' plane. Normally, an

array of closely spaced microphones is used and the precision of the measurement is determined by the wavelength and the sampling spatial frequency. Although pressure measurements occur at discrete points, if the plane is adequately sampled the far-field pressure can be calculated quite accurately by summing the pressure resulting from each point on (x', y').

To determine the acoustic intensity it is necessary to know the particle velocity in addition to the pressure. The particle velocity associated with the calculated pressure field can be derived using Euler's equation

$$\boldsymbol{v}(\boldsymbol{r}) = \frac{-i}{\omega\rho_0}\nabla p(\boldsymbol{r})\,, \tag{5.12}$$

where ρ_0 is the density of air. In Cartesian coordinates this can be written as

$$\boldsymbol{v}(\boldsymbol{r}) = \frac{1}{\rho_0 ck}\mathcal{F}^{-1}\left\{\left(\hat{\imath}k_x + \hat{\jmath}k_y + \hat{k}k_z\right) \times P\left(k_x, k_y, z'\right)\mathrm{e}^{\mathrm{i}k_z(z-z')}\right\}, \tag{5.13}$$

where as usual $k = \sqrt{k_x^2 + k_y^2 + k_z^2}$.

Near-Field Acoustic Holography

Determining the pressure and vector velocity at a point where $z > z'$ is known as a forward problem. In this way one may determine the acoustic intensity at any point in the direction of travel of the disturbance. However, in the study of musical instruments it is often more useful to determine the pressure and velocity fields that created the disturbance. That is, we are normally interested in the surface of the instrument. To investigate the radiation from the surface of an instrument one must calculate the desired quantities at a plane where $z < z'$. This is known as an inverse problem. Typically the measurement is made in the far field and for obvious reasons the resolution is limited to the acoustic wavelength. However, this resolution limit can be overcome by placing the recording microphones in the near field of the radiator, which is defined as being significantly less than a wavelength distant from the source. In practice the distance will normally be less than approximately $1/8$ wavelength. Solving the inverse problem using measurements made in the near field is referred to as near-field acoustic holography (NAH).

The requirement for making the measurements in the near field stems from the necessity to detect evanescent waves. These waves are generated by subsonic bending waves in the object, for example the top plate of a guitar, which have shorter wavelengths than the acoustic wavelength of the sound in the air. This forces the bending waves to move at a slower speed than sound waves of the same frequency do in air, and the shorter wavelengths provide increased precision of the measurement. These evanescent waves decay exponentially as they propagate away from the surface; therefore the microphones must be close to the surface to detect them. Assuming that the surrounding air stays in contact with the surface, the surface velocity of the source can be determined by calculating the velocity of the air at the surface of the object. This can be calculated from the pressure field measured at some plane a distance z' from the surface.

Theoretically, one can solve the inverse problem in a manner similar to the method one uses to solve the forward problem. That is, measure the pressure at some plane a distance z' from the surface located at $z = 0$, and then compute the angular spectrum by calculating the Fourier transform (5.6), multiply by an inverse propagator to propagate the field from z' to $z = 0$ and then perform an inverse Fourier transform to determine the real quantity. Normally the measured quantity is the pressure in the z'-plane and the calculated quantity is the velocity at $z = 0$.

There are several subtleties in this process that are beyond the scope of this review and the interested reader should consult [5.8, 9]. However, it is important to note two points. First, the velocity at the surface of the radiator can be calculated using (5.13), so that the velocity normal to the surface is given by

$$v_z(x, y, 0) = \frac{k_z}{\rho_0 ck}\mathcal{F}^{-1}\left\{P\left(k_x, k_y, z'\right)\mathrm{e}^{-\mathrm{i}k_z z'}\right\}. \tag{5.14}$$

As usual, $k_z = \sqrt{k^2 - k_x^2 - k_y^2}$, but in the case of evanescent waves k_z is purely imaginary because $k_x + k_y > k$. Therefore, the exponent is real and the wave decays exponentially with increasing z. The second point to note is that the measuring process is discrete in nature and introduces high spatial frequencies that are not present in the sound. These must be eliminated by applying a low-pass filter during the calculation. Other issues also complicate the measurement, including microphone spacing and noise; these are all addressed in general in [5.8] with some of the more practical aspects of applying NAH being addressed in [5.10]. Naturally, there is no reason to assume that a Cartesian coordinate system is the logical choice in every instance, but the theory is independent of the coordinate system used and implementation in spherical and cylindrical systems is also common.

An example of the intensity field around a tenor steel pan derived by NAH is shown in Fig. 5.2 [5.11].

A similar example of the sound field around a violin can be found in [5.12].

While we have addressed the physics of NAH and microphones in general, once the signal has been detected there is a significant amount of signal processing that can be applied. In addition to NAH, microphone arrays can be used for enhancing directivity, differential detection, localization and echo reduction, among many other things. The signal processing involved in these applications has been addressed in several places and a good overview can be found in [5.13, 14].

Equivalent Sources Calculations

As noted above, the sound field radiated by a musical instrument can be reconstructed using measurements made with microphone arrays by NAH, but these same measurements can be used in a different manner to achieve the same goal [5.15]. One example is an equivalent sources calculation, sometimes referred to as the source simulation technique, which is a procedure that represents the radiating surface as a collection of individual sources. These virtual sources are arranged such that the linear superposition of the radiation from all of them creates a field that is equivalent to the measured pressure field. Once this is accomplished, the field at the surface of the radiating body (or any other point in space) can be calculated by superimposing the field resulting from each source in the plane of interest. While the technique of equivalent sources was first proposed and demonstrated in 1989 [5.16], new techniques for solving the inverse problem are still being published.

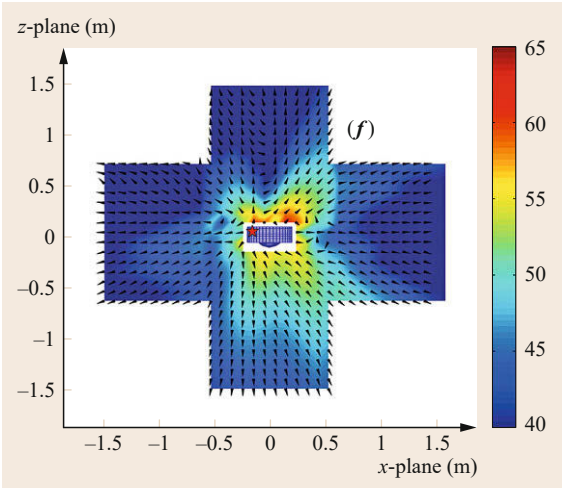


Fig. 5.2 Active acoustic intensity near a tenor steel pan derived from NAH measurements (after [5.11]). The active intensity indicates the magnitude and direction of the acoustic energy that propagates to the far field

Although the implementation of equivalent source methods can sometimes be tricky, the theory is straightforward. One merely posits a series of radiators on or behind the surface of the object and adjusts the phases and amplitudes so that the pressure recorded in the plane of the array matches the calculated pressure field. The radiators can be a series of monopoles, dipoles, or radiators with complex shapes. The correct choice of the virtual radiators is often critical to ensuring that the algorithm converges to a solution.

Assuming that all possible fields created by the real source can be represented by some collection of virtual sources on or behind the surface of interest, if the amplitudes of each of the sources is known then the pressure at any point in space can be calculated through (5.11). The problem is that the pressure is measured at some plane z' and not at the position of the virtual sources. That is, an inverse problem must be solved before (5.11) can be applied. This inverse problem is significantly different than the one that must be solved when performing NAH.

In an equivalent source calculation, rather than back propagating the measured acoustic field using (5.14), the geometry of the virtual radiators is assumed and the amplitudes are adjusted to match the measured field at the position of the microphone array. Assuming that there are N virtual monopole radiators at some point behind the surface of the actual radiator, then the pressure at each point in the plane of the virtual radiators can be designated as $p_n(x, y, 0)$. In the plane of the microphone array containing M microphones, the pressure at each microphone is $p_m(x', y', z')$. Then the problem is to ensure that

$$p_m(x', y', z') = \mathcal{F}^{-1}\left\{\mathcal{F}\left\{\sum_{n=1}^{N} p_n(x, y, 0)\right\} \times e^{ik_z(z-z')}\right\} \quad (5.15)$$

when all that is known is $p_m(x', y', z')$ for all M microphones. The ease with which this inverse problem can be solved depends upon the choice of sources as well as the method of solution and can be quite difficult. Furthermore, since there will be noise in the measurements the problem is ill posed and regularization is necessary. However, there are several algorithms available for solving these problems and multiple source methods can be very useful in the context of musical acoustics [5.17].

If the virtual sources are not assumed to be behind the radiating surface one must still solve the inverse problem, however, a simple application of some variation of (5.15) is not the most accurate or efficient procedure. When assuming that the sources are on the

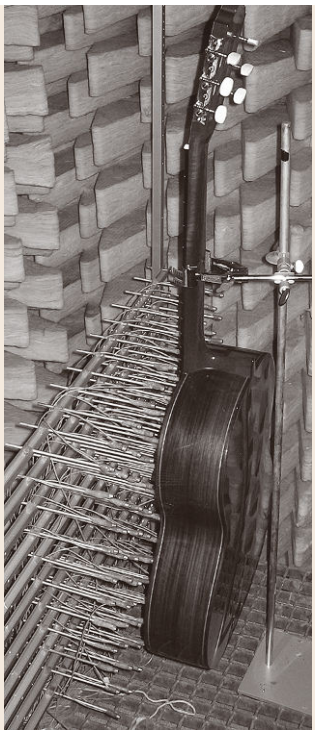


Fig. 5.3 Experimental arrangement of a near-field microphone array used to measure the radiation pattern of a guitar (after [5.18]). The array is positioned approximately 3 cm from the top plate of the guitar

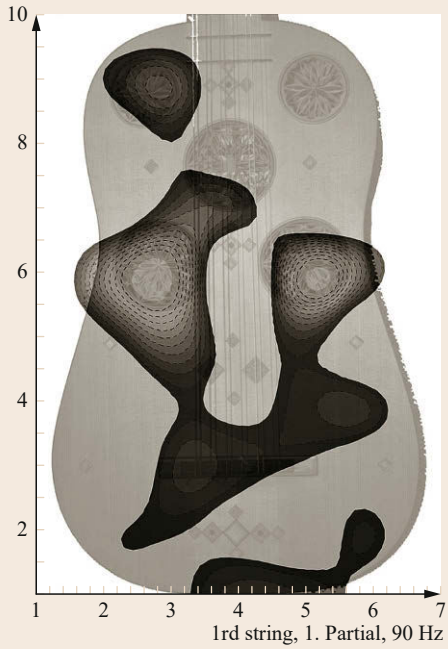


Fig. 5.4 Example of a radiation pattern calculated for a Yaish vihuela using the minimum energy method (after [5.18]). The experimental arrangement is shown in Fig. 5.3

radiating surface, one can assume the radiation is represented by an array of monopoles that are arranged so that one is under each microphone of the array. The array is placed in the near field of the radiation to achieve subwavelength resolution. The radiation pattern can then be calculated using minimum-energy methods to determine the amplitude of the monopole radiators [5.19]. An example of a typical experimental arrangement for recording the near-field radiation is shown in Fig. 5.3 and the reconstruction of the radiation pattern of a Yaish vihuela using this technique is shown in Fig. 5.4 [5.18].

5.2 Measurement of Deflection

Many musical instruments produce sound by vibration of some portion of the instrument itself. Percussion instruments are the most easily recognizable instruments that fall into this category, however, the strings of all acoustic stringed instruments are connected to some type of moving membrane or plate that efficiently transfers the energy from the string to the air. Examples include piano soundboards, violin and guitar top plates and backs, and the membrane head of the banjo. When investigating these instruments it is important to understand the motion of the individual parts and therefore visualizing the deflection shape is often necessary. Several methods for visualizing deflection shapes are addressed below.

5.2.1 Chladni Patterns

The most simple method of investigating the deflection shape of a vibrating object is to sprinkle sand on it while it is in motion. This method was pioneered by Ernst Chladni over 200 years ago and it still makes an impressive demonstration [5.20]. These so-called *Chladni patterns* do not make the deflection visible, but because the sand migrates to the nodal regions it makes the nodes visible. When the pattern of the motion is simple, observing the Chladni pattern is an effective method of determining the deflection shape. Indeed, if one can be sure that the motion represents a modal pattern, this technique is often adequate to determine

which mode is being excited if the mode shapes are known from theoretical considerations. Several examples of Chladni patterns can be found in a classic book by *Waller* [5.21]. An example of a Chladni pattern showing one of the modal patterns of a square flat plate is shown in Fig. 5.5 [5.22]. Note that it is also possible to use powder instead of sand to create the pattern, but in so doing the antinodes become visible because the powder is attracted to them instead of the nodal regions [5.23].

Although Chladni patterns have not been commonly used in the investigation of musical instruments since the advent of more precise optical methods, there are examples in the literature of the technique being used effectively. For example, the first few deflection shapes of a piano soundboard are shown using Chladni patterns in [5.24], and investigations of violin top plates and mode shapes in drum heads can also be found in the literature [5.25, 26]. However, there are limitations to using Chladni patterns for analysis of musical instruments.

The most obvious limitation of using Chladni patterns for scientific investigations is that the part of the instrument being investigated must be flat and horizontal. This means that, for example, only unstrung piano soundboards and violin top plates that are not mounted to an actual instrument can be investigated. Furthermore, if the instrument can be damaged by the sand it is not an economically efficient technique. Finally, there is no indication of the relative displacement of nonnodal regions. Chladni patterns only indicate the position of nodal regions, which is often not enough information to be useful. For these reasons Chladni patterns are seldom used for scientific investigation, although they are often used for demonstration purposes. Usually, more sophisticated techniques are employed, such as those discussed below.

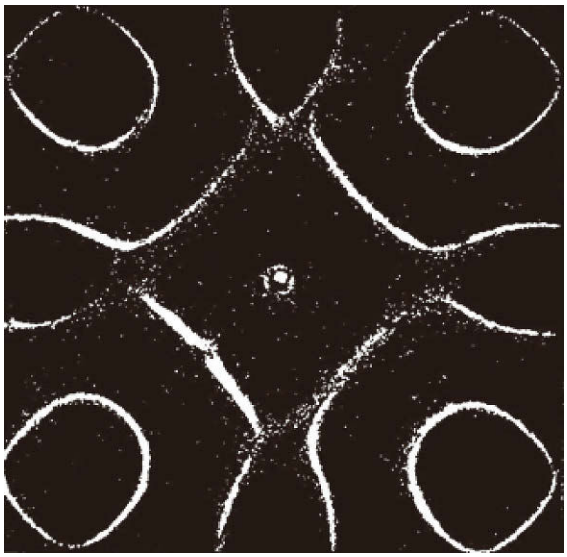


Fig. 5.5 Photograph of a Chladni pattern of one mode of a vibrating flat plate. The white sand has migrated to the nodal lines making them visible

5.2.2 Holographic Methods

With the invention of the laser in the mid-twentieth century several optical methods became available that are of use to musical acousticians. Most of these have been refined over the past 50 years and are now commonly used and easily implemented, but probably the most useful techniques that are of specific interest to musical acousticians are the processes involving optical interferometry. Although interferometry is possible without using lasers, the high coherence, directionality and power afforded by using a laser has made it the only source used for interferometry in modern research.

Interferometers rely on the principle of superposition to glean information from an optical signal. Somewhere in every interferometer two optical waves are combined and the interference between them is used to determine some aspect of the object under study. If the object under study is not transparent, one of the optical signals must be reflected from the object, which is the usual case for a musical instrument. In this case, the information that can be derived from the interferometer is the magnitude of the displacement or velocity of a point on the object.

There are numerous types of interferometers that can be applied to the study of musical instruments, but the most useful are those that provide a two-dimensional image of the surface vibration. In this section we discuss holographic interferometry, an interferometric method that can be used to produce such images. Two other interferometric techniques, electronic speckle pattern interferometry (ESPI) and laser Doppler vibrometry (LDV), will be discussed in following sections.

Holographic interferometry is used to determine the part of an object that is moving as well as the amplitude of the motion. It produces a two-dimensional image of the motion and has a sensitivity that is on the order of a fraction of the wavelength of the light being used. Therefore, amplitudes exceeding 0.01 mm are normally difficult to image using holographic techniques.

Because of the sensitivity, holographic interferometry is of little use in analyzing the motion of a musical instrument while being played because the amplitude of vibration typically exceeds the maximum resolvable displacement. Additionally, the act of playing the instrument produces whole-body motion that normally

exceeds the limitation of the resolution. Therefore, holographic methods are typically used under controlled circumstances with the instrument mounted securely and vibrations driven by some mechanism that produces very small amplitudes of the motion. For example, it is common to use the sound from a loudspeaker or a small piezoelectric element to drive the oscillations. Several different methods of holographic interferometry are addressed below.

Holographic Interferometry

Holographic interferometry has been used extensively to image the deflection shapes of many musical instruments, including bells [5.27], drums [5.28], guitars [5.29], clarinet reeds [5.30] and even wine glasses [5.31]. The process relies on the fact that when a holographic image of an object is superimposed on an image of the object being illuminated by light from the same laser, the final image is a superposition of the two images. Because the light from both images is coherent, any variation in position of the object that occurred after the creation of the hologram will result in an interference pattern. The lines of interference represent contours of equal displacement. The basic principles of holographic interferometry are described below, however, an extensive review of the physics involved can be found in [5.32].

The process of holographic interferometry begins with making a hologram of the object of interest. A typical arrangement for making a hologram is shown in Fig. 5.6. Light from a laser is split into two beams, which we refer to as the illuminating beam and the reference beam. The illuminating beam is directed toward the object of interest and the light reflected from the object falls on a photographic plate. We will refer to this reflected light as the object beam.

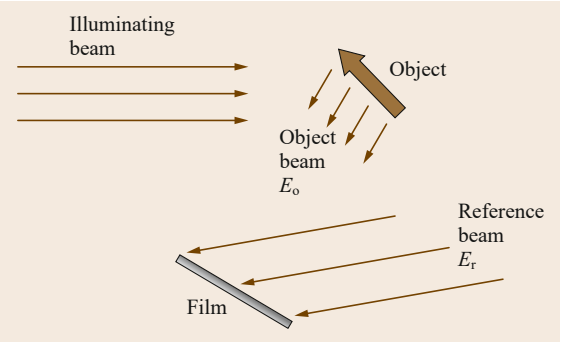


Fig. 5.6 Schematic diagram of the arrangement for making a hologram. The illuminating beam and the reference beam interfere on the film to produce a diffraction pattern, which can later be used to reconstruct an image of the object. Both beams originate from the same laser

The reference beam is also directed toward the film, where the two coherent beams interfere. The interference pattern is recorded on the film, which when developed results in a series of closely spaced dark fringes. These fringes form a diffraction pattern that when illuminated by the original laser produces a holographic image of the object. If the film is replaced in the same position in which it was originally illuminated, then the holographic image appears at the position of the original object. Therefore, imaging the object through the hologram results in the formation of two superimposed images of the object. Since the light forming the two images is coherent, a steady-state interference pattern occurs that can be photographed as a real image. If the object has not moved and the film is replaced in exactly the same place then there will be no visible fringes, however, any movement of the object will result in fringes representing contours of equal displacement.

To understand the process in more detail consider the formation of the hologram on the photographic plate. Assuming that the reference beam is of uniform intensity and linearly polarized, light from the reference beam falling on the photographic plate can be represented by the electric field

$$E_r = A_r e^{i(\boldsymbol{k}_r \cdot \boldsymbol{r})} , \tag{5.16}$$

where $\boldsymbol{k}_r$ is the wave vector of the light, $\boldsymbol{r}$ is the spatial coordinate, A_r is the amplitude, and the time-varying optical frequency has been suppressed because all of the light is coherent and assumed to have the same frequency. The electric field of the light reflected from the object and incident on the same position on the film can be written as

$$E_o(\xi) = A_o(\xi) e^{i(\boldsymbol{k}_o \cdot \boldsymbol{r} + \phi(\xi))} , \tag{5.17}$$

where ξ is the spatial coordinate of a point on the film, $\phi(\xi)$ represents the phase shift due to the different path lengths from the object to the plate, A_o is the amplitude, and $|\boldsymbol{k}_o| = |\boldsymbol{k}_r|$. Since these two fields are incident on the plate at the same position they interfere and produce an intensity pattern that is recorded by the photographic emulsion. Once the latent image is developed the plate has a transmission function given by

$$T = \beta \left| A_r + A_o(\xi) e^{i\phi(\xi)} \right|^2 , \tag{5.18}$$

where β is a constant determined by the response of the photographic emulsion to the light. The plate is a diffraction grating with the grating spacing determined by the relative phases of the two incident beams at the position of the film.

By replacing the plate with the interference pattern into its original position the intensity of the diffraction grating is imposed on the reference beam forming a field given by

$$E_{\mathrm{d}}(\xi) = A_{\mathrm{r}} \mathrm{e}^{\mathrm{i}(\boldsymbol{k}_{\mathrm{r}} \cdot \boldsymbol{r})}(1 - T) \,, \tag{5.19}$$

which can be written as

$$E_{\mathrm{d}}(\xi) = A_{\mathrm{r}} \mathrm{e}^{\mathrm{i}(\boldsymbol{k}_{\mathrm{r}} \cdot \boldsymbol{r})} \left\{ 1 - \beta \left[A_{\mathrm{r}}^2 + A_{\mathrm{o}}^2(\xi) + A_{\mathrm{r}} A_{\mathrm{o}}(\xi) \mathrm{e}^{\mathrm{i}([\boldsymbol{k}_{\mathrm{r}} - \boldsymbol{k}_{\mathrm{o}}] \cdot \boldsymbol{r} - \phi(\xi))} + A_{\mathrm{r}} A_{\mathrm{o}}(\xi) \mathrm{e}^{-\mathrm{i}([\boldsymbol{k}_{\mathrm{r}} - \boldsymbol{k}_{\mathrm{o}}] \cdot \boldsymbol{r} - \phi(\xi))} \right] \right\} . \tag{5.20}$$

When (5.20) is expanded it becomes obvious that the fifth term is

$$E_{\mathrm{h}}(\boldsymbol{\xi}) = \beta A_{\mathrm{r}}^2 A_{\mathrm{o}}(\boldsymbol{\xi}) \mathrm{e}^{\mathrm{i}(\boldsymbol{k}_{\mathrm{o}} \cdot \boldsymbol{r} + \phi(\xi))} \,, \tag{5.21}$$

which is a replica of the image of the object multiplied by a constant. This is the holographic image. The other terms represent light with different wave vectors and therefore the light attributable to these terms is traveling in other directions and these beams are easily distinguished from the hologram.

If the original object has not been moved, the light from the object and the light diffracted from the hologram produce two identical superimposed images of the object. However, any displacement of the object results in a change in phase of the object beam, which can be represented by a phase shift at the film plane with spatial dependence, $\varphi(\boldsymbol{\xi})$. The object beam can then be written as

$$E'_{\mathrm{o}}(\boldsymbol{\xi}) = E_{\mathrm{o}}(\boldsymbol{\xi}) \mathrm{e}^{\mathrm{i}(\boldsymbol{k}_{\mathrm{o}} \cdot \boldsymbol{r} + \phi(\xi) + \varphi(\boldsymbol{\xi}))} \,. \tag{5.22}$$

The resulting image can be determined by the superposition of this wave with the holographic image E_{h}, which results in an intensity pattern showing interference fringes given by

$$I_{\mathrm{i}}(\boldsymbol{\xi}) = \eta\{1 + \cos(\varphi(\xi))\} \,, \tag{5.23}$$

where η is a constant.

By imaging the object and its holographic image simultaneously, the phase difference between the two can be related to the displacement of the object and the two-dimensional coordinate ξ can be mapped onto the surface of the object. The phase difference is then a function of the displacement of the object and $\varphi(\xi)$ is given by

$$\varphi(x, y) = |\boldsymbol{k}_{\mathbf{o}}| \Delta(x, y)[\cos(\theta_{\mathrm{i}}) + \cos(\theta_{\mathrm{r}})] \,, \tag{5.24}$$

where $\Delta(x, y)$ is the amplitude of the displacement from the original position at the point (x, y) on the surface of the object and θ_{i} and θ_{r} are the angles of incidence and reflection from the object respectively. Contour lines represent subsequent displacements of

$$\Delta_{\mathrm{c}} = \frac{\lambda_{\mathrm{L}}}{\cos(\theta_{\mathrm{i}}) + \cos(\theta_{\mathrm{r}})} \,, \tag{5.25}$$

where λ_{L} is the wavelength of the light.

It is important to note that replacement of the holographic plate into its original position is critical, and rather difficult. Similarly, the entire arrangement must be isolated from ambient motion such that the magnitude of any motion of the object or film is significantly less than the wavelength of the light. This normally requires active isolation of the entire system. Thermal stability is also important or expansion and contraction of the object will create carrier fringes that make it difficult or impossible to determine the motion of interest. However, with the proper equipment and patience holographic interferometry produces high-quality interferograms that unambiguously indicate minute motion of the object of interest.

From the derivation above it is clear that small displacements of the object will be represented by fringes of equal displacement on the final image, however, the static displacement of an object is rarely important when investigating musical instruments. Instead, identifying the deflection shape of a vibrating object is usually of interest. Therefore, we must consider the case of an object vibrating with an angular frequency of ω and period $T = 2\pi/\omega$. These vibrations are typically driven externally by a mechanical component connected to a signal generator so that the point of excitation can be unambiguously determined. However, acoustical excitation by sound from a loudspeaker is useful if it is desirable not to have single-point excitation; excitation at a single point will not excite modes with nodes at the point of contact. To investigate the harmonic motion of musical instruments there are three holographic methods that are commonly used to image the harmonic displacement of an object: double-pulsed, time-averaged, and real-time.

Double-Pulsed Holographic Interferometry

The method of double-pulsed holographic interferometry is relatively easy to understand. When using this technique a hologram is recorded using a pulsed laser with a pulse length that is short compared to the period of oscillation of the object. At some later time, typically $T/2$, another laser pulse records a second latent image on the film. When the hologram is developed and reconstructed, two superimposed images appear. The two images interfere producing fringes representing the displacement of the object that occurred between the

two laser pulses. The contour lines are separated by a distance given by (5.25).

The primary advantage of double-pulsed holographic interferometry is that it is not necessary to replace the developed film back into the same position that it was when the hologram was constructed. Also, the requirements for isolation from ambient vibrations are significantly reduced since the object only needs to be stable in the time period between the two pulses. Finally, once the image is recorded and developed it can be analyzed at a later time and in a different location. Any continuous-wave laser having the same wavelength as the recording laser will produce the interferogram that can be used for analysis.

Unfortunately, the disadvantages of double-pulsed holography for the experimentalist are significant. Pulsed lasers are typically much more expensive than continuous-wave lasers, the timing and triggering mechanisms can be complicated, and having a continuous-wave laser of the same wavelength is convenient for viewing the interferogram. For these reasons it is more common to use time-averaged or real-time holographic interferometry to investigate musical instruments.

Real-Time Holographic Interferometry

Real-time holographic interferometry uses the arrangement shown in Fig. 5.6, with the interferogram being produced by the superposition of an image of the object and the hologram that results from placing the film plate back in its original position. This is similar to the arrangement of the more common holographic interferometer described above. However, once the holographic plate has been placed back into its original position the object is made to undergo harmonic motion with angular frequency ω. This produces an oscillating interference pattern with a period T as the object executes the motion. Obviously nodal areas undergo no such motion and the magnitude of the displacement will vary with position. The method of inducing the oscillation can be either mechanical or acoustical stimulation depending on the information desired.

Once the object is oscillating, the phase of the object beam given by (5.24) becomes time dependent and the phase difference between the holographic image and the real-time image of the object is given by

$$\varphi(\xi, t) = \varphi(\xi)\sin(\omega t)\,. \tag{5.26}$$

Under this condition, (5.22) becomes

$$E_o'(\xi, t) = E_o(\xi)e^{i(\boldsymbol{k}_o\cdot\boldsymbol{r}+\phi(\xi)+\varphi(\xi)\sin(\omega t))} \tag{5.27}$$

and the electric field is given by the superposition of E_h and E_o', where E_h is still given by (5.21). The image created by this superposition is then viewed by eye, or recorded by either film or a charge-coupled device (CCD) array.

The intensity pattern created by the superposition of the two fields, which is recorded by either the eye or other means is given by

$$I(\xi, t) = \eta\left\{|E_o(\xi)|^2 + |E_h(\xi)|^2 + E_o(\xi)E_h(\xi)\cos[\varphi(\xi)\sin(\omega t)]\right\}. \tag{5.28}$$

To perform real-time holographic interferometry it is necessary that the recording medium has an integration time that exceeds several periods of vibration. This is easily achieved since the recording medium is usually either the eye, a CCD array, or film and the oscillations of interest typically exceed 100 Hz. The resulting irradiance that will produce the visible pattern that will be the final image is then given by

$$\langle I_i(\xi)\rangle = I_o(\xi) + I_h(\xi) + \sqrt{I_o(\xi)I_h(\xi)} \times \frac{1}{T}\int_0^T \cos[\varphi(\xi)\sin(\omega t)]\mathrm{d}t\,, \tag{5.29}$$

where $\langle I_i(\xi)\rangle$ represents the time-average irradiance of the image.

The integral in (5.29) is one method of calculating the zero-order Bessel function of the first kind, therefore, the time-averaged image can be written as

$$\langle I_i(\xi)\rangle = I_o(\xi) + I_h(\xi) + \sqrt{I_o(\xi)I_h(\xi)}J_0(\varphi(\xi))\,, \tag{5.30}$$

where $\varphi(\xi)$ is given by (5.24). Maximum contrast occurs when $I_r = I_o$, in which case the intensity of the interferogram is given by

$$I_i = 2I_o(\xi)[1 + J_0(\varphi(\xi))]\,. \tag{5.31}$$

As with static holographic interferometry, it is necessary to replace the hologram precisely in the position it was in when the hologram was recorded. This is a difficult process. However, the advantage to real-time holographic interferometry is that the frequency of oscillation of the object, ω, can be changed while the interferogram is being observed. Therefore, one can view the deflection shape of the object vibrating at different frequencies. If information at only a single frequency is of interest, and that frequency is known, it is not necessary to go to the trouble of repositioning the holographic plate. Instead, the process known as time-averaged holographic interferometry is useful.

Time-Averaged Holographic Interferometry

Time-averaged holographic interferometry uses the same arrangement as double-pulsed interferometry; that

is, the hologram is recorded and processed and available for viewing at any later time. However, instead of using two pulses to record the holographic image at two different displacements, the image is recorded while the object is undergoing harmonic motion. That is, the object is set into harmonic motion and the hologram is recorded over a time that is long compared to the period of oscillation of the object. Using the same procedure described above for deriving (5.30), it is straightforward to show that under these conditions (5.31) becomes

$$I_{\mathrm{i}} = I_{\mathrm{o}}(\xi)[J_0(\varphi(\xi))]^2 \,. \tag{5.32}$$

The primary disadvantage of time-averaged holographic interferometry is that you must know the frequency of interest before recording the hologram. It is also important that the amplitude of vibration be small enough such that the fringes of equal displacement are visible. In real-time holographic interferometry the frequency and amplitude of vibration can be adjusted while viewing the interferogram. However, there are significant advantages to time-averaged holographic interferometry. These include the ease of reconstruction, which does not require precise replacement of the recording medium, and the improved contrast. After the displacement of the object exceeds the distance necessary to produce the first fringe, the contrast of the real-time interferogram is reduced significantly. Only the intensity between the fringes is reduced in time-averaged interferometry. Because the intensity of the dark fringes in the time-averaged case is always zero, the contrast between light and dark fringes remains high. An example of a holographic interferogram showing one deflection shape of a vibrating guitar body can be seen in Fig. 5.7 [5.33] .

The advantages of holographic interferometry were identified quite early by acousticians [5.34]. However, as noted above, the process is tedious, time consuming, and requires excellent stability and extensive optical equipment. For standard and real-time holographic interferometry it also requires proper alignment of the holographic plate back into its original position. For these reasons holographic interferometry has not been as prevalent in the study of musical instruments as one might think. However, there is another technique that produces similar results and requires less time, expense and effort. We address this technique in the next section.

5.2.3 Electronic Speckle Pattern Interferometry (ESPI)

Electronic speckle pattern interferometry (ESPI) is a technique that produces results similar to real-time holographic interferometry, but circumvents many of the problems. Of particular importance is the elimination of the necessity of producing and replacing the holographic plate.

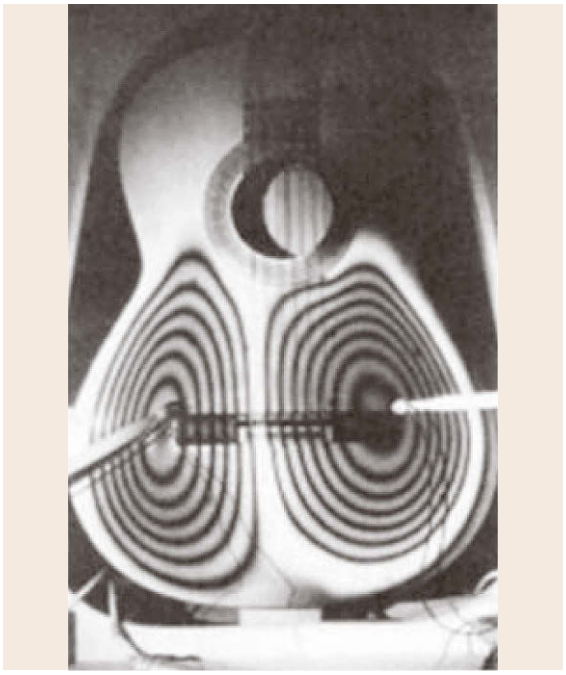


Fig. 5.7 Time-averaged holographic interferogram of a vibrating guitar body (after [5.33]). The fringes indicate contours of equal displacement. The vibrations are driven by a coil placed near a small magnet attached to the guitar

An electronic speckle pattern interferogram of a vibrating guitar body produced using ESPI is shown in Fig. 5.8. The similarity of the images, which is evident by comparing Figs. 5.7 and 5.8, causes the process of ESPI to be incorrectly referred to by the name *TV holography*. However, although the image appears similar to one produced by real-time or time-averaged holographic interferometry, and it can be viewed on a television monitor or computer screen, ESPI does not involve the production of a diffraction grating (i. e., a hologram). Therefore, the process is fundamentally different from holography.

Although holographic interferometry is still used for analysis in some laboratories, it has largely been replaced in the study of musical instruments by ESPI. The process of speckle pattern interferometry was first developed using film as the recording mechanism, but since the process involves subtracting images, film was quickly replaced by analog and then digital video. Currently, almost all ESPI is performed using a digital CCD array to capture the images, with image subtraction performed digitally in a computer. Electronic speckle pattern interferograms have been used to study

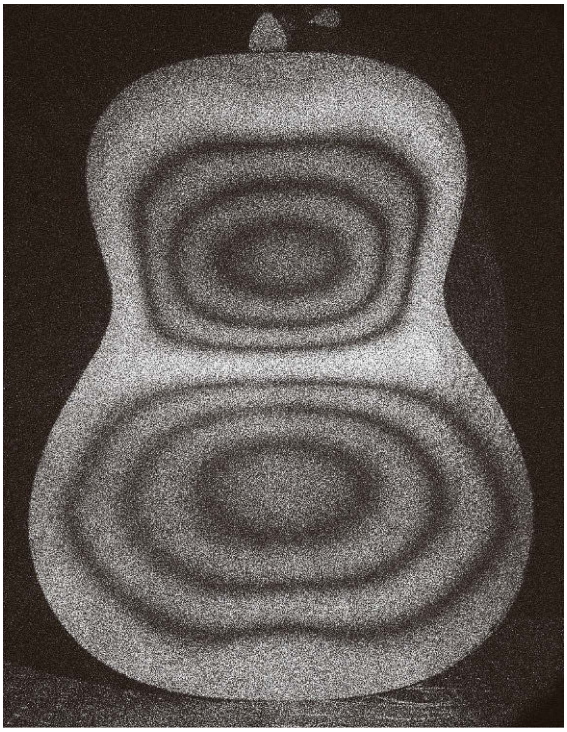


Fig. 5.8 Electronic speckle pattern interferogram of a vibrating guitar body showing fringes of equal displacement. The vibrations are driven by a speaker located approximately 1 m from the guitar

steel pans [5.35], banjos [5.36], trumpets [5.37], several percussion instruments [5.38–41], and even piano soundboards [5.42].

There are several different ways to construct an ESPI system, with each variation having distinct advantages and disadvantages. However, an understanding of the process can be gained by considering one of the simplest arrangements, a schematic diagram of which is shown in Fig. 5.9. Light from a laser is divided into two beams using a beam splitter, which is normally a semisilvered mirror that reflects a portion of the light while transmitting the rest. One of the beams is directed toward the object of interest off of which it reflects, and the other is directed toward a diffusing screen. As with holographic interferometry, the beam reflected from the object is usually referred to as the object beam and the latter as the reference beam. An image of the object is captured by a digital camera by viewing it through a second semisilvered mirror in such a way that the image of the object is superimposed on the image from the diffusing screen. The signal from the camera is then sent to a computer for processing.

Because the light used to illuminate the object and screen is highly coherent, the image projected onto the

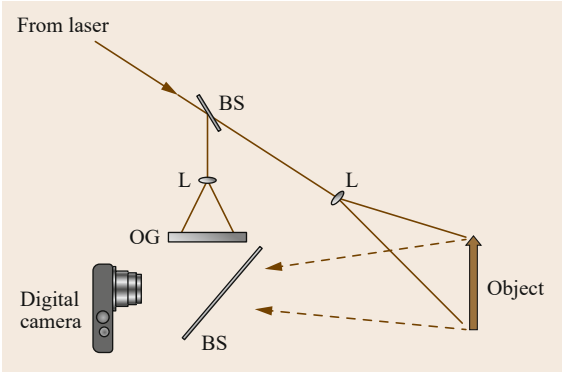


Fig. 5.9 Schematic of a simple arrangement for electronic speckle pattern interferometry with components: L = lens, BS = beam splitter, OG = opal glass. The lens on the camera is focused on the object; the opal glass plate provides the necessary reference beam, which reaches the camera by reflection from the beam splitter

recording array is speckled due to the diffraction from the edges of the aperture of the lens. The variation of this *objective speckle pattern* is complex, but the mean diameter of the speckle is approximately given by

$$d \approx (1 + M)\lambda_{\mathrm{L}} F \,, \tag{5.33}$$

where λ_{L} is the wavelength of the light from the laser, M is the magnification of the image and F is the aperture ratio of the lens (i. e., the f-number). It is normally advantageous to ensure that the mean diameter of the speckle is approximately the same size as the pixels that make up the recording array. For a thorough review of speckle within the context of ESPI, see [5.43].

To understand ESPI it is useful to consider only one pixel on the recording array, on which light from the reference beam with irradiance I_{r} interferes with light reflected from the object with irradiance I_{o}. The irradiance of the light on the pixel is then given by the usual form for interference between two coherent beams, i. e.,

$$I_1 = I_{\mathrm{r}} + I_{\mathrm{o}} + 2\sqrt{I_{\mathrm{r}} I_{\mathrm{o}}}\cos(\phi) \,, \tag{5.34}$$

where the phase angle ϕ is due to the differing optical path lengths of the two beams from the laser to the image plane. This image is recorded and digitally stored.

At some later time, after the portion of the object that is imaged onto the pixel has been translated a distance Δ, the irradiance on the image plane is given by

$$I_2 = I_{\mathrm{r}} + I_{\mathrm{o}} + 2\sqrt{I_{\mathrm{r}} I_{\mathrm{o}}}\cos(\phi + \varphi) \,, \tag{5.35}$$

where φ is given by

$$\varphi = 2\Delta|\boldsymbol{k}_{\mathbf{o}}| \,. \tag{5.36}$$

This image is also then digitally stored.

The ESPI process involves digitally subtracting the two images represented by (5.34) and (5.35) and taking the absolute value. The resulting image can then be displayed on a computer monitor. After subtraction the displayed pixel intensity is given by

$$I_{1,2} = \eta|\cos(\phi+\varphi)-\cos(\phi)|\,, \tag{5.37}$$

where η is a constant determined by the relative intensities of the two beams and the display parameters of the monitor. It is obvious from (5.37) that when $\varphi = 2n\pi$, where n is an integer, then the displayed intensity of the pixel will be zero. Similarly, when $\varphi = (2n+1)\pi$ the displayed intensity will be maximum. Therefore, the displacement of the object at the point imaged onto the pixel can be determined. Unfortunately, an image of the entire object is comprised of many pixels, all with a different value of ϕ. That is, because the value of ϕ varies from pixel to pixel in a random manner it is not obvious that the entire image will result in a meaningful interferogram. However, it can be shown that if ϕ is a random variable, on average the maximum irradiance will occur when $\varphi = 2n\pi$ and the minimum will occur when $\varphi = (2n+1)\pi$. A proof can be found in Appendix E of [5.32].

Due to the high calculation speed of modern computers it is possible to perform the image subtraction in almost real time. Therefore, it is possible to record an image of the object and then subtract subsequent images in real time and observe interferograms of the changing motion. If the displacement of the object occurs on a timescale that is long compared to the integration time of the camera, then this is a useful technique for determining displacements on the order of the wavelength of the light. Typically excellent interferograms can be achieved for motions ranging from $\approx 0.25{-}5\,\mu\text{m}$ if the light from the laser is in the visible portion of the electromagnetic spectrum.

If after storing an image of the static object the object begins oscillating with a period that is small compared to the integration time of the detector, as is common when studying musical instruments, then the second recorded image is the result of the time average over several oscillations. The process of determining the irradiance on the pixel in this case is similar to that outlined in Sect. 5.2.2, *Real-Time Holographic Interferometry*. By integrating the oscillating image over a large number of periods the displayed intensity of the subtracted images is given by

$$I_{1,2} = \eta[1-J_0(\varphi)]\,. \tag{5.38}$$

This response is similar to the response of real-time holographic interferometry with the exception that nodal areas show as black instead of white.

As with holographic interferometry, the stability requirements for ESPI are stringent. The object and all optical components must be stable to within a fraction of the wavelength of the light or the two speckled images become uncorrelated and the interference fringes are not visible. This normally requires active isolation of the object and the optical system, or it requires that the images be captured and stored within a time that is short compared to the timescale of the ambient motion.

Although this method of ESPI has been used in the past to study musical instruments, and it is useful to explain the basic principles of ESPI, there are better arrangements for acoustical studies. It is important to note one arrangement that is of particular interest to musical acousticians because it affords the possibility of imaging large objects that are typically not stable enough to analyze using either ESPI or holographic interferometry. This method, known as *decorrelated electronic speckle pattern interferometry* (DESPI), requires a linear phase shift of one beam during the process of recording the image. The phase shift can be caused by whole-body motion of the object that is linear on the timescale of the integration of the detector, effectively inducing a linear phase shift on the illuminating beam. Alternatively, the motion can be imposed on one beam by the linear motion of a mirror that reflects either the reference or illumination beam, or by introducing an electrical or mechanical phase-shifting device in one beam. When using DESPI it is not necessary to store an image of the object prior to the onset of vibration. Instead, two images of the object that were captured during vibration are subtracted. This significantly relaxes the restriction on the magnitude of the ambient motion and often eliminates the need for active isolation of the system. An analysis of decorrelated ESPI can be found in [5.44].

The contrast of the interferogram obtained using DESPI is dependent upon how rapidly the phase of the beam is changing, either due to motion of the object or an imposed path length difference on one beam, and the integration time of the detector. But there is a wide tolerance for motion and the imposition of this motion can enhance both the contrast and the precision of the interferogram. A comparison of two images of a vibrating plate is shown in Fig. 5.10, where the image obtained through the ESPI system described above is shown in Fig. 5.10a and the same plate imaged using DESPI is shown in Fig. 5.10b [5.44].

In the example shown in Fig. 5.10b the linear motion was imposed on one of the beams of the interferometer and was carefully controlled to produce maximum contrast. However, in many cases ambient motion can produce the same result. Indeed, the ambient motion that normally limits the usefulness of ESPI

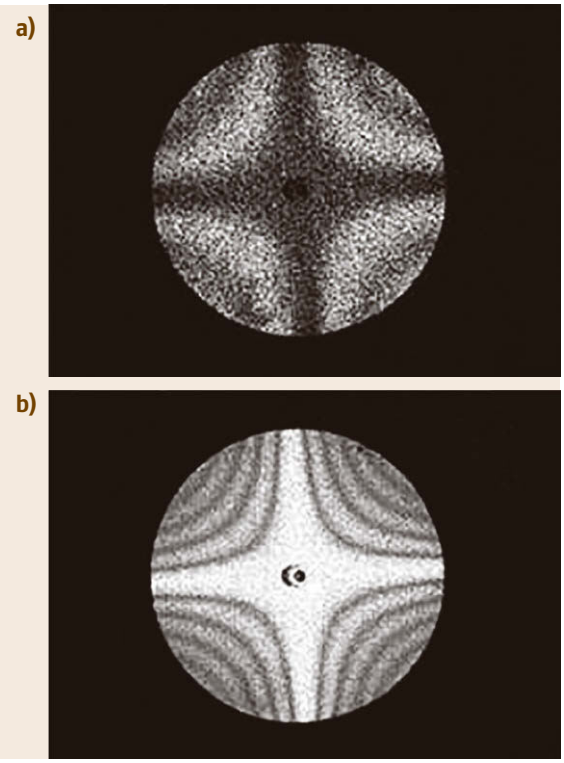


Fig. 5.10a,b Interferograms of a vibrating circular plate using (a) a simple ESPI system and (b) the same plate using DESPI (after [5.44]). Both plates have the same amplitude and frequency of vibration

for large objects can be used to produce excellent interferograms using DESPI. In such cases the integration time of the detector can be used to ensure that the magnitude of the ambient motion is sufficient to decorrelate the illuminating and reference beams. An example is shown in Fig. 5.11, where the deflection shape of a vibrating piano soundboard is made visible using DESPI. A similar image using ESPI would be extremely difficult to produce.

The contrast of interferograms produced by ESPI and DESPI is not as high as those produced by holographic interferometry. The speckle nature of the image is also more pronounced. However, usually the quality of the image is sufficient to obtain the same information that can be obtained from a time-averaged holographic interferogram. For example, compare the image of a vibrating guitar obtained by DESPI shown in Fig. 5.8 with a similar one obtained through time-averaged holographic interferometry shown in Fig. 5.7.

While ESPI is becoming more common in the analysis of musical instruments, the cost of a commercial ESPI system is significant. However, the system shown in Fig. 5.9 is not difficult to construct if one has experience with optical experimental techniques, and the cost of the components can be as little as one-tenth that of a commercially available system. Those wishing to build a simple ESPI system at relatively low cost should consult [5.45, 46]. An even more simple and inexpensive variation is described in [5.22], but this arrangement produces inferior results and is primarily intended for educational use.

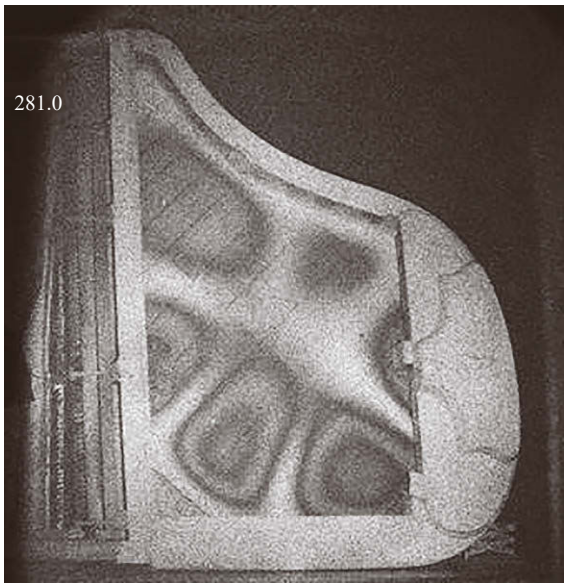


Fig. 5.11 DESPI interferogram of a piano soundboard excited acoustically by a distant speaker. The ambient motion of the piano was sufficient to decorrelate the illuminating and reference beams without imposing a controlled phase shift as part of the experimental arrangement

Note that the recording mechanism for speckle pattern interferometry does not necessarily need to be electronic. It can be accomplished using film rather than an electronic camera as the recording mechanism. Unfortunately, the effort required to use film is similar to that necessary for holographic interferometry and significantly greater than what is required when using an electronic recording mechanism. The prevalence of low-cost electronic cameras and computers has rendered film-based speckle pattern interferometry largely obsolete.

5.2.4 Laser Doppler Vibrometry

Another optical technique that is not normally classified as interferometry, but in detail relies on the physics of interfering optical waves, is laser Doppler vibrometry (LDV). LDV utilizes the frequency shift of light reflected from a moving source to determine the velocity of a vibrating surface. Although the Doppler effect

is most often associated with sound, electromagnetic waves can also experience Doppler shifts.

LDV uses light from a laser that is reflected from the point of interest on a moving object. This reflected light is mixed with the original light to produce a moving interference pattern. The speed of the fringe motion is proportional to the velocity of the object from which the beam is reflected. The laser is usually colocated with the receiver and the associated electronics are sealed with the optics into a case.

Laser Doppler vibrometry is a logical extension of laser Doppler velocimetry, which was in use as early as 1964 [5.47]. Laser Doppler velocimetry relies on the Doppler shift of a moving surface or particles in a fluid to determine the velocity of the body or fluid. Laser Doppler vibrometry relies on the same effect, but the object of interest is a harmonically vibrating solid surface and Fourier techniques can be used to isolate the vibrations within a given frequency band. LDV has been used in studying percussion instruments, [5.39, 48] the bodies of stringed instruments, [5.49] vibrations of the bells of brass instruments [5.50], and the motion of organ pipes [5.51].

To understand the operation of a laser Doppler vibrometer one must begin with the realization that the light from a laser has a narrow linewidth. Due to the narrow linewidth it is acceptable to assume that the laser has a frequency of ν_L and a wavelength of λ_L. In reality all lasers have a finite linewidth, which limits the coherence of the light and hence the distance between the object and the laser. The limitations associated with the finite coherence length can be found elsewhere and is a detail that is not necessary to understand the operation of a LDV system [5.52].

In a LDV system the light from a laser is divided into two beams by a semisurfaced mirror. One beam is directed toward the object of interest. The second beam is directed toward a second beam splitter where it is mixed with the light reflected from the object. The resulting interference pattern is imaged onto a detector. If the object is moving toward the laser then the reflected beam will have a shorter wavelength than the incident object beam because each successive phase front must travel a shorter distance than the last one.

Let the period of the electromagnetic wave incident on the object be given by $T = 1/\nu_\mathrm{L}$. A wavefront from the laser will move a distance $\lambda_\mathrm{L} = c_0 T$ during one period, where c_0 is the speed of light in air. However, during the time T an object moving at velocity $\boldsymbol{v}$ toward the laser will have moved a distance of $|\boldsymbol{v}|T$ and therefore the distance between two successive wavefronts of light reflected off of the moving object will be

$$\lambda' = c_0 T - 2|\boldsymbol{v}|T\,, \tag{5.39}$$

and the frequency of that light will be Doppler shifted to

$$\nu' = \frac{c_0}{\lambda'}\,. \tag{5.40}$$

Combining equations (5.39) and (5.40) yields

$$\nu' = \frac{\nu_\mathrm{L}}{1 - \frac{|\boldsymbol{v}|}{c_0}}\,. \tag{5.41}$$

The velocities of interest for the musical acoustician are normally on the order of 1 m/s or less. Therefore, the task becomes determining a shift in frequency of approximately one in 10^8. To do this the reflected light is mixed with the reference beam at the second beam splitter. When combined at the second beam splitter, the reference beam of frequency ν_0 and the reflected beam at frequency ν' create a changing interference pattern with a frequency of $\Delta\nu = \nu - \nu'$, or

$$\Delta\nu = \frac{-2|\boldsymbol{v}|}{\lambda_\mathrm{L}}\,. \tag{5.42}$$

If the laser is in the near-infrared region, as many are, it will have a wavelength of approximately 1 μm and therefore a typical difference frequency may be on the order of a few megahertz, which is easily detected by a photodiode.

The simplified description above is adequate to understand how a LDV system works, however, there is one other aspect that deserves mention. It is clear from (5.42) that in the situation described there is no way to determine if the object is moving toward or away from the laser. In either case the difference frequency would be the same. To solve this problem, and to increase the signal-to-noise ratio, one of the beams is usually frequency modulated by a Bragg cell so that there is a fixed carrier frequency from which $\Delta\nu$ can be added or subtracted.

As described above, this system can be used to measure the velocity of any object and hence describes laser Doppler velocimetry. However, since the motion of musical instruments is almost always periodic, the analysis can be extended to assume that the displacement of the point of interest is given by

$$x = \Delta_x \sin(\omega t)\,, \tag{5.43}$$

where Δ_x is the amplitude of the motion and ω is the acoustic angular frequency. The velocity is then given by

$$v(t) = \frac{\mathrm{d}x}{\mathrm{d}t} = \Delta_x \omega \cos(\omega t)\,, \tag{5.44}$$

and the Doppler shift is given by

$$\Delta\nu(t) = \frac{-2\Delta_x\omega}{\lambda_L}\cos(\omega t)\,. \tag{5.45}$$

A fast Fourier transform algorithm can then be used to determine the velocity as a function of frequency of vibration, in which case the process is termed laser Doppler vibrometry (LDV). Dividing the result by ω yields the amplitude of the displacement as a function of vibrational frequency.

Unlike Chladni patterns, holographic interferometry or ESPI, LDV provides the amplitude of the vibration at only a single a point on the object of interest. If it is desirable to measure the motion over a large area the laser must be scanned across the object in a grid pattern, stopping at each point for a period of time determined by the desired precision of the measurement. The motion at each point is analyzed individually and then the motion of the entire body is reconstructed using a computer algorithm. Scanning is normally achieved by a pair of actuated mirrors within the system.

5.2.5 Accelerometers

Deflection measurements using accelerometers has become common in the field of musical acoustics. These devices are inexpensive and robust, and they can provide real-time information about the motion of an object in three dimensions. Although an accelerometer only provides a voltage that is proportional to the acceleration, the velocity and position can be determined by numerical integration of the signal.

Types of Accelerometers

There are several types of accelerometers, including piezoelectric, piezoresistive, capacitive and ones using strain gages. All accelerometers produce an output that is proportional to the acceleration of the device and can be obtained in configurations that are sensitive to changes in velocity in one, two or three dimensions. Regardless of the sensing mechanism, accelerometers consist of a mass attached to a sensor. An applied force on the accelerometer produces an acceleration of the mass, which results a change in the electrical characteristics of the sensor.

A piezoelectric crystal will produce a charge proportional to the stress, a capacitive sensor will produce a change in capacitance due to a varying distance between two plates, and a piezoresistive element and strain gages typically produce an output voltage through the change in resistance of a Wheatstone bridge arrangement. There are also accelerometers that use the Hall effect and magnetic induction. Although the resistive accelerometers provide excellent sub-sonic response, in most applications associated with musical acoustics it is the response at audible frequencies that is of interest. Therefore piezoelectric and micro-electric mechanical system (MEMS) capacitive accelerometers appear to be more common. We will briefly describe how these work before addressing their application.

A simplified diagram of a piezoelectric accelerometer is shown in Fig. 5.12. As with the piezoelectric microphone described in Sect. 5.1.1, the operation of a piezoelectric accelerometer depends on the fact that piezoelectric crystals produce a polarizing charge when they are under stress. This charge results in a potential difference across the crystal that is proportional to the stress. Inside the accelerometer housing the crystal is placed between a surface attached to the housing and a movable mass m. When it is subjected to a force F in the direction that the mass can move, a potential difference of V is produced. This potential is proportional to F/m, which according to Newton's second law is the acceleration.

As with electret microphones, capacitive accelerometers rely on the fact that the capacitance of a parallel plate capacitor is given by (5.1). Typically, a MEMS accelerometer will have several *fingers* on a movable structure that produce a capacitance as shown in Fig. 5.13. The movable beams between the static plates can be modeled as a mass m on a spring with spring constant k. For the simple case of a mass on a spring the restoring force of the spring is given by

$$F_s = -kx\,, \tag{5.46}$$

where x is the displacement of the mass from its equilibrium position. The distance between the plates on either side of the central conductor (i. e., the finger) is then

$$\Delta = d_0 \pm x\,, \tag{5.47}$$

where d_0 is the equilibrium distance between the plates when the accelerometer is not undergoing acceleration.

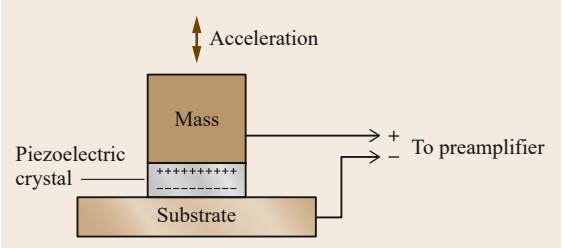


Fig. 5.12 Simplified diagram of a piezoelectric accelerometer. The piezoelectric crystal produces a potential difference that is proportional to the strain produced by the acceleration of the inertial mass

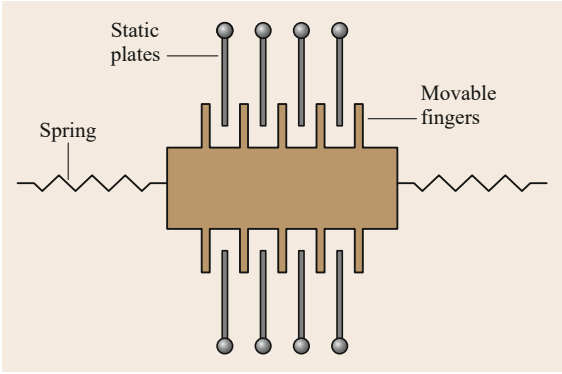


Fig. 5.13 Schematic diagram of a MEMS accelerometer. The acceleration is determined by measuring the change in capacitance that results from the fingers moving away from the equilibrium position

Detection is typically accomplished by measuring the difference in the capacitance between the plates closest to the movable finger and those further away [5.53]. When the device experiences an acceleration, the applied force results in the fingers moving closer to one set of plates than the other. Since the capacitors are all identical and wired in parallel, the capacitance is given by (5.1) multiplied by the number of capacitors N, assuming that one can ignore the edge effects and that the central fingers extend the length of the two plates. Substituting (5.47) into (5.1) and subtracting the two capacitances results in the difference between the capacitances when the accelerometer is subjected to a force. This results in an equation that indicates that the difference in capacitance is nonlinearly related to the displacement of the mass

$$\Delta C = 2N\varepsilon_0 A \frac{x}{d_0^2 - x^2} \,. \tag{5.48}$$

The nonlinearity is, however, negligible if $x \ll d_0$.

Rearranging (5.48) and ignoring the term quadratic in x reveals that the difference in capacitance is linearly related to the displacement of the fingers

$$\Delta C \approx x\left(\frac{2N\varepsilon_0 A}{d_0^2}\right) \tag{5.49}$$

and the acceleration is given by

$$a \approx \frac{-kd_0}{2N}\frac{\Delta C}{C_0} \,, \tag{5.50}$$

where C_0 is the capacitance when the sensor is experiencing no acceleration.

The sensitivity of an accelerometer is measured in units of volts per unit of acceleration. The normal unit of acceleration is equivalent to the acceleration due to the Earth's gravity, therefore the sensitivity is usually quoted in mV/g, where $g = 9.8\,\mathrm{m/s^2}$.

Measurements with Accelerometers

In the study of musical acoustics accelerometers can be used in a variety of ways, but they are typically used for determining the modal structure and the mechanical impedance (or admittance) of musical instruments [5.54–57]. Modal analysis will be briefly addressed here; measurements of the mechanical impedance are addressed in the following section.

Knowing the acceleration of a structure at the point where an accelerometer is attached is typically valuable only if something is known about the exciting force. Therefore, accelerometers are usually used in conjunction with a force sensor. There are two methods of determining the modal structure of an object using accelerometers: using multiple accelerometers placed at several points on the object, and using a single accelerometer while exciting the vibrations at different points serially. In each case it is the frequency response function (FRF) between the point of excitation and the accelerometer position that is measured. These measurements can be used to reconstruct the deflection shape of the object.

When multiple accelerometers are available and the object is small enough such that the distance between them is less than the distance between modal antinodes, it is convenient to drive the object either by impulse or steady-state excitation and record the motion at several different points. The accelerometers must be sensitive to the direction of displacement and are usually mounted so that the acceleration normal to the surface of the object produces a signal. The displacement is found by integrating the acceleration twice with respect to time.

To determine the mode shapes the position of each accelerometer must be known. To facilitate the analysis they are often placed symmetrically on the object. The multiple signals from the accelerometers can then be aggregated in a computer and used to determine the deflection and mode shapes. This technique was used in [5.56] to determine the mode shapes of a Himalayan singing bowl, but this work also demonstrated one of the limitations of this technique; the mass of the accelerometers altered the resonance frequencies of the instrument.

For modal analysis it is the displacement as a function of frequency that is of interest. If the driving mechanism is steady state and narrow band, the frequency may be swept as a function of time and the individual measurements aggregated at the frequencies of interest. If the excitation mechanism is broad band,

as when the body is excited by an impulse, the Fourier transform is calculated prior to analysis.

Multiple accelerometers can also be used to determine the whole-body motion and eliminate the effects of any modal vibrations if the body has simple symmetry, such as the bell of a brass instrument [5.50]. To do this, multiple accelerometers are attached to the structure in a symmetric pattern and vibrations are driven over a broad band, either acoustically or by contact with a shaker. The phase and amplitude of the accelerometer signals are recorded and then averaged in the complex plane. In this way the vibrations due to modal resonances are eliminated due to the fact that the antinodes are symmetric and adjacent ones are π out of phase. The resulting signal contains only the effects of whole-body and asymmetric motions.

For large instruments, such as pianos, it is more common to attach a single accelerometer to some point and excite the vibrations at points on the instrument in a grid pattern. The signal from the accelerometer and the exciting instrument, often a force hammer, can then be used to determine the mode shapes of the structure. An example of the result of this type of measurement can be found in [5.54].

The theory is the same in either case, therefore here we will only consider the case where a single accelerometer is affixed at a point (x_0, y_0) and N points at (x_n, y_n) are excited by a broad-band strike from a force hammer. It is assumed that after the strike the motion can be described by a set of damped orthogonal harmonic modes and that nonlinear coupling between the modes is negligible.

The FRF between the position of the accelerometer (x_0, y_0) and the exciting strike at (x_n, y_n) is determined by the mode amplitude at each point and the frequency-dependent modal damping γ_m, which determines the quality factor of the resonance. Denoting the mode shapes as $\Phi_\mathrm{m}(x, y)$, the FRF is given by

$$H_{0,n}(\omega) = \sum_{m=1}^{\infty} \frac{\Phi_m(x_0, y_0)\Phi_m(x_n, y_n)}{(\omega_m^2 - \omega^2) + 2i\gamma_m\omega\omega_m} , \qquad (5.51)$$

where ω_m is the resonant frequency of the m-th mode. Typically, only a limited number of modes are of interest and m is truncated such that the angular frequency does not exceed some predetermined maximum.

The mode shapes can be determined from the measurement of $H(\omega)$ by averaging measurements from several different points. Assuming that only M modes are of interest and measurements are made at N unique positions, then

$$H_{0,N}(\omega) = \sum_{m=1}^{M} \frac{\sum_{n=1}^{N} \Phi_m(x_0, y_0)\Phi_m(x_n, y_n)}{(\omega_m^2 - \omega^2) + 2i\gamma_m\omega\omega_m} . \qquad (5.52)$$

An optimization algorithm is then used to determine the mode shapes Φ_m by comparing the computed value of $H_{0,N}$ with the value derived from measurements.

When performing modal analysis in this manner it is important to ensure that the frequency spectrum of the exciting impact contains the frequencies of interest and that the accelerometer is not placed on a node of the mode of interest. It is also important to ensure that the excitation mechanism is similar at each point and, in the case of impact excitation, that there are not multiple strikes. There are several methods available to ensure that the structure is excited in an appropriate manner [5.58].

5.3 Measurement of Impedance

All musical instruments consist of one or more oscillators and one or more resonators. The final sound is largely determined by the spectrum of the oscillator and the input impedance of the resonators. Therefore, much effort goes into measuring the input impedance of musical instruments.

The input impedance of wind instruments is defined as the input pressure divided by the volume flow. In other instruments it is the mechanical impedance at the driving point that is important, which is defined as the applied force divided by the velocity. In either case, the impedance determines the efficiency of the system and it is a function of frequency. Some of the methods that can be used to measure both types of impedance are discussed here. In each case measurements are made either by using a swept sine function or a random noise generator to produce a signal. If a random noise generator is used, the Fourier transform of the signal is first calculated before determining the impedance to provide frequency discrimination.

5.3.1 Mechanical Impedance

The impedance that is of interest for many instruments, such as percussion and stringed instruments, is the impedance at the point of energy transfer from the string or striking mechanism to the resonating body. Some instruments have multiple interactions that are of

interest and therefore there is not one impedance measurement that characterizes the response. The sound generated by a piano, for example, depends on the impedance of the string during the interaction with the hammer as well as the impedance of the bridge where it meets the string. Both impedances are mechanical impedances.

The mechanical impedance at the point of excitation, referred to as the driving point impedance, is normally measured by applying a force at the same point that the acceleration is measured. To determine the mechanical impedance, the signal from an accelerometer is integrated to provide the velocity v in the direction of the exciting force. The impedance is defined by

$$Z = \frac{F}{v}\,, \tag{5.53}$$

where F is the applied force. Since the force and velocity are not generally in phase, Z is normally a complex value. Often the admittance, which is the reciprocal of the impedance, is used for analysis rather than the impedance.

The force sensor is usually made from a piezoelectric crystal that is placed between the driving mechanism and the object of interest. Compression of the crystal produces a potential difference proportional to the stress, as described in Sect. 5.2.5. The difference between an accelerometer and a force sensor is that an accelerometer has the crystal attached to an inertial mass that is free to move while in the force sensor the crystal is placed directly in line between the driving element and the object of interest.

An impedance head consists of a driving mechanism called a shaker, which is usually driven electromagnetically by a coil and magnet arrangement, similar to the arrangement found in a common speaker. The shaker may be driven by a swept harmonic signal or by a noise signal containing all of the frequencies of interest. An accelerometer is placed between the driving mechanism and the force sensor, and the force sensor is placed as close to the object as is possible. This placement of the force sensor is important because the measured impedance is affected by any mass m that exists between the force sensor and the object. When there is a mass between the force sensor and the object, the measured impedance Z_m is related to the actual impedance Z by

$$Z_m = Z + i\omega m\,, \tag{5.54}$$

and therefore it is important to minimize m to obtain an accurate measurement.

The mechanical impedance can also be measured by impacting the surface of the object with a hammer containing an integrated sensor rather than driving it with a shaker. As is the case when a shaker is driven by a noise signal, when an impact is used to excite the motion it is necessary to Fourier transform the force and acceleration signals. If the driver is a shaker driven by a harmonic signal that is swept through the frequencies of interest over time, the appropriate frequency dependence can be directly substituted for the time dependence. In each case the velocity can be determined at each frequency by dividing the acceleration by $i\omega$. A direct application of (5.53) will then yield the impedance spectrum.

The velocity at the point of excitation can also be measured by other means. LDV can be used and is very convenient, but by necessity the measurement of the velocity cannot be exactly at the point of excitation since the driver must be located there. However, if the surface of interest is thin and can be accessed from both sides, such as is possible with a piano soundboard, the measurement of velocity can be made on the opposite side of the structure. Any measurement of the displacement can also be used to determine the velocity.

5.3.2 Impedance of Wind Instruments

Measuring the mechanical impedance is relatively straightforward, however, measuring the input impedance of a wind instrument is more complicated. The input impedance of a wind instrument is a complex value defined as the ratio of the pressure p and the volume flow u at the point of input

$$Z = \frac{p}{u}\,. \tag{5.55}$$

In wind instruments the input is usually defined as the point where the oscillator meets the resonator, i. e., the mouthpiece.

While it is relatively easy to measure both the force and velocity of a point on a mechanical system, the simultaneous measurement of the pressure and volume flow at the input of a wind instrument can be quite difficult. Rather than measuring the flow, often a known flow is imposed on the system by a membrane with very high impedance. The high impedance ensures that the input impedance of the instrument under test negligibly affects the flow. In practice a piezoelectric disk works well for this purpose [5.59]. However, it is also possible to couple a low impedance driver, such as a speaker, through a small capillary tube. This arrangement can provide a close approximation to a high impedance driver and it was the first known method for making direct measurements of the input impedance of brass

instruments [5.60]. Later developments provided corrections [5.61] and this method is currently used in the commercially available brass instrument analysis system (BIAS) [5.62]. An example of the impedance of a B-flat trumpet as a function of frequency measured using the BIAS system is shown in Fig. 5.14.

Provided the driving mechanism is unaffected by the presence of the attached instrument, the input impedance can be determined simply by measuring the pressure at the input, which can be accomplished using a small microphone. The problem with this arrangement is that the microphone and the driving diaphragm must be at the same place. Any displacement of the microphone from the driving point results in an error, which can sometimes be significant, but these errors can usually be accounted for theoretically [5.59].

Methods exist for measuring the flow directly, for example a hot-wire anemometer can be use [5.63], but normally the volume flow is determined by inducing a known flow or by deducing it from a pressure gradient determined using two or more microphones [5.64, 65]. This method has been extensively improved upon since the first report of its use [5.66–68] and is in common use for measuring the input impedance of wind instruments.

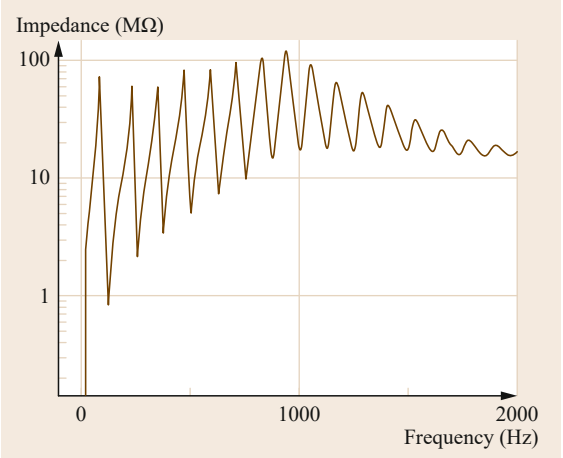


Fig. 5.14 Input impedance of a B-flat trumpet with all of the valves open measured using the BIAS impedance head. Playable notes occur where the impedance is a maximum

5.4 Conclusions

While the measurement techniques discussed here encompass many of the measurement techniques important to musical acousticians, it is by no means an exhaustive treatment of the subject. This chapter is intended only as an overview of some of the more common experimental techniques and should form the basis for further investigation.

It is likely that the technology available for measuring the physical parameters of musical instruments will continue to rapidly advance. Therefore, as new systems and techniques become commercially available in the future many of the techniques discussed here will be easier to implement than they currently are. Musical acousticians would be wise to keep pace with these developments.

References

5.1 D.R. Raichel: *The Science and Applications of Acoustics* (Springer, New York 2000) pp. 168–175

5.2 L.E. Kinsler, A.R. Frey, A.B. Coppens, J.V. Sanders: *Fundamentals of Acoustics*, 4th edn. (Wiley, New York 2000) pp. 416–428

5.3 M.W. Hoffman, C. Pinkelman, X.F. Lu, Z. Li: Real-time and off-line comparisons of standard array configurations containing three and four microphones, J. Acoust. Soc. Am. **107**, 3560–3563 (2000)

5.4 R. Streicher, W. Dooley: Basic stereo microphone perspectives-a review, J. Audio Eng. Soc. **33**, 548–556 (1985)

5.5 M. Park, B. Rafaely: Sound-field analysis by plane-wave decomposition using spherical microphone array, J. Acoust. Soc. Am. **118**, 3094–3103 (2005)

5.6 N. Huleihel, B. Rafaely: Spherical array processing for acoustic analysis using room impulse responses and time-domain smoothing, J. Acoust. Soc. Am. **133**, 3395–4007 (2013)

5.7 E.G. Williams, J.D. Maynard: Holographic imaging without the wavelength resolution limit, Phys. Rev. Lett. **45**, 554–557 (1980)

5.8 E.G. Williams: *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, London 1999)

5.9 J.D. Maynard, E.G. Williams, Y. Lee: Nearfield acoustic holography: I. Theory of generalized hologrphy and the development of NAH, J. Acoust. Soc. Am. **78**, 1395–1413 (1985)

5.10 S. Dumbacher, D. Brown, J. Blough, R. Bono: Practical aspects of making NAH measurements. In: *Proc.*

Noise and Vibration Conference and Exposition, Warrendale (1999)
5.11 F. Muddeen, B. Copeland: Sound radiation from caribbean steelpans using nearfiled acoustical holography, J. Acoust. Soc. Am. **131**, 1558–1595 (2012)
5.12 L.M. Wang, C.B. Burroughs: Acoustic radiation from bowed violins, J. Acoust. Soc. Am. **110**, 543–555 (2001)
5.13 J. Benesty, J. Chen, Y. Huang (Eds.): *Microphone Array Signal Processing* (Springer, Berlin, Heidelberg 2008)
5.14 M. Brandstein, D. Ward (Eds.): *Microphone Arrays: Signal Processing Techniques and Applications* (Springer, New York 2001)
5.15 M.B.S. Magalhães, R.A. Tenenbaum: Sound sources reconstruction techniques: A review of their evolution and new trends, Acta Acust. united with Acust. **90**, 199–220 (2004)
5.16 G.H. Koopmann, L. Song, J.B. Fahnline: A method for computing acoustic fields based on the principle of wave superposition, J. Acoust. Soc. Am. **86**, 2433–2438 (1989)
5.17 R. Bader: *Microphone Arrays* (Springer, Berlin, Heidelberg 2014)
5.18 R. Bader: Radiation characteristics of multiple and single sound hole vihuelas and a classical guitar, J. Acoust. Soc. Am. **131**, 819–827 (2012)
5.19 R. Bader: Reconstruction of radiating sound fields using minimum energy method, J. Acoust. Soc. Am. **127**, 300–308 (2010)
5.20 T. Rossing: Chladni's law for vibrating plates, Am. J. Phys. **50**, 271–274 (1982)
5.21 D. Waller: *Chladni Figures: A Study in Symmetry* (Bell, London 1961)
5.22 T.R. Moore, A.E. Cannaday, S.A. Zietlow: A simple and inexpensive optical technique to help students visualize mode shapes, J. Acoust. Soc. Am. **131**, 2480–2487 (2012)
5.23 J.R. Comer, M.J. Shepard, P.N. Henriksen, R.D. Ramsier: Chladni plates revisited, Am. J. Phys. **72**, 1345–1346 (2004)
5.24 H.A. Conklin: Design and tone in the mechanoiacoustic piano. Part II. Piano structure, J. Acoust. Soc. Am. **100**, 695–708 (1996)
5.25 N.E. Molin, L.E. Lindgren, E.V. Jansson: Parameters of violin plates and their influence on the plate modes, J. Acoust. Soc. Am. **83**, 281–291 (1988)
5.26 P.G.M. Richardson, E.R. Toulson, D.J.E. Nunn: Analysis and manipulation of modal ratios of cylindrical drums, J. Acoust. Soc. Am. **131**, 907–913 (2012)
5.27 T.D. Rossing, A. Perrier: Modal analysis of a Korean bell, J. Acoust. Soc. Am. **94**, 2431–2433 (1993)
5.28 T. Rossing, I. Bork, H. Zhao, D.O. Fystrom: Acoustics of snare drums, J. Acoust. Soc. Am. **92**, 84–94 (1992)
5.29 T.J. Hill, B.E. Richardson, S.J. Richardson: Acoustical parameters for the characterization of the classical guitar, Acta Acust. united with Acust. **90**, 335–348 (2004)
5.30 M.L. Facchinetti, X. Boutillon, A. Constantinescu: Numerical and experimental modal analysis of the reed and pipe of a clarinet, J. Acoust. Soc. Am. **113**, 2874–2883 (2003)
5.31 G. Jundt, A. Radu, E. Fort, J. Duda, H. Vach, N. Fletcher: Vibrational modes of partly filled wine glasses, J. Acoust. Soc. Am. **119**, 3793–3798 (2006)
5.32 R. Jones, C. Wykes: *Holographic and Speckle Pattern Interferometry* (Cambridge Univ. Press, Cambridge 1989)
5.33 B. Richardson: The acoustical development of the guitar, J. Catgut Acoust. Soc. **2**, 1–10 (1994)
5.34 G.M. Brown, R.M. Grant, G.W. Stroke: Theory of holographic interferometry, J. Acoust. Soc. Am. **45**, 1166–1179 (1969)
5.35 B. Copeland, A. Morrison, T. Rossing: Sound radiation from caribbean steelpans, J. Acoust. Soc. Am. **117**, 375–383 (2005)
5.36 L.A. Stephey, T.R. Moore: Experimental investigation of an american five-string banjo, J. Acoust. Soc. Am. **124**, 3276–3283 (2008)
5.37 T.R. Moore, J.D. Kaplon, G.D. McDowall, K.A. Martin: Vibrational modes of trumpet bells, J. Sound Vib. **254**, 777–786 (2002)
5.38 R. Worland: Normal modes of a musical drumhead under non-uniform tension, J. Acoust. Soc. Am. **127**, 525–533 (2010)
5.39 A.E. Cannaday, B.C. August, T.R. Moore: Tuning the nigerian slit gong, J. Acoust. Soc. Am. **131**, 1566–1573 (2012)
5.40 B.M. Deutsch, C.L. Ramirez, T.R. Moore: The dynamics and tuning of orchestral crotales, J. Acoust. Soc. Am. **116**, 2427–2433 (2004)
5.41 R. Worland: Musical acoustics of orchestral water crotales, J. Acoust. Soc. Am. **131**, 935–944 (2012)
5.42 T.R. Moore, S.A. Zietlow: Interferometric studies of a piano soundboard, J. Acoust. Soc. Am. **119**, 1783–1793 (2006)
5.43 A.E. Ennos: *Speckle Interferometry* (Springer, New York 1984) pp. 203–253, ed. by C. Dainty
5.44 T.R. Moore, J.J. Skubal: Time-averaged electronic speckle pattern interferometry in the presence of ambient motion. Part 1. Theory and experiments, Appl. Opt. **47**, 4640–4648 (2008)
5.45 T.R. Moore: A simple design for an electronic speckle pattern interferometer, Am. J. Phys. **72**, 1380–1384 (2004)
5.46 T.R. Moore: A simple design for an electronic speckle pattern interferometer, Am. J. Phys. **73**, 189 (2005)
5.47 Y. Yeh, H.Z. Cummins: Localized fluid flow measurements with an he-ne laser spectrometer, Appl. Phys. Lett. **4**, 176–178 (1964)
5.48 T. Ryan, P. O'Malley, J. Vignola, J. Judge: Conformal scanning laser doppler vibrometer measurement of tenor steelpan response to impluse excitation, J. Acoust. Soc. Am. **132**, 3494–3501 (2012)
5.49 E. Skrodzka, A. Lapa, B.B. Linde, E. Rosenfeld: Modal parameters of two incomplete and complete guitars differing in the bracing pattern of the soundboard, J. Acoust. Soc. Am. **130**, 2186–2194 (2011)
5.50 V. Chatziioannou, W. Kausel, T. Moore: The effect of wall vibrations on the air column inside trumpet bells. In: *Proc. Acoustics Nantes Conf. EAA, Nantes* (2012) pp. 2243–2248
5.51 E. De Lauro, S. De Martino, E. Esposito, M. Falanga, E.P. Tomasini: Analogical model for mechanical vibrations in flue organ pipes inferred by independent component analysis, J. Acoust. Soc. Am. **122**, 2413–

2424 (2007)
5.52 E. Hecht: *Optics*, 4th edn. (Addison Wesley, San Francisco 2002) pp. 560–578
5.53 L.E. Lyshevski: *MEMS and NEMS: Systems, Devices and Structures* (CRC, Boca Raton 2001)
5.54 H. Suzuki: Vibration and sound radiation of a piano soundboard, J. Acoust. Soc. Am. **80**, 1573–1582 (1986)
5.55 J. Berthaut, M.N. Ichchou, L. Jézéquel: Piano soundboard: structural behavior, numerical and experimental study in the modal range, Appl. Acoust. **64**, 1113–1136 (2003)
5.56 O. Inácio, L.L. Henrique, J. Antunes: The dynamics of tibetan singing bowls, Acta Acust. united with Acust. **92**, 637–653 (2006)
5.57 C. Waltham, A. Kotlicki: Vibrational characteristics of harp soundboards, J. Acoust. Soc. Am. **124**, 1774–1780 (2008)
5.58 D.J. Ewins: *Modal Testing: Theory, Practice and Application* (Research Studies, Baldock 2000) pp. 25–286
5.59 A.H. Benade, M.I. Ibisi: Survey of impedance methods and a new piezo-disk-driven impedance head for air columns, J. Acoust. Soc. Am. **81**, 1152–1167 (1987)
5.60 J.C. Webster: An electrical method of measuring the intonation of cup-mouthpiece instruments, J. Acoust. Soc. Am. **19**, 902–906 (1947)
5.61 J. Agulló, J. Badrinas: Improving the accuracy of the cappillary based technique for measuring the acoustic impedance of wind instruments, Acustica **59**, 76–83 (1985)
5.62 W. Kausel: Bore reconstruction of tubular ducts from its acoustic input impedance curve. In: *Proc. IEEE Instrument Measurement Technol. Conf., New York* (2003) pp. 993–998
5.63 S. Elliott, J. Bowsher, P. Watkinson: Input and transfer response of brass wind instruments, J. Acoust. Soc. Am. **72**, 1747–1760 (1982)
5.64 J.Y. Chung, D.A. Blaser: Transfer function method of measurring in-duct acoustic properties I. Theory, J. Acoust. Soc. Am. **68**, 907–913 (1980)
5.65 J.Y. Chung, D.A. Blaser: Transfer function method of measurring in-duct acoustic properties II. Experiment, J. Acoust. Soc. Am. **68**, 914–921 (1980)
5.66 V. Gibiat, F. Laloë: Acoustical impedance measurements by the two-microphone-three-calibration (TMTC) method, J. Acoust. Soc. Am. **88**, 2533–2545 (1990)
5.67 P.-P. Dalmont: Acoustic impedance measurement, Part I: A review, J. Sound. Vib. **243**, 427–439 (2001)
5.68 M. van Walstijn, D.M. Campbell, J. Kemp, D. Sharp: Wideband measurement of the acoustic impedance of tubular objects, Acta Acust. united with Acust. **91**, 590–604 (2005)

6. Some Observations on the Physics of Stringed Instruments

Nicholas Giordano

We provide a general introduction to stringed instruments, focusing on the piano, guitar, and violin. These are representative of instruments in which the strings are excited by striking (the piano), plucking (the guitar), and bowing (the violin). We begin by discussing, in a general way, the strings and soundboards, and how these couple to the surrounding air to generate sound. Important features specific to these instruments are then discussed, with particular attention to the different ways the strings are set into motion, key differences in the way the soundboards vibrate, and the effects of these differences on the resulting musical tones.

6.1 Three Classes of Stringed Instruments

A very wide variety of musical instruments employ strings as the central vibrating element; that is, as the vibrating element that determines the component frequencies of the musical tone produced by the instrument. It is indeed a challenge to describe all of the important aspects of all stringed instruments and we will not attempt such an ambitious task in this chapter. (The references given throughout this chapter should provide the interested reader with an entryway to more extensive information, as do a number of excellent books [6.1–7]). Our goal instead will be to give an overview of the important features common to all stringed instruments and then highlight the key differences between the three main classes of stringed instruments. These three classes are:

1. Instruments in which the strings are struck, such as the piano
2. Instruments in which the strings are plucked, such as the guitar
3. Instruments in which the strings are bowed, such as the violin.

Our goal will be to understand how and why the tones produced by these instruments differ.

6.2 Common Components and Issues

A generic stringed instrument is composed of strings connected to a soundboard or top plate (we will use these two terms interchangeably). A string is set into motion by either striking with a hammer, plucking, or bowing. As they vibrate, the strings produce a time-varying force that sets the soundboard into motion. Because of their very small surface area, the strings themselves produce a negligible amount of sound. It is instead the motion of the soundboard that dominates sound production. When considering the soundboard motion, we may also need to take into account the way the soundboard couples to the surrounding air. For a piano, it is a good approximation to treat the soundboard as simply a vibrating surface driven by only the strings.

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_6

In contrast, guitars and violins have an enclosure coupled to the outside through a hole, forming a Helmholtz resonator and we must consider the coupled modes of this resonator (and the air inside it) and the soundboard itself.

6.2.1 Strings

As a rough approximation, musical instrument strings can be treated as ideal flexible strings described by the familiar wave equation

$$\frac{\partial^2 y}{\partial t^2} = v_s^2 \frac{\partial^2 y}{\partial x^2} \,, \tag{6.1}$$

where x is position measured along the string, $v_s = \sqrt{T/\rho}$ is the speed of a transverse wave on the string with mass per unit length ρ and tension T, and y is the transverse displacement of the string. (The origin and properties of the solutions of this equation are reviewed in this volume by *Kausel* in Chap. 2). There are, of course, two transverse directions and we will consider both of them in due course. For now we will only be concerned with one of the transverse directions. In a typical musical instrument one end of the string is held rigidly while the other is attached to a bridge that attaches to the soundboard. The bridge moves with the soundboard, so that end of the string is not stationary, but this motion is sufficiently small that we can, to a first approximation, consider the string to be fixed at both ends. (It is not hard to include the effect of soundboard motion on the string and this is done in some of the piano models referenced in this chapter.) The vibrational modes of an ideal string held in this way are the usual standing waves, with frequencies that form a harmonic series $f_n = nf_1$ with n an integer and the fundamental frequency $f_1 = v_s/(2L)$ for a string of length L.

This picture of an ideal flexible string described by (6.1) with its perfectly harmonic modes is very familiar, but it does not give an adequate description of musical instrument strings. Two key features of real strings are missing from (6.1): damping and stiffness. Both of these effects are included in the more realistic wave equation [6.2, 7–10]

$$\frac{\partial^2 y}{\partial t^2} = v_s^2 \left[\frac{\partial^2 y}{\partial x^2} - \epsilon \frac{\partial^4 y}{\partial x^4} \right] - \alpha_1 \frac{\partial y}{\partial t} + \alpha_2 \frac{\partial^3 y}{\partial t^3} \,. \tag{6.2}$$

String stiffness is proportional to the Young's modulus E_s of the string material, with $\epsilon = r_s^2 \sqrt{E_s/\rho}/v_s^2$ for a string of radius r_s; this result can be derived from Newton's second law [6.2]. Damping is not as straightforward to characterize and can be described in several different ways. In (6.2) damping is accounted for with two terms. One of these terms gives the usual damping proportional to the string velocity while the other involves a higher time derivative of y. The second term is necessary for describing the frequency-dependent damping found in real strings. Such frequency-dependent damping can also be described mathematically in other ways [6.11]. As a final note concerning (6.2), this relation does not explicitly contain the force of the hammer on the string (that term would appear on the right in (6.2)) nor does it account explicitly for the motion of the end of the string attached to the bridge.

Roughly speaking, damping has a small yet discernable effect on the tonal properties, but the tonal decay is usually dominated by other sources of energy loss (often to the soundboard) rather than to the internal damping of the string.

String stiffness can be very important in instruments such as pianos, where the strings are relatively thick and are made of steel. For pianos, string stiffness affects the standing wave frequencies, shifting them above the values found for an ideal string, with

$$f_n \approx nf_1(1 + \beta n^2) \,, \tag{6.3}$$

where β is proportional to ϵ in (6.2) [6.9]. The deviations from a harmonic spectrum in (6.3) have an important affect on piano tones and on the way a piano is tuned, as we will explain below. Stiffness is much less important with the relatively thin strings found in acoustic guitars and violins, especially when these strings are composed of nylon or a similar material. Since the vibrational modes of a piano string are not harmonically related, they are usually referred to as *partials* instead of *harmonics*.

6.2.2 Soundboards

The soundboards in pianos, guitars, and violins are generally made from spruce. This choice is generally justified by noting that its ratio of Young's modulus (for the *strongest* direction, along the grain) to density is extremely large. As with most woods, the elastic constants of spruce depend strongly on direction and in the general case a total of 27 different constants are required to completely describe its elastic behavior. The pieces of spruce used in musical instruments are *quarter cut* sections [6.6], which allows a description with a much smaller number of Young's moduli, Poisson ratios, and shear moduli. Because soundboards are thin compared to the vibration amplitudes of interest, they can be modeled as *thin*

plates [6.12] (Chap. 2), although a more general treatment is sometimes used [6.11]. The equation of motion for a thin plate has the form [6.13]

$$\rho_b h_b \frac{\partial^2 u_b}{\partial t^2} = -D_x \frac{\partial^4 u_b}{\partial x^4} - (D_x \nu_y + D_y \nu_x + 4D_{xy}) \frac{\partial^4 u_b}{\partial x^2 \partial y^2} - D_y \frac{\partial^4 u_b}{\partial y^4} - \beta \frac{\partial u_b}{\partial t} , \quad (6.4)$$

where ρ_b is the density of the soundboard and u_b is the displacement of the soundboard in the perpendicular (z) direction. The rigidity factors D_x, D_y, etc. in (6.4) depend on the Young's moduli (E), Poisson ratios (ν), shear moduli (G), and plate thickness (h_b), with

$$D_x = \frac{h_b^3 E_x}{12(1-\nu_x \nu_y)} ,$$
$$D_y = \frac{h_b^3 E_y}{12(1-\nu_x \nu_y)} ,$$
$$D_{xy} = \frac{h_b^3 G_{xy}}{12} . \quad (6.5)$$

Equation (6.4) also contains a damping term (proportional to β). The form of the damping term is largely heuristic and is intended to describe the internal damping in the soundboard (other forms for this term can and have been used).

The equation of motion (6.4) is a simplified description of a full instrument, as it ignores the forces exerted by the strings on the soundboard. These would contribute an extra term on the right in (6.4), as would the force resulting from the pressure in the surrounding air. (The force of the air on a piano soundboard is only a small effect and to the best of our knowledge has not been included in any piano modeling. This force is much more important for the thinner soundboards found in violins and guitars.) Equation (6.4) also omits the effect of the soundboard crown. This is an important effect which is complicated to model and is included in only the more recent modeling work [6.14–16].

The modeling of soundboard vibrations is made challenging by several facts. First, the term proportional to the fourth derivative in (6.4) makes soundboards highly dispersive which complicates numerical solutions. Second, the rigidity factors in (6.4) are functions of position, due to the presence of soundboard ribs (narrow strips of wood glued to the underside of the soundboard and which generally run perpendicular to the grain) and one or more bridges, along with the fact that the thickness of the soundboard usually varies with position, being thinner at the edges of the board. A full discussion of these complications is beyond the scope of this chapter; we refer the reader to the literature for a full discussion [6.12, 13, 17]. For our purposes in this chapter (6.4) is useful for describing the vibrational properties of a soundboard in general terms. Such a description involves two quantities: the mechanical admittance as a function of frequency and the modal deflection shapes. The mechanical admittance is the ratio of the soundboard velocity to the magnitude of an applied force, with the force (and resulting velocity) being harmonic. Modal deflection shapes are the deflection pattern of the soundboard at a particular frequency, typically the frequency of one of the normal modes. We will give examples of these when we discuss specific instruments. A third complication has to do with the soundboard crown. This crown requires a nonlinear description (note that while (6.4) is highly dispersive, it is still linear). These nonlinear effects are now being understood [6.14–16], but more work needs to be done in this area. In particular, to the best of our knowledge, the effect of the soundboard crown on soundboard damping remains to be understood.

6.2.3 Sound Generation

Stringed instruments generate sound by virtue of the air displaced due to soundboard vibrations. In normal situations, these vibrations are small enough that the behavior is described by the relations of linear acoustics. For our purposes it is convenient to express these relations as [6.18]

$$\rho_a \frac{\partial v_x}{\partial t} = -\frac{\partial p}{\partial x} , \quad (6.6)$$
$$\rho_a \frac{\partial v_y}{\partial t} = -\frac{\partial p}{\partial y} ,$$
$$\rho_a \frac{\partial v_z}{\partial t} = -\frac{\partial p}{\partial z} ,$$
$$\frac{\partial p}{\partial t} = \rho_a c_a^2 \left[-\frac{\partial v_x}{\partial x} - \frac{\partial v_y}{\partial y} - \frac{\partial v_z}{\partial z} \right] , \quad (6.7)$$

where p is the pressure, the velocity components (of the air) are v_x, v_y and v_z, ρ_a is the density, and c_a is the speed of sound in air.

The solution of these equations is numerically straightforward, as discussed in numerous places [6.19–21]. A crucial part of the problem that does not appear explicitly in (6.6) is the source of the sound excitation. This excitation enters through the boundary conditions, which are connected with the motion of the body of the instrument. For the piano, the motion of the soundboard imposes a boundary condition on the air ve-

locity on the surface of the soundboard. For the guitar and violin, the motion of the air produces a significant force on the soundboard which must be included in the equations of motion of the soundboard, (6.4). This will be discussed further in the following sections.

6.3 The Story of Three Instruments

We will now describe how the general ideas from the previous section – in which a stringed instrument is viewed (or modeled) as a combination of string(s), soundboard, and air – apply to three specific instruments. Limitations of space will prevent us from going into great depth in many important areas; we encourage the interested reader to use the references to explore the instruments in much greater detail. For excellent general references we recommend for pianos the papers by *Conklin* [6.22–24], along with [6.6, 11], for guitars we recommend [6.3, 14, 25], and for the violin we suggest [6.1]. For general introductions to all three instruments we recommend [6.2, 7].

6.3.1 Piano

Piano Hammers and Strings

Modern pianos have 88 notes. Notes in the midrange and treble employ three steel strings, while bass notes involve (typically) two or one string(s) which are composed of a copper string wound around a steel core [6.24]. This design increases the string mass without adding excessively to the string stiffness, thus keeping the inharmonicity in (6.3) to an acceptably low level.

Piano strings are set into motion by the blow from a hammer. A modern piano hammer is a felt-covered mallet with a mass of typically 3−10 g, with the heavier hammers used for the bass strings. Felt is one of the most complex (and difficult to model) materials used in any stringed instrument and our understanding of the force–compression properties of piano hammers is purely empirical. The interaction of a piano hammer with a string is described by a force–compression function $f_h(z_h)$ where f_h is the hammer string force and z_h is the amount the felt is compressed during the hammer–string collision. There have been a great many studies of piano hammers (see [6.9, 10, 26–35] for a partial list!). The experiments have shown that $f_h(z_h)$ is an extremely nonlinear function and can have significant hysteresis. The nonlinearity is illustrated in Fig. 6.1, which shows the force–compression function for a typical hammer from the middle range of a piano measured with the hammer striking a rigidly held string. The force–compression function is described approximately by a power law $f_h \approx z_h^{\alpha}$, with an exponent value α typically between 2.5 and 5. These exponent values are distinctly larger than the value $\alpha = 1$ expected for an object that follows Hooke's law and the value $\alpha = 3/2$ based on the Hertz solution for the contact force involving elastic bodies. This nonlinearity means that the hammer becomes effectively stiffer when the compression is increased and has an important effect on the tone color for the following reason. When a note is played softly, the hammer has a relatively low velocity just prior to its collision with the string; the felt compression during the collision is then small and the hammer is effectively soft. A collision involving a soft object results in a spectrum dominated by the lowest frequencies with little weight at high frequencies (i. e., at the higher partials of the string). Conversely, when a note is played loudly the hammer velocity is high, resulting in a much larger compression and an effectively stiffer hammer. The collision with this stiffer hammer produces a sound with much more weight in the high partials as compared to the behavior with a soft hammer. In this way, the hammer nonlinearity in Fig. 6.1 gives a change in tone color along with the change in loudness of a piano tone. This is a key to the expressiveness of the instrument.

Another important contribution to the tone color involves the string stiffness. As mentioned above in connection with (6.3), string stiffness causes a string's

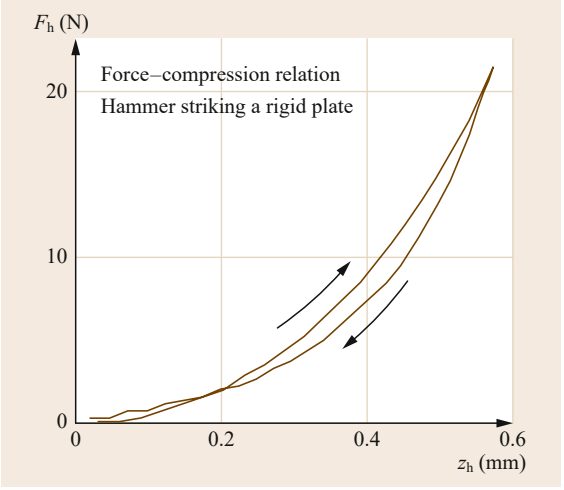


Fig. 6.1 Force–compression characteristic for a piano hammer from the middle range of a piano. Here z is the amount that the hammer felt is compressed as it collides with a rigidly held string (after [6.35])

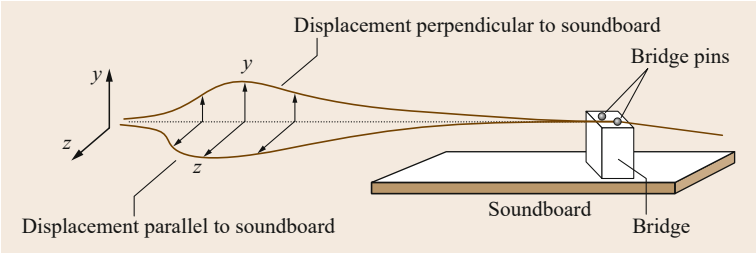


Fig. 6.2 A piano string can undergo transverse vibrations in two directions: perpendicular to the soundboard (along y) and parallel to the plane of the soundboard (along z). These two motions are coupled through the motion of the soundboard and bridge (after [6.6])

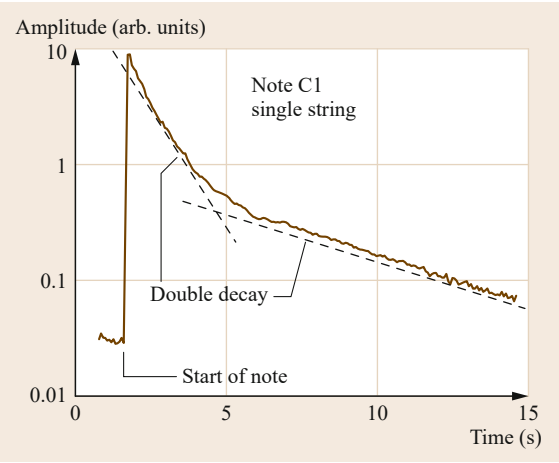


Fig. 6.3 Decay of the sound for the note C1 on a grand piano. This is the lowest C on the piano and involves the motion of a single string (after [6.6])

vibrational frequencies to deviate from a purely harmonic spectrum. One might have thought that the *best* piano tones would be ones in which the deviations from harmonicity are smallest. However, listening tests show that human listeners seem to prefer a small inharmonicity [6.36]. (See also [6.37] for a study of inharmonicity in historical instruments.) The reason for this is not clear and involves the perception of musical tones, a topic that we will not attempt to explore here. The existence of inharmonicity has important implications for the tuning of a piano. Consider two notes that are nominally an octave apart; that is, two notes that would be an octave apart on an ideal musical scale, which we can denote as notes C1 and C2. In the ideal case, the frequency of the second harmonic of C1 would be exactly twice the frequency of the fundamental of C1 and would equal the frequency of the fundamental of C2. For a real string with inharmonicity, we have according to (6.3) that the frequency of the second partial of C1 is slightly greater than twice the fundamental frequency of C1. In order for C1 and C2 to *sound good* when played together, the second partial of C1 should coincide with the fundamental of C2 and this means that the separation between these two notes must be slightly greater than the ideal factor-of-two separation. This is called a stretched octave and the presence of string stiffness means that all octaves on a well-tuned piano must be stretched.

This is a problem that was explored systematically some years ago and led to a surprising result. In a survey of the tunings produced by a collection of skilled piano tuners, it was found that all arrived at approximately the same stretched tuning; i. e., the same deviations from an ideal tuning [6.38–40]. This tuning is now called the Railsback stretch and it seems to be the tuning that *sounds best* to a typical listener [6.40]. The precise form for the preferred stretch tuning appears to be connected with (and explained by) our perceptual judgment of consonance and dissonance [6.41].

In our discussion of string motion we characterized that motion in terms of a single variable $y(x,t)$ in (6.2) where y is the displacement of the string in the direction perpendicular to the soundboard and parallel to the force exerted by the hammer (Fig. 6.2). String motion in the direction parallel to the soundboard and perpendicular to the axis of the string (the z-direction in Fig. 6.2) is also possible. Since the hammer force is along y, one might think that motion parallel to the soundboard will be negligible. That is not the case, however, because motion of the end of the string connected to the soundboard (through the bridge) couples the motions in these two perpendicular directions due to rocking motion of the bridge. This leads to an interesting time dependence for the decay of the string vibration and of the sound, as first discussed by *Weinreich* [6.42]. The *double decay* that Weinreich described is illustrated in Fig. 6.3, which shows the decay of the sound from the note C1 (the lowest C on the piano). The initial decay is dominated by the decay of vibrational motion that is initially along y. This decay is rapid because the soundboard is most compliant to forces in this direction. Eventually the vibration along z grows (because of the coupling of energy from the motion along y) and the z-vibration comes to dominate. The decay of this z-vibration is slower than that of the initial (y) vibration because the soundboard is less compliant in that direction, reducing the rate of energy loss from the string. This double decay effect gives the long persistence of a piano tone. The note considered in Fig. 6.3 is produced by a sin-

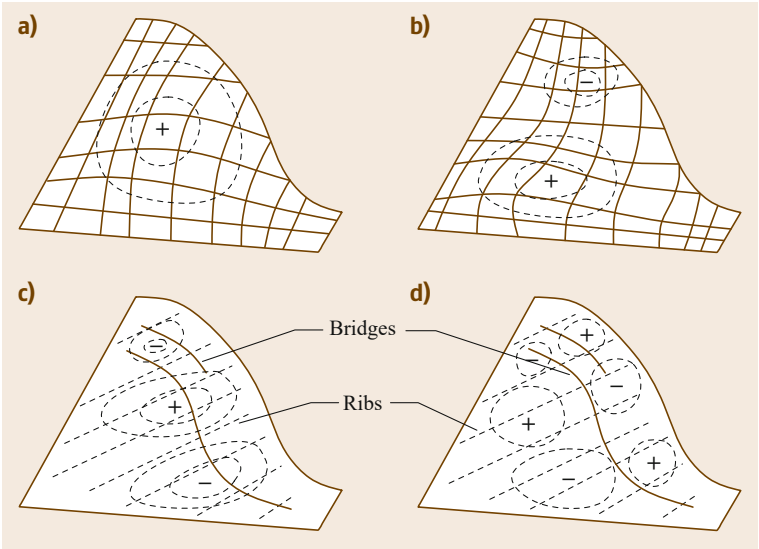


Fig. 6.4 **(a)** The lowest frequency mode of a piano soundboard is a breathing mode, which has its maximum amplitude near the center of the board. For this mode the soundboard undergoes a simple up and down motion with no nodal lines. **(b)** The second soundboard mode is one in which two parts of the board move out of phase, with a nodal line between the two regions. **(c)** The ribs and bridges affect the locations of the nodal lines. This is the third soundboard mode and has two nodal lines. **(d)** Mode with five nodal lines (after [6.6])

gle string. In notes that involve two or three strings one observes a more complicated decay, as the strings beat against each other in addition to the vibrations along y and z. This beating produces a sort of vibrato [6.6, 42], which is apparently a pleasing effect for most listeners.

Piano Soundboards and Sound Generation

Roughly speaking, one can view a piano soundboard as a large speaker as it moves in response to forces from the strings, and to understand the efficiency with which this speaker generates sound one must understand its vibrational modes. In this article we will refer to these as modes of the *soundboard* and early soundboard models [6.12] considered the motion of the soundboard alone. A more accurate description (and more recent modeling) must include the effect of the strings and thus describe the modes of the coupled soundboard–string system. At the level of our discussion here, the strings make a relatively small contribution, but the most sophisticated piano models do now include this effect.

At low frequencies, below a few hundred Hz, it is useful to think about the individual soundboard modes. The lowest is a breathing mode, illustrated in Fig. 6.4a, in which the soundboard moves uniformly up and down, with the largest amplitude near the center. The frequency of this mode for a typical piano is around 50–100 Hz (with larger pianos having a lower frequency for this mode). At higher frequencies one finds modes with two or more antinodes separated by nodal lines; the second mode (with one nodal line) is shown schematically in Fig. 6.4b. As mentioned above in connection with the equation of motion for a soundboard (6.4), soundboards contain ribs and bridges which make the rigidity nonuniform. This affects the deflection shapes and the positions of nodal lines and antinodes. In general and especially at high frequencies, the antinodes will tend to be found between the ribs and somewhat off the bridges (Fig. 6.4c,d). Deflection shapes measured using the Chladni method for grand piano soundboards have been reported by *Conklin* [6.23].

Another way to characterize the soundboard motion is in terms of the mechanical impedance $Z(f)$ which is the ratio of the force applied to the soundboard at a particular frequency to the soundboard velocity at that frequency, or equivalently, the mechanical admittance which is the reciprocal of $Z(f)$. In these measurements, the force and velocity are usually applied and measured at a point on the bridge. Typical results for $Z(f)$ are shown in Fig. 6.5 for the soundboard of an upright piano. The large dip near 100 Hz indicates the fundamental mode, since there is a strong response (a large soundboard velocity) at that frequency. The other large dips in $Z(f)$ indicate additional modes. As the frequency is increased, the average separation between modes approaches a constant that is approximately equal to the fundamental frequency, leading to an approximately constant $Z(f)$ until high frequencies where effects due to ribs and bridges can cause a decrease in the impedance.

At the lowest frequencies (well below the fundamental mode) the impedance increases monotonically and the soundboard vibration is small, resulting in a weak tone. In addition, the efficiency of sound production becomes very small at low frequency, as the sound wavelength becomes much greater than the size of the soundboard. These two effects lead to a phenomenon known as the *missing fundamental*. This effect is illustrated in Fig. 6.6, which shows the spectrum for

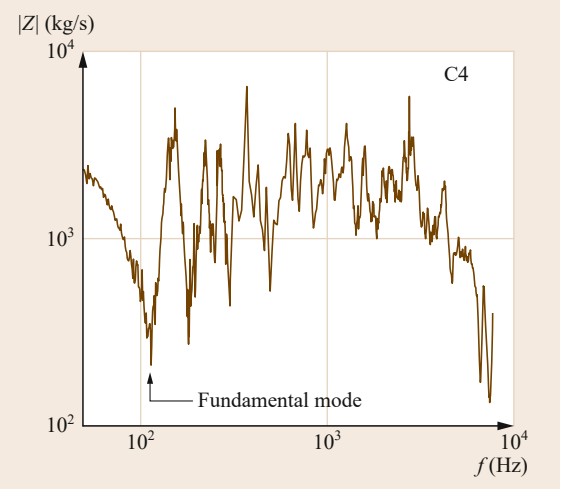


Fig. 6.5 Mechanical impedance of the piano soundboard from an upright piano. The force was applied at the bridge at the point where the strings for middle C are connected and the soundboard velocity was measured at the same location (after [6.43])

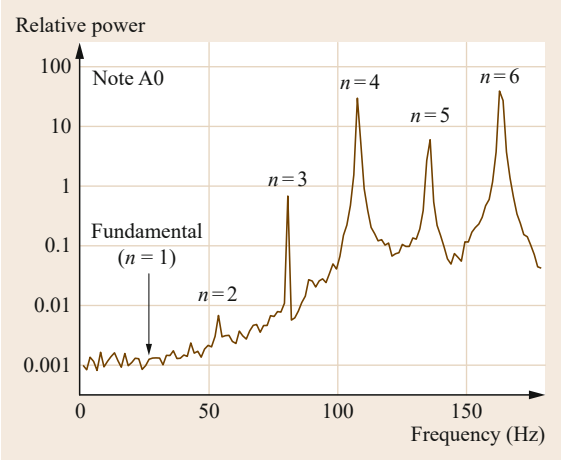


Fig. 6.6 Spectrum of the note A0 for a grand piano. (This is the lowest note on the piano). A number of partials are indicated; the sound power at the fundamental (the first partial) is negligible (after [6.6])

the lowest note on a piano (fundamental frequency 27.5 Hz). A number of partials are indicated and it is seen that the power at the fundamental is extremely small, with the strongest partials found for partial numbers of order 6 or even higher. An interesting point here is that the tone in Fig. 6.6 is perceived to have a pitch corresponding to a fundamental frequency of 27.5 Hz, even though there is negligible sound power at that frequency. One explanation for this phenomenon is that the human nervous system is somehow sensitive to the periodicity of the signal rather than the spectral content [6.44], but it does not appear that this hypothesis can explain fully how our nervous system reacts to tones with missing fundamental components [6.45].

6.3.2 Guitar

Figure 6.7 shows a schematic diagram of an acoustic guitar. In a very general way this is similar to the structure of a piano, with strings that are held rigidly at one end (at the nut, near the tuning pins at the top) and which are connected at their other ends to a bridge which rests on the soundboard (here called the body). The soundboard is typically made of spruce with the sides and back made of a hardwood, usually maple. The strings are often nylon (for a classical guitar) but can be steel. In the bass the strings often have a wire core wrapped with nylon or a thinner wire, to increase the mass without substantially increasing the string stiffness. There are several ways that the bridge can be held in place. In some cases it is glued to the top plate

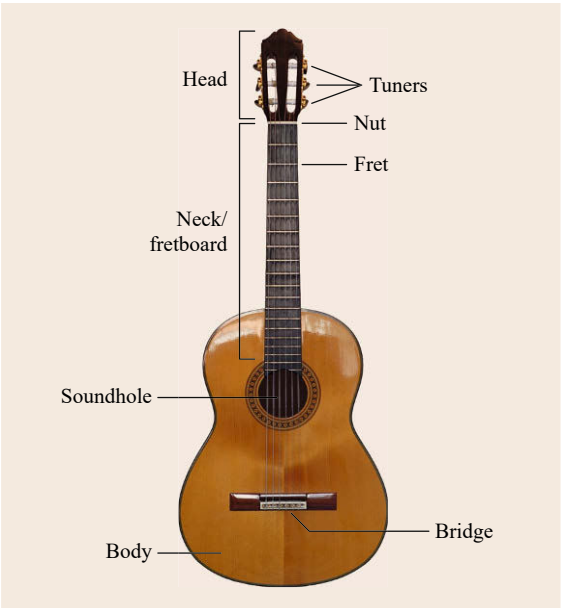


Fig. 6.7 Sketch of a guitar showing its principal components

and the strings are fastened to the bridge. This produces a torque on the bridge which can distort the top plate. Another arrangement, which reduces this torque, has the ends of the strings fastened to the bottom edge of the guitar. With this arrangement it is also possible for the bridge to be held in place by the tension of the strings [6.25].

Vibrations of the strings produce a force on the bridge that sets the soundboard into motion. An impor-

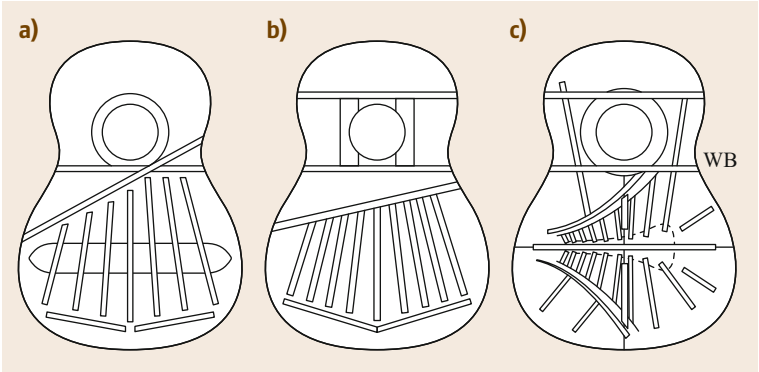


Fig. 6.8a–c Bracing patterns for several classical guitars: **(a)** Ramirez (Spanish), **(b)** Fleta (Spanish), **(c)** Eban (Unites States) (after [6.5])

tant difference between the guitar and the piano is that the soundboard forms one surface of an air cavity and air in this cavity is set into motion by the soundboard. Because of the soundhole, this forms a Helmholtz resonator [6.2] and a number of the important modes of the guitar body are combination modes of this Helmholtz resonator together with vibrations of the soundboard, sides and back. The underside of the top plate is reinforced with ribs adding strength, especially in the across-the-grain direction (perpendicular to the neck). Unlike the case of the piano, for which the arrangement of ribs is fairly standard, many different rib patterns are employed by different guitar makers, a few of which are shown in Fig. 6.8. Other common patterns are described by *French* [6.4].

Plucked Strings – Spectra and More

To understand the spectrum of a guitar tone one must first consider the motion of a plucked string [6.46]. Figure 6.9 shows the behavior for two cases, a string plucked at its center and a string plucked very close to one end. Here we assume for simplicity that the string is perfectly flexible and that the plucking force is applied to an extremely narrow portion of the string. (More realistic situations can be treated but do not change our basic picture; see the references for a discussion of such cases.) With these assumptions the string has a sharp kink at the plucking point. After the string is released this kink splits into two separate kinks that propagate in opposite directions and these travel back and forth along the string as they reflect from the ends. The resulting force from the string on the bridge as a function of time is also shown in Fig. 6.9. This force is equal to the string tension multiplied by the sine of the angle (at the bridge) that the string makes with the axis of the undisplaced string. The kinks maintain their shape as they propagate, and as can be seen from Fig. 6.9 the kink angle is constant for long periods, changing sign when the kinks reflect, leading to a square wave time dependence for the force on the bridge. When the string is plucked at its center the two portions of the square wave are of equal length (Fig. 6.9a), while if the string is plucked near one end the relative times spent in the high and low portions vary according to the location of the plucking point relative to the end of the string (Fig. 6.9c).

As very rough approximations, the velocity of the guitar top plate is proportional to the applied force and the sound produced by the vibrating top plate is proportional to its velocity. As a result, the spectrum of the resulting tone is roughly the same as the spectrum of the force on the bridge. The spectra of the two force–time square waves in Fig. 6.9 are also shown in the figure. In general, for a string plucked $1/n$ of the way from an end, the n-th harmonic and its overtones (frequencies $m \times n$ where m is an integer) are absent. In addition, for n large and for a string plucked very near one end, the amplitudes of the harmonics with orders less than n decrease slowly with n. The timbre of a guitar tone thus depends strongly on the plucking point and plucking near the end of the string enhances the high harmonics.

To this point we have not specified the direction in which the string is plucked. Unlike the piano, for which the string is excited in the direction perpendicular to the soundboard, a guitar player can control the direction of the pluck. In general, a pluck perpendicular to the soundboard will initially excite strong soundboard motion and produce a fast initial decay. Rocking motion of the bridge will then lead to string motion parallel to the soundboard and string motion in that direction will decay more slowly, producing a double decay like that found with the piano. It is also possible for a player to pluck a string in the parallel direction leading to a slower overall decay of the sound. In addition to affecting the time dependence of a tone, these different string motions will excite the soundboard modes differently, thus affecting the timbre. In this way, the player has significant control over several important aspects of the guitar tone.

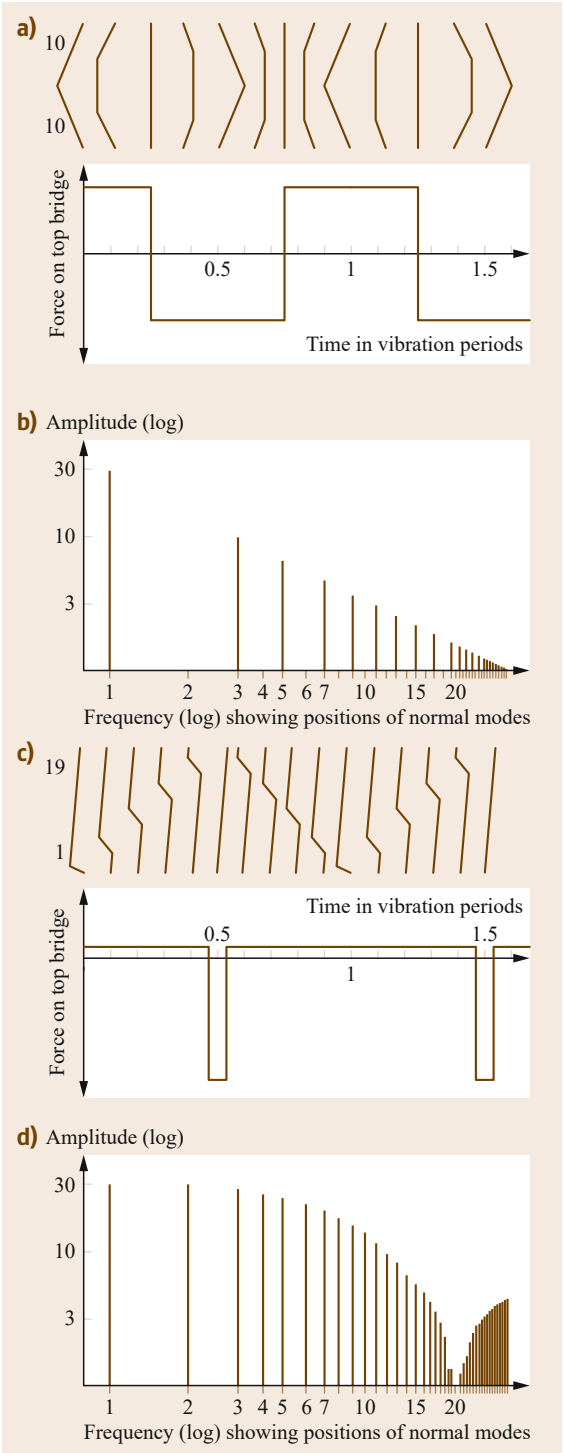


Fig. 6.9 **(a)** Motion of a string plucked at its center. The string profile is shown at the *top*, at various times during the course of its motion, with the axis of the string aligned vertically. **(b)** Spectrum of the force exerted by the plucked string in **(a)** on the support at one end. **(c)** Same as **(a)** but for a string plucked 1/20 of the way from one end. **(d)** Spectrum for the string plucked 1/20 of the way from one end (after [6.46]) ►

While we are concerned in this article mainly with the acoustic guitar, we should mention that the decay and timbre for an electric guitar behave somewhat differently. First, the time decay is very slow, since the bridge of an electric guitar is essentially rigid. Indeed, the desire for a very slow decay was a major motivation for the invention of the electric guitar. Second, the pickup of an electric guitar couples differently to the string motion than do the bridge and soundboard of an acoustic guitar. Third, the timbre is strongly shaped by the characteristics of the amplifier used to convert the signal from the electric guitar pickup coils into sound.

Vibrational Motion of the Top Plate

The response of the guitar body to the force produced by a plucked string depends on the vibrational modes of the body. Figure 6.10 shows the deflection shapes for several modes of an acoustic guitar measured using holographic techniques [6.47]. These deflection patterns show the locations of the antinodes and nodal lines for several of the lowest frequency modes and were measured by driving the top plate at the indicated frequencies. The lowest (fundamental) mode has a single antinode (Fig. 6.10a) and the number of antinodes increases as one progresses to higher modes, a few of which are shown here. Typically the lowest top plate mode has a frequency near or just above 100 Hz, which is above the fundamental frequency of the lowest bass string (82.5 Hz). While we have referred to the modes in Fig. 6.10 as being the modes of the *body* they are actually the modes of the combination of body and the air inside [6.14, 48, 49]. At low frequencies the body and air form a Helmholtz resonator while at higher frequencies one must take account of the nodes and antinodes of the air motion inside the body. In any event, the fact that the lowest mode of the body/air is not far from the frequency of the lowest bass string means that there is a strong response and hence strong sound generation even for the lowest notes, unlike the case with the piano.

These results for the deflection shapes illustrate several important points:

1. Most of the top plate motion is in the region below the soundhole, near the bridge. This is expected, since the string forces act at the bridge.
2. The deflection shapes are often close to being symmetric with respect to the axis defined by the strings and neck, but do exhibit distinct asymmetries. These

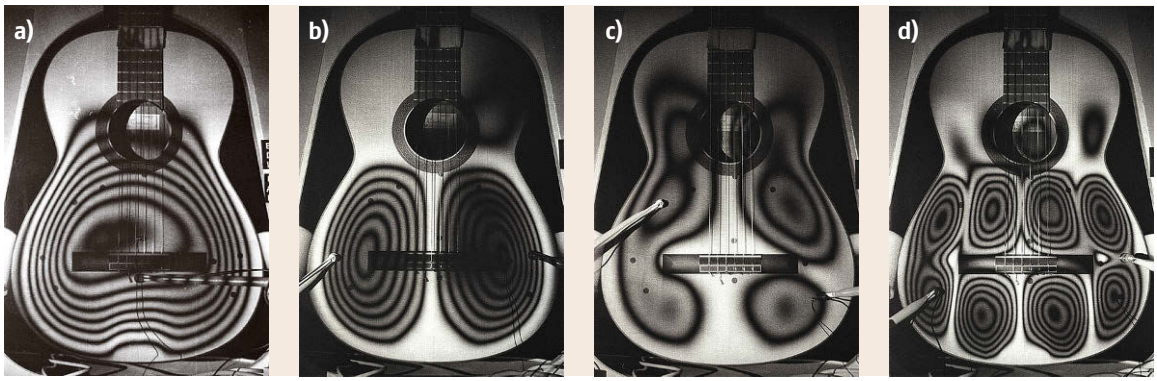


Fig. 6.10a–d Modal deflection shapes showing the top plate motion for several modes of an acoustic guitar. **(a)** The lowest mode, frequency 103 Hz. **(b)** 268 Hz, **(c)** 628 Hz, **(d)** 1010 Hz (after [6.47])

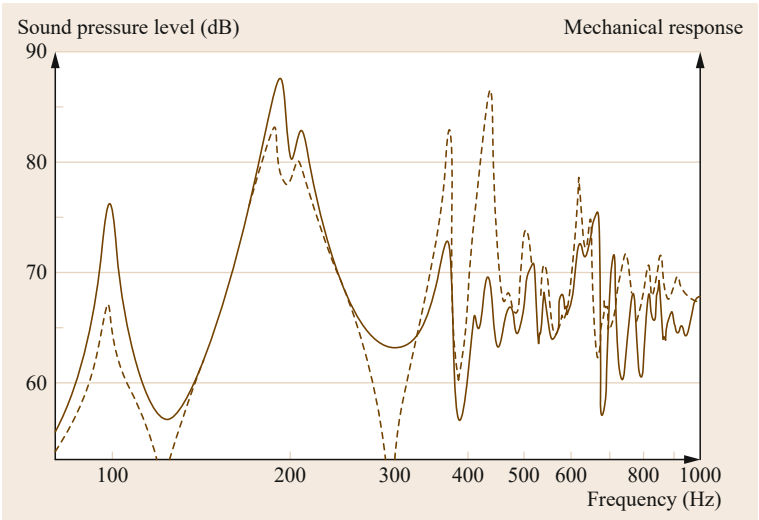


Fig. 6.11 Sound pressure level (*solid curve*) and bridge acceleration (*dashed curve*) as functions of frequency for a harmonic (single frequency) force applied to the bridge, for a Martin D-28 guitar. The force had an amplitude of 0.15 N and the sound was measured 1 m in front of the guitar (after [6.5])

asymmetries are generally caused by an asymmetric rib pattern (Fig. 6.8). Indeed, different rib patterns are presumably designed to emphasize (strengthen) or de-emphasize the sound of particular frequencies and strings, through their effect on the vibrational modes of the soundboard.

3. To understand the contribution of a particular mode to the sound production one must take account of the nature of the driving force.

For example, if a string is plucked in the direction parallel to the top plate, the force from the bridge may give what is effectively a torque on the soundboard, preferentially exciting certain modes.

Sound Generation

The deflection shapes in Fig. 6.10 give insight into the motion of the different soundboard modes, but to get a full picture of the sound generation it is useful to also examine the sound pressure as a function of frequency while driving the bridge at different frequencies [6.5]. Figure 6.11 shows results for both the sound pressure level (solid curve) and the mechanical response of the bridge (dashed curve) as a function of frequency. One would expect the two curves to be similar, as one would expect to find strong sound production at frequencies that produce large soundboard motion. The peaks of the two curves do correspond fairly well at the lower frequencies. However, we do not expect the correspondence to be perfect, since there will be modes at which the bridge motion is small (perhaps dominated by rocking motion) yet the sound production is large. In addition, the radiativity can be very different for different modes. Both of these effects should increase in importance as the frequency is increased and this is probably why the two curves in Fig. 6.11 do not agree as well at higher frequencies.

We should add that measurements like those in Fig. 6.11 have been carried out with different gases in the guitar cavity to determine which modes have significant contributions from Helmholtz resonances of the cavity. Those experiments along with detailed modal calculations have shown that some modes (typically those at the lowest frequencies) contain strong mixtures of the Helmholtz and body (structural) modes [6.48].

6.3.3 Modeling

Models of essentially all components of the guitar have been developed [6.3, 49] and a rather complete picture of all aspects of guitar function is now available. One current focus of modeling is to study how different bracing patterns, different bridge configurations, top plate thicknesses, etc. affect the tone, with the goal of helping guitar makers achieve specific tonal objectives.

6.3.4 Violin

Figure 6.12a gives a sketch of the basic structure of a violin. As with the guitar, the top plate is generally made from spruce while the sides and back are maple. The f-holes allow air to flow into and out of the body as the top plate vibrates, forming a Helmholtz resonator and (as with the guitar) vibrations of the top plate can couple strongly to the Helmholtz resonances. One difference between guitars and violins concerns the structure inside the cavity (Fig. 6.12b). Unlike the guitar, there is no system of ribs on the top plate. Instead, there is a soundpost that connects the top plate with the back and which is held in place by the overall rigidity of the two plates. There is also a single *brace* called the bass bar, that is glued to the underside of the top plate and runs beneath the bass side of the bridge parallel to the strings. We note that the bridge is not glued to the top plate, but is held in place by string tension.

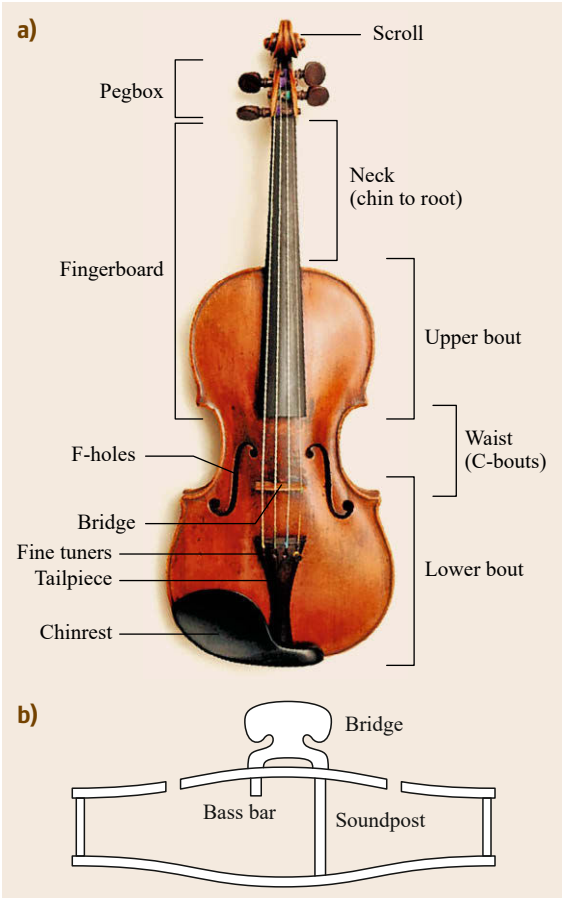


Fig. 6.12 **(a)** Principal components of a violin. Photo from wikimedia commons. **(b)** Schematic cross section of a violin body showing the sound post and bass bar

The Bowed String

An understanding of the motion of a bowed string was first given by Helmholtz, who showed how the stick–slip frictional interaction of a bow with a string can give rise to periodic motion, with the period of a freely vibrating string. It is indeed amazing that a bowed string can vibrate at the frequency of a free string, even while the string is in constant contact with the heavy bow. The nature of this Helmholtz motion is explained in Fig. 6.13. As the bow moves with an approximately constant speed, the string spends a portion of each period sticking to the bow and they move together with a common speed (panels c–h on the left in Fig. 6.13). Eventually sticking cannot be maintained as the required force exceeds the maximum force of static friction and the string slips back rapidly (panels a–c) whereupon the string is *recaptured* by the bow and another interval of sticking occurs. The corresponding string motion is shown in the right panels in Fig. 6.13 which shows how a sharp kink travels back and forth along the string. The force of the string on the bridge is proportional to the sine of the angle that the string makes with its axis and thus varies in a sawtooth manner with time. Note that the force on the bridge is parallel to the plane of the top plate, so the bridge undergoes a rocking motion as the string is bowed.

While the stick–slip motion in Fig. 6.13 yields periodic motion with the desired frequency, this is not the only result that can be obtained with a bowed string. Indeed, anyone who has listened to a beginning player knows that pleasing periodic motion can be quite diffi-

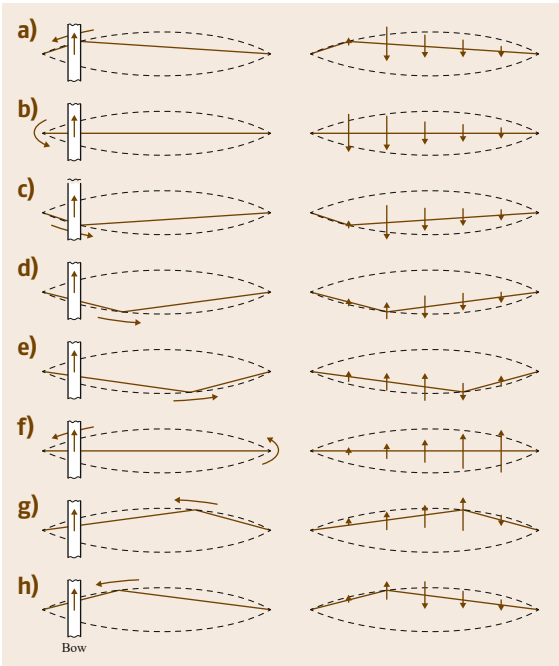


Fig. 6.13a–h The stick–slip frictional interaction of a violin bow (shown on the *left* as a *rectangle* with an *upward arrow* denoting its upward motion) with a violin string. This interaction produces an excitation in which two straight sections of the string meet at a kink, with the kink traveling back and forth on the string (after [6.5])

cult to produce, with creaks and squeals being a much more common result for a beginner. This problem was explored in a famous work by *Schelleng* [6.50] who showed that *good* Helmholtz motion, that is, periodic stick-slip motion with a period equal to the period of the freely vibrating string, is obtained only for a limited range in the parameter space defined by the bow force and the position of the bow as sketched in Fig. 6.14. Here the bow force is the downward force of the bow on the string and the bow position is measured as a fraction of the string length from the bridge.

The Schelleng diagram in Fig. 6.14 is one key to understanding why a violin can be difficult to play, and to a (possible) strategy for making a violin as *playable* as possible – e.g., by making the *desirable* region in Fig. 6.14 as large as possible. But there is another important aspect to the playability of a violin, namely, how quickly Helmholtz motion is established. It has been shown that expert violinists can establish perfect Helmholtz motion within the first few periods or so of the initiation of bowing and that listeners judge tones to be poor if the pre-Helmholtz phase of the tone lasts longer than about 50–90 ms (depending on the precise nature of the tone during that phase [6.51]); this cor-

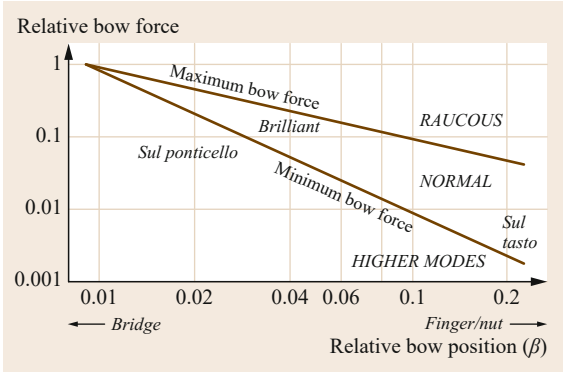


Fig. 6.14 *Schelleng diagram* showing the combination of bow force and bow position that produce Helmholtz motion (the region labeled *normal* and *brilliant*). Outside these regions other vibrational motion of the string is produced (after [6.5])

responds to between 10 and 18 periods of the ideal Helmholtz motion. This result leads naturally to the question of what combination of bowing parameters enables a player to establish perfect Helmholtz motion as quickly as possible, *when starting with the string at rest*. This problem has been studied by *Guettler* who considered the behavior starting from realistic initial conditions with realistic bowing gestures [6.52]. He showed that certain combinations of bow force and bow acceleration (Fig. 6.15) yield the quickest path to the Helmholtz motion when the bow starts from rest. This work has yielded new ways to think about the playability of an instrument.

Modes of the Violin Body and What Makes a Good Violin

The motion of the violin body has been studied using holography and other similar methods to obtain pictures of the deflection shapes of the top plate modes, like those discussed above for the guitar [6.53–55]. As with the guitar, it has been found that several of the lowest modes are mixtures of body modes and resonances of the air in the cavity [6.56]. Studies of these modes have been used to address several of the central questions in violin acoustics, including *what properties make a good violin?* and *what can I measure to tell if I have a good violin?* Let us take the second question first. Figure 6.16 shows the mobility of a violin bridge as a function of frequency. This quantity can be measured by applying a harmonic force to the bridge in the direction parallel to the top plate and perpendicular to the strings and measuring the resulting velocity of the bridge in the same direction. (Note that this mobility is the inverse of the impedance plotted in Fig. 6.5) Each peak in Fig. 6.16 corresponds to a vibrational mode; studies of

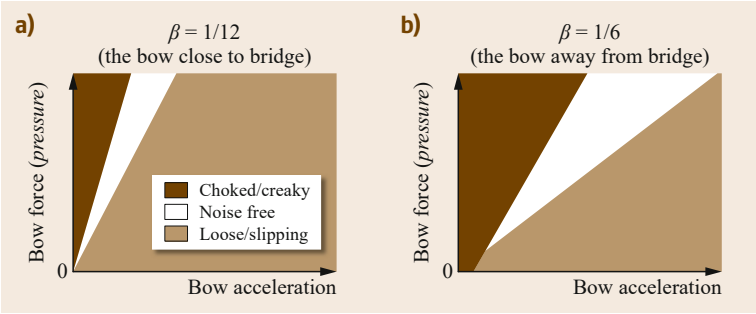


Fig. 6.15a,b Diagrams showing the combinations of bow force and bow acceleration that produce good Helmholtz motion quickly after initiation of bowing. For two different locations of the bow relative to the bridge: **(a)** close to the bridge; **(b)** relatively far from the bridge (after [6.5, 52])

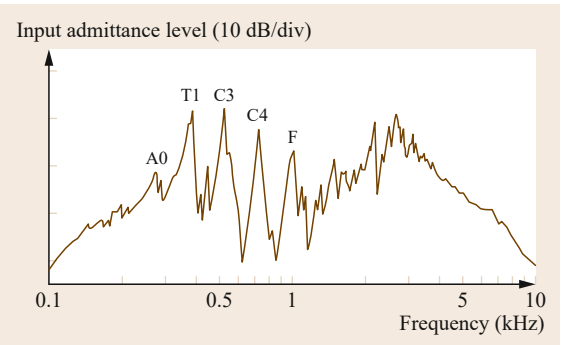


Fig. 6.16 Mobility (also called the admittance) of a violin bridge as a function of frequency (after [6.5, 57, 58])

violins containing different gases have shown that the mode labeled *A*0 in Fig. 6.16 involves strong mixing with a Helmholtz resonance, similar to that found with the guitar. In addition, certain features of the mobility curve, including the so-called *bridge hill* (labeled F in Fig. 6.16) have been found to correlate well with violin quality [6.57, 58]. As this term implies, the vibrational modes of the bridge play a role in determining the mobility at frequencies above a few kHz.

The modes of the top and bottom plates have also been identified as playing an important role in construction of a violin. As discussed by *Hutchins* [6.53, 54], many violin makers listen to the tap tones of the top and bottom plates while these plates are being brought to final form. In order to produce a high-quality violin, it is desirable to have certain relationships between the mode frequencies of these two plates. It is believed that this approach was practiced by the legendary violin makers.

The problem of understanding the origin of violin quality fascinates violinists and nonviolinists alike. That question can be stated as *what is the secret of master violin makers such as Stradavari?* and *why are the legendary (and old) violins best*? The first question remains unsettled (despite many claims about special varnish, etc.), but the second question has been addressed in some intriguing recent studies. Careful double blind comparisons of old and new violins have suggested that the violins made by the best modern makers are at least equal to the best old instruments [6.59]. This result will probably be hard for some to accept.

6.3.5 Violins are Complicated

Because of space limitations, this article has not been able to cover many interesting issues. This has been especially the case for the violin. Indeed, while the violin has been studied more than any other instrument, the physics of bowed strings and the complex shape and structure of the violin lead to many complications that are still the object of intense study [6.60]. These complications include torsional motion of the string due to the torque produced during bowing, devising realistic descriptions of the frictional force between the bow and string, the effects of finite bow width, string stiffness and string damping, the effect of the body modes on string motion and the different radiativities of different vibrational modes of the violin body and of different violins [6.61]; and this is by no means a complete list!

6.4 Summary

This chapter has reviewed some essential elements of stringed instruments. We have tried to emphasize the common features of all stringed instruments while highlighting the important differences. Because of space limitations we have had to omit many interesting issues. Hopefully the references contained herein will give the interested reader a window into further study of this important class of instruments.

Acknowledgments. I am grateful to A. Askenfelt, R. Bader, A. Chaigne, T. Rossing, and G. Weinreich for many enlightening discussions and to the authors who have graciously allowed me to show figures with their results. I also thank R. Bader for inviting me to contribute to this chapter.

References

6.1 L. Cremer: *The Physics of the Violin* (MIT Press, Cambridge 1985)
6.2 N.H. Fletcher, T.D. Rossing: *The Physics of Musical Instruments* (Springer, New York 1991)
6.3 R. Bader: *Computational Mechanics of the Classical Guitar* (Springer, Berlin 2008)
6.4 R.M. French: *Engineering the Guitar* (Springer, New York 2009)
6.5 T.D. Rossing (Ed.): *The Science of String Instruments* (Springer, New York 2010)
6.6 N.J. Giordano: *Physics of the Piano* (Oxford University Press, Oxford 2010)
6.7 A. Chaigne, J. Kergomard: *Acoustique des Instruments de Musique* (Belin, Paris 2010)
6.8 A. Chaigne: On the use of finite differences for musical synthesis. Application to plucked string instruments, J. Acoustique **5**, 181–211 (1992)
6.9 A. Chaigne, A. Askenfelt: Numerical simulations of piano strings. I. Physical model for a struck string using finite difference methods, J. Acoust. Soc. Am. **95**, 1112–1118 (1994)
6.10 A. Chaigne, A. Askenfelt: Numerical simulations of piano strings. II. Comparisons with measurements and systematic exploration of some hammer-string parameters, J. Acoust. Soc. Am. **95**, 1631–1640 (1994)
6.11 J. Chabassier, A. Chaigne, P. Joly: Modeling and simulation of a grand piano, J. Acoust. Soc. Am. **134**, 648–665 (2013)
6.12 N. Giordano: Simple model of a piano soundboard, J. Acoust. Soc. Am. **102**, 1159–1168 (1997)
6.13 W. Strutt (Lord Rayleigh): *Theory of Sound* (Dover, New York 1945)
6.14 E. Bechache, A. Chaigne, G. Deveraux, P. Joly: Numerical simulation of a guitar, Comput. Struct. **83**, 107–126 (2005)
6.15 A. Mamou-Mani, J. Frelat, C. Besnainou: Numerical simulation of a piano soundboard under downbearing, J. Acoust. Soc. Am. **123**, 2401–2406 (2008)
6.16 A. Mamou-mani, J. Frelat, C. Besaniou: Prestressed soundboards: Analytical approach using simple systems including geometric nonlinearity, Acta Acoust. united Acust. **95**, 915–928 (2009)
6.17 A. Chaigne, B. Cotté, R. Viggiano: Dynamical properties of piano soundboards, J. Acoust. Soc. Am. **133**, 2456–2466 (2013)
6.18 P.M. Morse, K.U. Ingard: *Theoretical Acoustics* (McGraw-Hill, Princeton 1968)
6.19 D. Botteldooren: Acoustical finite-difference time-domain simulation in a quasi-Cartesian grid, J. Acoust. Soc. Am. **95**, 2313–2319 (1994)
6.20 D. Botteldooren: Finite-difference time-domain simulation of low-frequency room acoustic problems, J. Acoust. Soc. Am. **98**, 3302–3308 (1995)
6.21 N. Giordano, M. Jiang: Physical modeling of the piano, Eur. J. Appl. Signal Process. **7**, 926–933 (2004)
6.22 H.A. Conklin Jr.: Design and tone in the mechanoacoustic piano. Part I. Piano hammers and tonal effects, J. Acoust. Soc. Am. **99**, 3286–3296 (1996)
6.23 H.A. Conklin Jr.: Design and tone in the mechanoacoustic piano. Part II. Piano structure, J. Acoust. Soc. Am. **100**, 695–708 (1996)
6.24 H.A. Conklin Jr.: Design and tone in the mechanoacoustic piano. Part III. Piano and scale design, J. Acoust. Soc. Am. **100**, 1286–1298 (1996)
6.25 M. French: *Technology of the Guitar* (Springer, New York 2012)
6.26 D.E. Hall: Piano string excitation in the case of small hammer mass, J. Acoust. Soc. Am. **79**, 141–147 (1986)
6.27 D.E. Hall: Piano string excitation II: General solution for a hard narrow hammer, J. Acoust. Soc. Am. **81**, 535–546 (1987)
6.28 D.E. Hall: Piano string excitation III: General solution for a soft narrow hammer, J. Acoust. Soc. Am. **81**, 547–555 (1987)
6.29 D.E. Hall, A. Askenfelt: Piano string excitation V: Spectra for real hammers and strings, J. Acoust. Soc. Am. **83**, 1627–1638 (1987)
6.30 D.E. Hall: Piano string excitation VI: Nonlinear modeling, J. Acoust. Soc. Am. **92**, 95–105 (1992)
6.31 T. Yanagisawa, K. Nakamura, H. Aiko: Experimental study on force-time curve during the contact between hammer and piano string, J. Acoust. Soc. Jpn. **37**, 627–632 (1981)
6.32 T. Yanagisawa, K. Nakamura: Dynamic compression characteristics of piano hammer, Trans. Musical Acoust. Tech. Group Meet. Acoust. Soc. Jpn. **1**, 14–17 (1982)
6.33 T. Yanagisawa, K. Nakamura: Dynamic compression characteristics of piano hammer felt, J. Acoust. Soc. Jpn. **40**, 725–729 (1984)
6.34 A. Stulov: Hysteretic model of the grand piano hammer felt, J. Acoust. Soc. Am. **97**, 2577–2585 (1995)
6.35 N. Giordano, J.P. Winans II: Piano hammers and their force compression characteristics: Does a power law make sense?, J. Acoust. Soc. Am. **107**, 2248–2255 (2000)
6.36 H. Fletcher, E.D. Blackham, R. Stratton: Quality of piano tones, J. Acoust. Soc. Am. **34**, 749–761 (1962)
6.37 N. Giordano: Evolution of music wire and its impact on the development of the piano, Proc. Meet. Acoust. **12**, 035002 (2011)
6.38 O.L. Railsback: Scale temperament as applied to piano tuning, J. Acoust. Soc. Am. **9**, 274 (1938)
6.39 O.L. Railsback: A study of the tuning of pianos, J. Acoust. Soc. Am. **10**, 86 (1938)
6.40 O.H. Schuck, R.W. Young: Observations on the vibration of piano strings, J. Acoust. Soc. Am. **15**, 1–11

(1943)
6.41 N. Giordano: Explaining the Railsback stretch in terms of the inharmonicity of piano tones and sensory dissonance, J. Acoust. Soc. Am. **138**, 2359–2366 (2015)
6.42 G. Weinreich: Coupled piano strings, J. Acoust. Soc. Am. **62**, 1474–1484 (1977)
6.43 N. Giordano: Mechanical impedance of a piano soundboard, J. Acoust. Soc. Am. **103**, 2128–2133 (1998)
6.44 W.M. Hartmann: *Signals, Sound and Sensation* (AIP, Woodbury 1997)
6.45 J.G. Roederer: *Introduction to the Physics and Psychophysics of Music* (Springer, New York 2008)
6.46 N.H. Fletcher: *Physics and Music* (Heinemann Educational Australia, Port Melbourne 1976)
6.47 B.E. Richardson, G.W. Roberts: The adjustment of mode frequencies in a guitar: A study by means of holographic interferometry and finite element analysis. In: *Proc. SMAC 83. R. Swedish Acad. Music, Stockholm* (1985) pp. 285–302
6.48 M.J. Elejabarrieta, A. Ezcurra, C. Santamaria: Coupled modes of the resonance box of the guitar, J. Acoust. Soc. Am. **111**, 2284–2292 (2002)
6.49 G. Derveaux, A. Chaigne, P. Joly, E. Bécache: Time-domain simulation of a guitar: Model and method, J. Acoust. Soc. Am. **114**, 3368–3383 (2003)
6.50 J.C. Schelleng: The bowed string and the player, J. Acoust. Soc. Am. **53**, 26–41 (1973)
6.51 K. Guettler, A. Askenfelt: Acceptance limits for the duration of pre-Helmholtz transients in bowed string attacks, J. Acoust. Soc. Am. **101**, 2903–2913 (1998)
6.52 K. Guettler: On the creation of the Helmholtz motion in bowed strings, Acta Acoust. united Acust. **88**, 970–985 (2002)
6.53 C.M. Hutchins: Plate tuning for the violin maker, J. Catgut Acoust. Soc. **39**, 25–32 (1983)
6.54 C.M. Hutchins: Note for the violin maker in free plate mode tuning and plate stiffness, J. Catgut Acoust. Soc. **1**(II), 25–30 (1989)
6.55 C.M. Hutchins, A.S. Hopping, F.A. Saunders: Subharmonics and plate tap tones in violin acoustics, J. Acoust. Soc. Am. **32**, 1443–1449 (1960)
6.56 G. Bissinger, C.M. Hutchins: Evidence for the coupling between plate and enclosed air vibrations in violins, J. Catgut Acoust. Soc. **39**, 7–14 (1983)
6.57 E.V. Jansson: Admittance measurements of 25 high quality violins, Acta Acoust. united Acust. **83**, 337–341 (1997)
6.58 E.V. Jansson: Acoustics for Violin and Guitar Makers, http://www.speech.kth.se/music/acviguit4 (2002)
6.59 C. Fritz, J. Curtin, J. Poitevineau, P. Morrel-Samuels, F.-C. Tao: Player preferences amoung new and old violins, Proc. Nat. Acad. Sci. **109**, 760–763 (2012)
6.60 J. Woodhouse, P.M. Galluzzo: The bowed string as we know it today, Acta Acoust. united Acust. **90**, 579–589 (2004)
6.61 G. Bissinger: Structural acoustics of good and bad violins, J. Acoust. Soc. Am. **124**, 1764–1773 (2008)

7. Modeling of Wind Instruments

Benoit Fabre, Joël Gilbert, Avraham Hirschberg

Wind instruments driven by a constant pressure air reservoir produce a steady oscillation and associated sound waves. This self-sustained oscillation can be explained in terms of a lumped element feedback loop composed of an exciter, such as a reed-valve or an unstable jet, coupled to an acoustical air column resonator, usually a pipe. In this chapter this simplified model is used to classify wind instruments. Five prototype wind instruments are selected: the clarinet, the oboe, the harmonica, the trombone and the modern transverse flute. The elements of this feedback loop are described for each instrument. In simplified models the player is reduced to the role of a pressure reservoir. The player's control, also called the embouchure is however essential. This aspect is discussed briefly for each instrument.

7.1 A Classification of Wind Instruments

A wind instrument driven by an air flow from a reservoir at constant pressure will, for appropriate pressure ranges, oscillate at a constant pitch (frequency) and amplitude. This sustained oscillation will generate sound waves that reach the listener.

While the steady oscillation is very important under normal playing conditions, other aspects such as slow variations of control parameters, rapid transients, *vibrato* and broad-band turbulence flow noise are essential for musical sound quality. The discussion presented here will however mainly be based on the description of the steady oscillation. The other aspects, which are closely related to the player's control of the sound quality, will be introduced as complementary information.

A systematic description of wind instruments is provided by *Baines* [7.1]. The physics of wind instruments has been extensively discussed in many textbooks [7.2–11]. Some additional references will be provided for specific topics or can be found in a review paper [7.12].

The self-sustained oscillation in wind instruments is driven by a feedback loop (Fig. 7.1). A pressure reservoir is the energy source. One distinguishes *reed instruments* in which the air flow is modulated by an oscillating mechanical valve from the *flue instruments* in which mechanical vibrations are negligible (Fig. 7.2). The exciter is an amplifier. It is fed by acoustic oscillation in the resonator. It drives a sound source that reinforces the acoustic oscillation in the resonator. Under normal playing conditions the oscillation mode corresponds to a frequency (pitch) close to an acoustical resonance of the pipe. The pipe behaves then as a narrow-band filter. In order to reach a steady oscillation, a nonlinear saturation mechanism must exist to limit the amplitude of the oscillation of a feedback loop. Hence wind instruments are essentially nonlinear oscillators. This implies that next to stable limit cycle oscillation, they can display chaotic behavior [7.13, 14].

In reed instruments the flow control is due to separation of the flow from the wall, resulting in the formation of a jet downstream of the valve (Fig. 7.2a). The jet thickness is determined by the valve opening. In the jet, the kinetic energy of the flow is dissipated by a chaotic motion called turbulence. There is almost no pressure recovery. This makes the volume flux through the valve vary proportionally to the valve opening. When the valve oscillates the oscillating volume flux acts as a piston (speaker) on the air column in the pipe. This drives acoustic waves in the pipe. The resulting oscillating pressure in the pipe controls the valve oscillation, which closes the feedback loop.

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_7

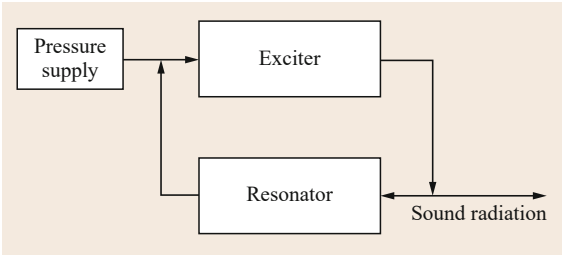


Fig. 7.1 A woodwind instrument uses a continuous pressure supply to generate an oscillating flow, which drives sound waves. This is obtained by a feedback loop consisting out of a nonlinear *exciter* (amplifier) coupled to an acoustical *resonator* (narrow-band filter), usually a pipe. One distinguishes *reed instruments* in which the *exciter* is an oscillating mechanical valve from *flue instruments* driven by a flow instability without significant mechanical vibration (Fig. 7.2)

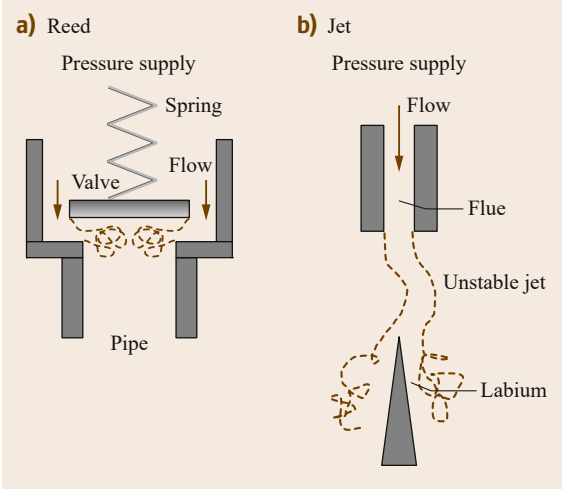


Fig. 7.2a,b One distinguishes *reed instruments* from *flue instruments* (Fig. 7.1): **(a)** The sound generation in reed instruments is driven by the mechanical oscillation of a valve controlling the flow entering a resonator (pipe). **(b)** Flue instruments are driven by an intrinsically unstable flow, usually a jet impinging on a sharp edge called the labium attached to the resonator (pipe)

In flue instruments the flow oscillation is a result of an instability of the air flow. The player produces a narrow relatively high-speed air jet. A *jet* is a narrow high-speed flow surrounded by stagnant fluid. Jets are intrinsically unsteady. Their movement is controlled by the oscillating flow associated with the acoustic oscillation of the resonator. This flow is in the direction perpendicular to the jet flow. It induces perturbations in the jet at the origin of the jet (flue exit), which are amplified as they are convected towards a sharp edge called the *labium*. It is the force of this edge on the flow that is the source of sound driving the acoustic oscillation in the resonator. Some authors refer to this jet as an air reed. This is rather confusing, as the sound generation mechanism is essentially different from that in reed instruments as will be explained in detail later (Sect. 7.6). Flue instruments appear all over the world in different cultures and have very different geometries and playing techniques. The flue instruments that have been most widely studied from the physical modeling point of view are the modern transverse flute, the recorder, the shakuhachi, the whistle and the flue organ pipe.

The oscillating mechanical valve in reed instruments is called a reed, because it is in some instruments made of a bamboo-like wood. Reeds can be a single striking reed (clarinet, saxophone, reed organ pipe), double reeds (oboe, bassoon), free reeds (harmonica) and lips (brass instruments). *Helmholtz* [7.2] distinguishes reeds that close when the difference $p_r = p_m - p_p$, between the supply (mouth) pressure p_m and the pipe inlet pressure p_p, is increased from reeds that open when p_r increases, respectively the *closing and opening reeds* (Fig. 7.3). Helmholtz demonstrated that the oscillation frequency of a closing reed will be lower than both the reed mechanical resonance frequency and the pipe acoustical resonance frequency. The opening reeds should have the opposite behavior; the oscillation frequency is higher than the reed and the pipe resonance frequencies. The closing reeds of instruments such as the clarinet, the oboe and the reed organ pipe do indeed display this behavior. One defines brass instruments as lip-driven wind instruments; they are not necessarily made of brass. The lips of a brass player do, in first approximation, behave as opening reeds. However, lips' behavior can be much more complex because they can display self-sustained oscillations involving at least two mechanical degrees of freedom, which will be discussed later. In this sense lips can behave as vocal folds.

One distinguishes striking reeds (single striking reeds, double reeds and lips) from free reeds. The motion of striking reeds is limited by full closure. The reeds sketched in Fig. 7.3 are striking reeds. In principle they can close completely. Free reeds are thin tongues attached on one side to a plate containing an aperture through which the reed can pass. They act as swinging doors. There always remains a thin slit opening between the tongue and the support plate. The reed never closes completely. When the tongue is fixed on the high pressure side of the plate, it will initially behave as a closing reed. At high pressures p_r, if the plate is thin enough, the tip of the tongue can pass to the low-pressure side of the plate so that the reed will start behaving as an opening reed.

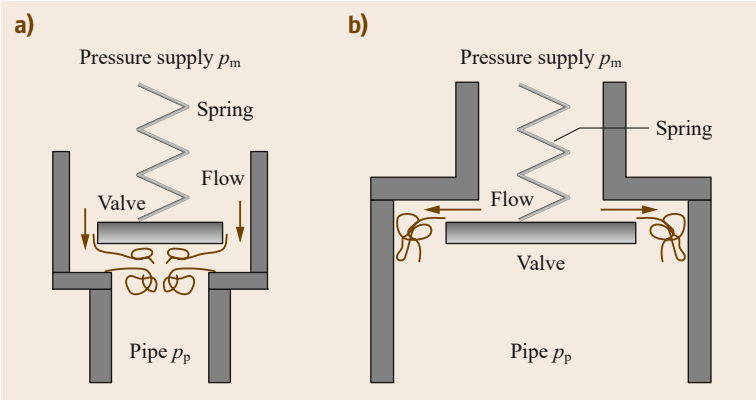


Fig. 7.3a,b One distinguishes **(a)** *closing reeds* from **(b)** *opening reeds*: A closing reed closes as the pressure difference $p_m - p_p$ increases. In opening reeds we have the opposite behavior. The clarinet and oboe have closing reeds. The lips of a brass player can in first-order approximation be described as an opening reed

An important parameter is the ratio f_{reed}/f_{pipe} of the reed mechanical resonance frequency and the acoustical pipe resonance frequency. In the normal playing range of the clarinet, saxophone, oboe and bassoon the reed frequency f_{reed} is much higher than the pitch f of the note played, which is close to the pipe frequency f_{pipe}. In most reed organ pipes and brass instruments f_{reed} is close to the pitch f of the note played. A brass player will adjust the lip vibration frequency f_{reed} by tuning the muscular tension. The player is able to do this without contact of the lips with the instrument or by placing their lips against the mouthpiece without a pipe [7.15]. In Asian free-reed instruments such as the sheng, sho and khaen, the pitch is determined by a pipe acoustical resonance frequency f_{pipe}. In Western free-reed instruments, such as the harmonica, harmonium and accordion, the pitch is close to f_{reed}.

In order to provide a more specific discussion, five prototype instruments have been selected: the clarinet, the oboe, the harmonica, the trombone and the flute. The following Sects. 7.2–7.6 are devoted to these instruments.

The actual geometry of the pipe of woodwinds is much more complex than a simple cylinder or a truncated cone. For example, in instruments with side holes, the closed holes form cavities that significantly affect the effective acoustic pipe length [7.6], while we will see that open holes promote sound radiation at high frequencies. However, most instruments display either a dominantly *cylindrical* or *conical* behavior. This explains some essential differences between these families of instruments that will be discussed later. The clarinet has a single reed and an approximately cylindrical pipe. The oboe is commonly played with a double reed and has a more or less conical pipe. The harmonica is a free-reed instrument with a less obvious coupling to a resonator. The trombone is a typical brass instrument with a large portion of cylindrical bore. The modern transverse flute designed by *Boehm* [7.16] is the most sophisticated flue instrument. The main pipe is almost cylindrical but the mouthpiece (head) as a complex geometry [7.17].

In our classification of wind instruments, we could also have considered the typical blowing pressure p_m and the volume flow Φ_r. The oboe is a high-pressure low-volume flow instrument. Typically mouth pressures of a few kPa are used. In the flute blowing pressures are commonly lower than 1 kPa but the volume flow is high. In brass instruments at low pitches the volume flow is high and the blowing pressure relatively low. For high pitches the blowing pressure is high and the flow relatively low. As a rule of the thumb the amplitude of the acoustical pressure oscillation within the pipe of the wind instruments is of the order of p_m.

The musical sound that reaches the listener is a minute fraction of the pressure oscillations within the pipe. Upon radiation there is a strong shift towards high frequencies, which are much more efficiently radiated than the fundamental oscillation frequency, driving the feedback loop. There is therefore no simple relationship between the amplitude (sound pressure level) of the radiated sound and the driving pressure p_m. Furthermore, the perceived sound pressure level depends on the spectral distribution of the sound.

7.2 The Clarinet

Globally the clarinet (with closed tone holes) is a cylindrical pipe terminated by a short bell (horn). Tone holes can be open; this reduces the effective acoustic length and increases the pitch. The acoustics of tone holes is

Fig. 7.4 (a) Clarinet (courtesy of Buffet-Crampon Co.) and (b) its mouthpiece with reed (courtesy of Vandoren Co.)

discussed in depth by *Nederveen* [7.6] and *Dalmont* et al. [7.18]. At the inlet, the mouthpiece opening is partially covered by a thin wedge-shaped slice of cane wood (reed). The thick end of the reed is rigidly attached to the mouthpiece (Fig. 7.4). The free end of the reed is pressed by the player's (lower) lip against the side walls of the mouthpiece opening – the lay. The curvature of the lay, the position of the lip on the reed and the lip pressure determine the initial height h_0 of the slit through which air is blown into the pipe.

The thin reed tip can be described as a spring with stiffness K_r, which bends under influence of the difference $p_r = p_m - p_p$ between the player's mouth pressure p_m and the pressure p_p in the mouthpiece, at the inlet of the pipe. In the framework of a lumped model approximation the reed is described as a mass-spring system

$$m_r \frac{d^2 h_r}{dt^2} + \gamma_r \frac{dh_r}{dt} + K_r(h_r - h_0) = -S_r p_r \,, \qquad (7.1)$$

where $h_r \geq 0$ is the slit opening at the tip of the reed, m_r is the effective mass of the reed, γ_r is the damping and S_r is the effective reed surface. The effective mass m_r is smaller than the actual mass of the reed to account for the fact that the displacement velocity is

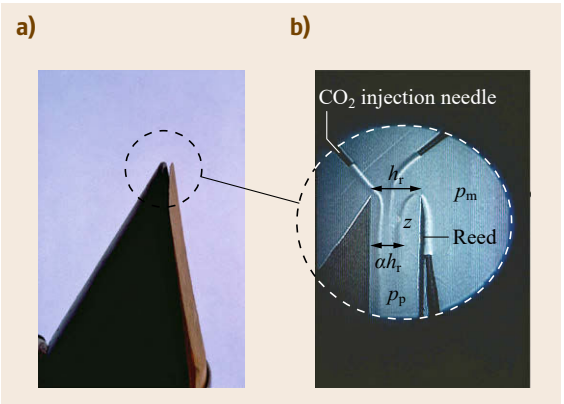


Fig. 7.5a,b Flow through reed channel of height h_r: (a) Tip of mouthpiece and reed (b) Schlieren flow visualization (after [7.9]); stream lines are made visible in an upscaled ($10\times$) static model by injection of carbon dioxide through three needles. We observe that the jet just downstream of the inlet of the reed channel is narrower than the reed channel height h_r by a factor α (vena contracta)

not uniform over the reed surface [7.6]. It varies between zero and dh_r/dt. For the same reason the surface S_r is an effective reed surface, which is not the actual reed surface. Typically the reed resonance frequency $f_{reed} = \sqrt{K_r/m_r}/(2\pi)$ is, for the lower register of a clarinet, much higher than the pitch f of the note played. Without sufficient damping by the player's lip the reed will oscillate close to its resonance frequency, resulting in a nasty *squeak* sound. Note that the drone pipes of a bagpipe are cylindrical and have single reeds. The frequency of these reeds is adjusted to approach the acoustical resonance frequency of the pipe, so that they can be played without lip contact with the reed.

The effective reed stiffness K_r is not constant and is not only determined by the material properties of the reed. In particular, due to the limitation of the reed movement by the lay the effective reed stiffness K_r is a strongly nonlinear function of h_r. The curvature of the lay, the exact shape of the reed and the player's lip position on the reed all have a strong influence on the playing comfort and the quality of the sound produced. Also moisture strongly affects the mechanical properties of wooden reeds.

In first approximation the volume flow Φ_r entering the pipe is given by

$$\Phi_r = \mathrm{sign}(p_r)\alpha W h_r \sqrt{\frac{2|p_r|}{\rho_0}} - A_r \frac{dh_r}{dt} \,, \qquad (7.2)$$

where W is the effective slit width, ρ_0 the air density and A_r the effective surface of the moving part of the reed. The first term of (7.2) is the quasistatic approximation

assuming the formation of a free jet of cross-sectional area $\alpha h_r W$ by flow separation at the sharp tip of the reed (Fig. 7.5). The vena contracta factor α is in the range $0.5 < \alpha < 1.0$ depending on the channel inlet geometry [7.19–21]. The frictionless equation of Bernoulli is used to estimate the velocity in the jet $U_j = \sqrt{2|p_r|/\rho_0}$. It is however important to realize that friction is assumed implicitly, because it causes flow separation, which results in the formation of the jet. The model assumes a dissipation of all the kinetic energy in the free jet by turbulent dissipation. Obviously when the reed is closing completely, beating on the lay, the slit opening will become so small that friction will eventually dominate the entire flow through the reed channel and the equation of Bernoulli should not be used [7.9, 20, 22]. The second term of (7.2), $A_r(dh_r/dt)$, is the volume flow due to the displacement of the reed. As the reed is fixed at one end, its displacement velocity varies between zero and the tip velocity (dh_r/dt). The effective acoustical surface area of the reed A_r is therefore smaller than the actual reed surface area. Inertia and this displacement flow become important at very high pitches. A measure for the importance of unsteadiness is the Strouhal number $\mathrm{Sr}_r = L_r f/\sqrt{2p_m/\rho_0}$ based on the length of the reed channel L_r. Significant deviation from the quasisteady behavior is expected for $\mathrm{Sr}_r \geq 10^{-1}$ as found in the simulations of *da Silva* et al. [7.20].

While (7.1) and (7.2) are essentially nonlinear, the acoustical response of the pipe and the respiratory system to the fluctuations in the volume flow are reasonably well described by linear acoustics. At moderately high amplitudes, the oscillating pressure is dominated by the first harmonic. One can therefore approximate the acoustical response of the system by a purely sinusoidal fluctuation $\hat{\Phi}_r \exp(i\omega t)$ of the reed volume flow around an averaged value $\overline{\Phi}_r$, where $\omega = 2\pi f$ and the complex notation is used ($i^2 = -1$). Similarly one has $p_k = \overline{p}_k + \hat{p}_k \exp(i\omega t)$ $(k = p, m)$, where the complex amplitudes $\hat{p}_k$ are related to the impedances Z_k $(k = p, m)$ of the pipe (p) and the player's respiratory system (m) by

$$\hat{\Phi}_r = \frac{\hat{p}_p}{Z_p} = -\frac{\hat{p}_m}{Z_m} \,. \tag{7.3}$$

This highly simplified model already indicates that the player can influence the oscillation of the instrument, by modifying the geometry of his vocal tract (mouth volume, tongue position, etc.). Hence while typically $|Z_m| \ll |Z_p|$ this allows some bending of the notes [7.23–25] in addition to changes in lip position and its pressure on the reed. Most physical models ignore this subtlety and assume $|Z_m| = 0$.

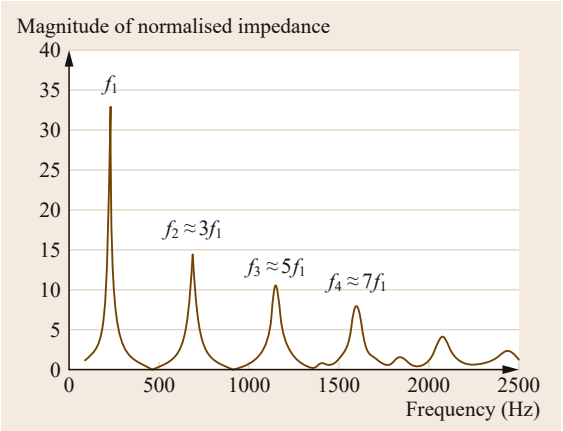


Fig. 7.6 Magnitude on a linear scale of normalized impedance $S_p|Z_p|/(\rho_0 c_0)$ of the pipe of a clarinet, with S_p the pipe inlet cross-section, as a function of the frequency for the note C4 (courtesy J.-P. Dalmont)

The amplitude (modulus) of the impedance of a cylindrical pipe $|Z_p|$ presents strong peaks at frequencies corresponding to resonance frequencies $f_n \simeq (2n-1)c_0/(4L)$, $(n = 1, 2, 3, \ldots)$ with c_0 the speed of sound and L the pipe length (Fig. 7.6). The period of oscillation at the first acoustical resonance (mode) f_1 is four times the travel time of the wave along the pipe. This is due to the fact that the traveling acoustic wave is inverted at the open pipe termination while it is reflected without inversion at the reed side.

The first impedance peak at frequency f_1 corresponding to the quarter-wavelength $(L \simeq c_0/(4f))$ resonance for a pipe length L is the strongest. The peak amplitude decreases with increasing frequency as a result of increasing viscous and thermal dissipation and radiation losses. Furthermore, the first Fourier components of the oscillating pressure in the mouthpiece are dominated by uneven multiples of the oscillation frequency. In order to make the instrument sound at $3f_1$ a register hole has to be opened, which lowers the first impedance peak at f_1.

In occidental music half tones are commonly used. However, for contemporary music quarter tones are used. A special clarinet with a side branch allows the player, by using a key, to shift the clarinet scale by a quarter tone [7.26]. Such a system was also designed for the modern transverse flute.

Linearization of (7.1) to (7.2) for small fluctuations $\hat{h}_r \exp(i\omega t)$ around h_0 yields in combination with (7.3) a homogeneous set of linear equations relating the complex amplitudes $\hat{h}_r, \hat{p}_p, \hat{p}_m$ and $\hat{\Phi}_r$. A nontrivial solution implies that the determinant of this set of equations should vanish, which yields a nonlinear characteristic equation. The real and imaginary parts of this equation

form a set of two equations. Assuming a steady harmonic oscillation, which implies a real frequency ω or equivalently $\mathrm{Im}\{\omega\} = 0$, one can obtain the threshold of the average pressure difference $\overline{p}_\mathrm{r} = \overline{p}_\mathrm{m} - \overline{p}_\mathrm{p}$ above which the instrument will start oscillating and the corresponding oscillation frequency $\mathrm{Re}\{\omega\}$. Alternatively, for given imposed pressure difference $\overline{p}_\mathrm{r} = \overline{p}_\mathrm{m} - \overline{p}_\mathrm{p}$ above the oscillation threshold, the real and imaginary parts of $\omega = \mathrm{Re}\{\omega\} + \mathrm{i}\,\mathrm{Im}\{\omega\}$ are the unknown. One finds for $\mathrm{Im}\{\omega\} < 0$ a complex frequency ω implying an exponential increase of the oscillation amplitude with time following $\exp(-\mathrm{Im}\{\omega\}t)$. As stated earlier, a steady oscillation amplitude is reached by nonlinear saturation of the oscillation amplitude and can therefore not be predicted by a linear model. A nonlinear saturation mechanism has to be taken into account to predict a steady oscillation.

As the amplitude increases a steady oscillation can be predicted by assuming the solution to be a function of a limited number of harmonics of the fundamental frequency. This is the so-called harmonic balance method [7.27, 28]. As stated before, for moderately high amplitudes near the oscillation threshold, an approximative harmonic solution can be considered. At higher amplitudes, more harmonics contribute to the oscillation. The sound becomes *richer* and *louder*.

One can also seek a solution of the equations in the time domain. Equation (7.3) becomes, in the time domain, an integral equation involving a convolution integral. As shown by *McIntyre* et al. [7.29] the model can also be written in terms of the impulse response of an extended semi-infinite pipe, which can be numerically more efficient. One calls this the reflection function approach. Alternatively, the acoustic response of the pipe can be discretized, solving directly for plane-wave propagation within the pipe. This so-called *wave guide* approach has become very sophisticated and is commonly used for sound synthesis by means of real-time solution. This provides so-called *virtual instruments* [7.30–34]. Systematic studies of the response of such models have been provided in the literature, displaying extremely complex behaviors, as do the actual instruments. In particular the virtual clarinet is a versatile musical instrument when it is controlled by a mouthpiece with a pressure sensor.

In the further discussion the reed dynamics are simplified by using the quasistatic approximation of Eqs. (7.1) and (7.2) to give (for $p_\mathrm{r} > 0$)

$$\Phi_\mathrm{r} = \alpha W \left[h_0 - \frac{S_\mathrm{r}}{K_\mathrm{r}} p_\mathrm{r} \right] \sqrt{\frac{2p_\mathrm{r}}{\rho_0}} \,. \tag{7.4}$$

In Fig. 7.7 this relationship between the volume flow Φ_r and the driving pressure difference $p_\mathrm{r} = p_\mathrm{m} - p_\mathrm{p}$

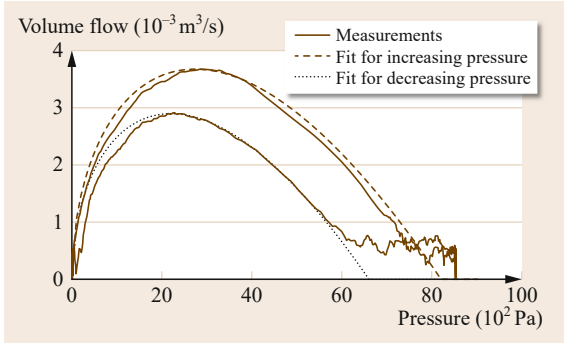


Fig. 7.7 Quasistatic volume flow Φ_r as a function of the driving pressure $p_\mathrm{r} = p_\mathrm{m} - p_\mathrm{p}$ for a closing reed with spring constant K_r, compared to experiments (after [7.19]). The measurements show an hysteresis: the volume flow is larger for increasing pressure than for decreasing pressure. This hysteresis is due to the viscoelastic properties of the reed (after [7.35])

for a constant reed stiffness K_r is compared to experiments [7.19]. Of course Φ_r vanishes at $p_\mathrm{r} = 0$. In other words when we do not blow into the pipe there is no flow. Above the critical pressure difference $p_\mathrm{c} = K_\mathrm{r} h_0 / S_\mathrm{r}$ the reed closes completely and the flow vanishes again. A maximum of reed flow is found, following (7.4), for $p_\mathrm{r} = p_\mathrm{c}/3$. The time-averaged acoustic power $\overline{P}_\mathrm{acoustic}$ generated by a harmonic oscillation of the pipe pressure is given by the time average over a period of oscillation of the product of the reed-volume flow fluctuation $\Phi_\mathrm{r}' = \Phi_\mathrm{r} - \overline{\Phi}_\mathrm{r} = \hat{\Phi}_\mathrm{r} \exp(\mathrm{i}\omega t)$ and the pipe-inlet pressure fluctuation $p_\mathrm{p}' = p_\mathrm{p} - \overline{p}_\mathrm{p} = \hat{p}_\mathrm{p} \exp(\mathrm{i}\omega t)$. The over bar indicates a time averaging over a period of oscillation. Hence the power is given by

$$\begin{aligned} \overline{P}_\mathrm{acoustic} &= \frac{\omega}{2\pi} \int_0^{2\pi/\omega} \Phi_\mathrm{r}' p_\mathrm{p}' \mathrm{d}t \\ &= \frac{1}{2} \left\{ \mathrm{Re}\{\hat{p}_\mathrm{p}\}\,\mathrm{Re}\{\hat{\Phi}_\mathrm{r}\} + \mathrm{Im}\{\hat{p}_\mathrm{p}\}\,\mathrm{Im}\{\hat{\Phi}_\mathrm{r}\} \right\} . \end{aligned} \tag{7.5}$$

For small oscillations, in linear approximation the reed volume flow fluctuation $\hat{\Phi}_\mathrm{r}$ is the product of the slope $\mathrm{d}\Phi_\mathrm{r}/\mathrm{d}p_\mathrm{r}$ of the volume flow characteristic and the amplitude of the pressure fluctuation $\hat{p}_\mathrm{r} = -\hat{p}_\mathrm{p}$, where the negative sign takes into account the fact that an increase of the pipe-inlet pressure p_p results in a decrease of the driving pressure $p_\mathrm{r} = p_\mathrm{m} - p_\mathrm{p}$ across the reed. So we find for the amplitude of the volume flux fluctuations

$$\hat{\Phi}_\mathrm{r} = -\left[\frac{\mathrm{d}\Phi_\mathrm{r}}{\mathrm{d}p_\mathrm{r}} \right] \hat{p}_\mathrm{p} \quad \text{and} \quad \overline{P}_\mathrm{acoustic} = -\frac{\left[\frac{\mathrm{d}\Phi_\mathrm{r}}{\mathrm{d}p_\mathrm{r}} \right] |\hat{p}_\mathrm{p}|^2}{2} \,.$$

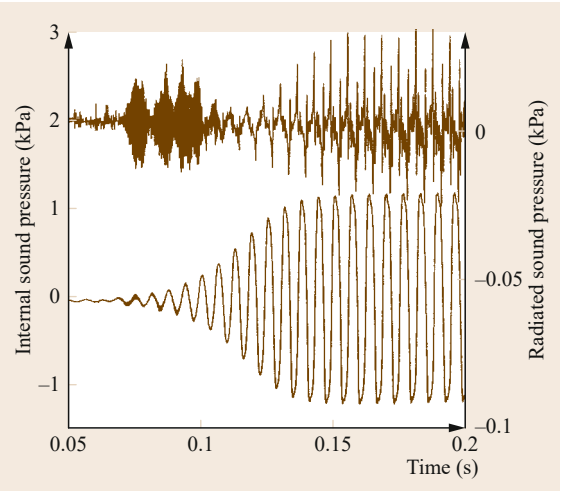


Fig. 7.8 Mouthpiece pressure and radiated pressure during the attack of a clarinet. We observe that a steady oscillation is reached within a few oscillation periods. Furthermore we see that the mouthpiece pressure is dominated by the fundamental, while the radiated sound (music) is much richer in higher harmonics

Hence for $0 < p_r < p_c/3$ as $d\Phi_r/dp_r > 0$ sound is absorbed. Sound is generated by the reed oscillation when $p_c/3 < p_r < p_c$. Indeed the musician will have to overcome a critical mouth pressure threshold of the order of $p_r = p_c/3$ to initiate oscillations. Furthermore at very high blowing pressures the reed indeed closes completely, eventually stopping any oscillation.

Some insight into the finite-amplitude dynamical behavior of the clarinet can be obtained by considering a lossless cylindrical resonator [7.11, 29]. The limit cycle oscillation of this system corresponds to a symmetric square-wave oscillation between the two points p_A and p_B corresponding to a fixed volume flow Φ_r on the two branches ($p < p_c/3$ and $p_c/3 < p < p_c$) of quasistatic reed characteristics. The actual large amplitude oscillation of the pressure in the mouthpiece of a clarinet approaches this behavior. Assuming frequency-independent energy losses upon reflection at the open pipe termination, one obtains an analytical model predicting many aspects of the dynamic response of the clarinet. Due to the fact that the transition from stationary flow to oscillating flow is a direct bifurcation, the player can play a note at low level. The instrument has a fast response and a large dynamical range, from pianissimo to fortissimo.

An essential property of the clarinet is that the limit cycle is reached within a few oscillations (Fig. 7.8). Attack transients are essential for the perception of musical sounds [7.36]. Note that the conical reed instruments, such as the oboe and the bassoon, have an even shorter attack transient than the clarinet, while flue instruments have very complex long attack transients of typically a few dozen oscillation periods [7.37, 38].

Until now we considered the pipe of the instrument as an acoustical resonator imposing the pitch of the note played. The effective acoustic length of the pipe is changed by opening tone holes along the pipe. A complex key system has been placed to facilitate this task. For the lowest harmonics of the fundamental oscillation frequency, the acoustic waves traveling down the pipe are strongly reflected at the open pipe termination. These acoustic waves are more strongly damped by viscous and thermal losses at the pipe walls than by radiation. However without radiation no musical sound would be generated by the instrument. There is a conflict between the necessity to keep acoustic waves traveling inside the pipe, regenerating the oscillation, and the need of radiation for the production of music. This apparent contradiction is resolved by the fact that the instrument is designed to radiate sound at high frequencies, for which our ear is most sensitive. At low frequencies $k_0^2 A < 1$ the real part of the radiation impedance Z_{rad} of an unflanged open pipe of cross-section A increases quadratically with the frequency $\mathrm{Re}\{Z_{rad}\} \simeq \rho_0 c_0 [k_0^2 A/(4\pi)]$. The wave number is defined by $k_0 = 2\pi f/c_0$, with c_0 the speed of sound, the propagation velocity of acoustic waves. For a given oscillation amplitude, this implies that the radiated acoustical power increases quadratically with the frequency for a small pipe outlet cross-section diameter compared to the wavelength.

While the bell of a clarinet has a catenoidal shape rather than an exponential shape [7.6], we will for simplicity discuss the acoustic behavior of an exponential, horn which has a similar behavior. At the lowest pitches a truncated exponential horn radiates select frequencies above a cutoff frequency of a few kHz. This acoustical behavior is predicted when considering plane-wave propagation in a pipe with exponentially varying cross-section $A(x) = A_0 \exp(\beta x)$ with β a positive constant. The linearized mass conservation, assuming a uniform acoustical pressure $p'(x)$ at position x along the pipe and neglecting friction and heat transfer, is

$$\frac{A}{c_0^2}\frac{\partial p'}{\partial t} + \rho_0 \frac{\partial A u'}{\partial x} = 0\,, \tag{7.6}$$

where u' is the fluid velocity. The x-component of the momentum equation is

$$\rho_0 \frac{\partial u'}{\partial t} + \frac{\partial p'}{\partial x} = 0\,. \tag{7.7}$$

This corresponds to the Lagrange–Webster approximation. Substitution of the plane outgoing wave solution

for the pressure $p' = \hat{p}\exp[\mathrm{i}(\omega t - kx)]$ and the velocity $u' = \hat{u}\exp[\mathrm{i}(\omega t - kx)]$, with $\omega = 2\pi f$, yields a set of linear equations for the amplitudes $\hat{p}$ and $\hat{u}$, which has only a nontrivial solution when its determinant vanishes. Using this characteristic equation one obtains

$$k = -\mathrm{i}\beta/2 + k_0\sqrt{1-(\beta/2k_0)^2}\,.$$

For $k_0 > \beta/2$ the wave will propagate and eventually be radiated at the outlet of the horn. Note that the amplitude of this wave decays following $\exp(-\beta x/2)$. This is a consequence of the conservation of the energy flux in the propagating wave as the pipe cross-section increases exponentially. For $k_0 < \beta/2$ the wave is evanescent; its amplitude will decay exponentially as $\exp[-\mathrm{i}kx]$, because k is purely imaginary. If the exponential horn is sufficiently long the low-frequency wave is reflected before it has reached the end of the horn. Hence the horn promotes the radiation of high frequencies $\omega > \beta c_0/2$ and keeps low frequency $\omega < \beta c_0/2$ waves within the instrument. The low frequencies take care of the regeneration of the oscillation. This very simple model can significantly be improved when considering the finite length of the horn and deviations from the plane-wave approximation. This more accurate theory becomes essential for brass instruments that have very large bells. Another function of the bell is to harmonize the resonances of the pipe for the lowest note of the clarinet [7.11].

When a few tone holes of the clarinet are open the horn becomes useless. The pipe segment with a series of open holes behaves as a row of Helmholtz resonators (lattice) with a similar radiation cutoff frequency behavior as an exponential horn [7.11, 39]. This cutoff frequency is about 1400 Hz in the example of clarinet impedance shown in Fig. 7.6. Hence the exact position and size of tone holes is not only important to determine the pitch but also the radiation [7.39, 40]. The radiation does not only involve the acoustical power radiated, but also the directivity of this radiation. In first-order approximation each tone hole behaves as a monopole sound source. The interference between the radiated sound from these different sources determines the directivity.

It is interesting to note that at *fortissimo* levels, the acoustic flow through open tone holes will be nonlinear. Vortex shedding will occur at sufficiently low Strouhal numbers $fd/|\hat{u}_{\mathrm{ac}}| < 1$ based on the tone hole diameter d and the amplitude $|\hat{u}_{\mathrm{ac}}|$ of the surface-averaged acoustic velocity passing through the tone hole. This so-called acoustical streaming induces a jet flow, which can blow the flame of a candle placed close to the tone hole [7.41, 42]. A similar jet flow also occurs inside the instrument, which can slightly affect the temperature profile along the pipe, because the jet flow entering the pipe transports air from the surroundings, which is cooler than the air blown by the musician [7.43]. The vortex shedding induces energy losses that for a given acoustical pressure amplitude in the instrument are inversely proportional to the third power of the tone hole cross-sectional area. Vortex shedding also generates higher harmonics of the fundamental. It is expected that details of the tone-hole design, such as sharp edges, will affect this. In a clarinet, the nonlinearity of the interaction of the reed with the lay is much more important for the sound quality than the vortex shedding at tone holes. It has been argued by *Bouasse* [7.4] that a horn at a pipe termination does not only enhance radiation but also reduces losses due to vortex shedding. This could be particularly significant for brass instruments at low pitches. In a flute or recorder the vortex shedding induced by the acoustic flow through the sharp edge (labium) of the mouth of the instrument appears to be important both for the saturation of the oscillation amplitude and for the sound quality.

7.3 The Oboe

While the clarinet has a *quasi*cylindrical pipe and a single reed the oboe has an almost conical pipe (of length L) that is truncated. The oboe is commonly played with a double reed (Fig. 7.9). The volume between the two reeds corresponds roughly to the volume of the missing part, of the truncated cone, of length r_0.

Single reeds have been developed for the oboe, but are rarely used. The quasisteady flow model described for the single reed of the clarinet appears to be quite reasonable for double reeds. There are some subtle differences between the volume flow characteristic of a double reed and that of a single reed [7.44]. The assumption that the static flow resistance downstream of the reed strongly influences these characteristics [7.9] has not been verified experimentally for the oboe [7.45] and bassoon [7.44]. Some ancient cylindrical double-reed instruments (Rausch pipes) do have a neck just downstream of the double reed, which does drastically influence the reed characteristics. This neck stabilizes the lowest register, which otherwise has a chaotic behavior [7.9].

The essential difference between the oboe and the clarinet is due to the conical shape of the pipe of the oboe. The acoustic field in a conical instrument is in

Fig. 7.9a,b Oboe (a) and its double reed (b) (courtesy of Buffet-Crampon Co.)

first-order approximation described for low frequencies by the modified solution of d'Alembert for harmonic spherical waves

$$rp' = A^+ \exp[\mathrm{i}(\omega t - k_0 r)] + A^- \exp[\mathrm{i}(\omega t + k_0 r)] \,, \tag{7.8}$$

where r is the distance from the apex of the cone and $A^\pm$ are the amplitudes of the outgoing (A^+) and incoming (A^-) waves. This solution is also obtained when using the Lagrange–Webster approximation (7.6) and (7.7), for a pipe cross-section A increasing linearly with r. The factor r at the left-hand side of (7.8) takes into account the energy conservation in the traveling waves, as done by the factor $\exp(-\beta x/2)$ for the exponential horn. The radial fluid particle velocity u'_r is found by substituting the wave solution (7.8) into the momentum equation (7.7) (with x replaced by r)

$$r^2 u'_\mathrm{r} = \frac{1}{\rho_0 c_0}\left\{\left(\frac{1}{\mathrm{i}k_0} + r\right) A^+ \exp[\mathrm{i}(\omega t - k_0 r)] + \left(\frac{1}{\mathrm{i}k_0} - r\right) A^- \exp[\mathrm{i}(\omega t + k_0 r)]\right\} . \tag{7.9}$$

As at the apex of a cone $\lim_{\mathrm{r}\to 0}[r^2 u'_\mathrm{r}] = 0$ we see from this equation that the pressure waves are inverted $A^+ = -A^-$ at the apex of a cone $r = 0$. This remains a fair approximation as long as the missing length of the cone r_0 remains short compared to the wavelength $k_0 r_0 \ll 1$. Combining this condition with the ideal open pipe termination condition $p'(L) = 0$ at $r = r_0 + L$, we find for a cone of length $r_0 + L$ the resonance frequencies $f_n \simeq nc_0/[2(r_0 + L)]$. Consequently conical instruments have maxima of the pipe impedance for each harmonic of the oscillation frequency, including the even harmonics. This is illustrated by the pipe inlet impedance amplitude $|Z_\mathrm{p}|$ shown in Fig. 7.10.

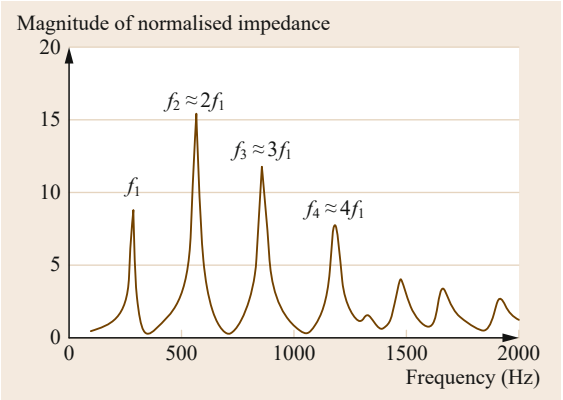


Fig. 7.10 Magnitude of the normalized impedance $S_\mathrm{p}|Z_\mathrm{p}|/(\rho_0 c_0)$, with S_p the pipe inlet cross-section, of the pipe of an oboe as a function of the frequency for the note D4 (courtesy J.-P. Dalmont)

Note that at low frequencies, the peak amplitudes of $|Z_\mathrm{p}|$ increase initially with the frequency. This is due to the increase in radiation impedance of spherical waves at the pipe inlet, as the frequency is increased $\mathrm{Re}\{Z_\mathrm{rad}\} \sim \rho_0 c_0 (k_0 r_0)^2/4$. At higher frequencies the peak in $|Z_\mathrm{p}|$ decreases as a result of increasing viscous/thermal losses and radiation losses, as was already observed in the clarinet. The increase in $|Z_\mathrm{p}|$ with the frequency for the lower frequencies implies that the instrument tends to sound spontaneously an octave higher than the first acoustic mode. Another important consequence of this frequency dependence of the impedance is that the oboe displays an inverse bifurcation. This means that upon a monotonous increase of mouth pressure, at the onset of oscillation the amplitude jumps suddenly towards a finite amplitude [7.46]. It is difficult to play this instrument at piano levels. Another difference with the clarinet is that the oboe will oscillate at the pipe resonance frequency even when the reed is not damped by the player's lips. However, without lip damping the inverse bifurcation behavior is more pronounced. Other conical instruments such as the saxophone and chanter of the bagpipe can also be played without lip contact with the reed.

As in the case of the clarinet, the simplified analytical model of *Chaigne* and *Kergomard* [7.11] provides insight into the limit cycle behavior of the instrument. In a clarinet the time intervals when the reed is closed or opened during an oscillation cycle are roughly equal. In an oboe with a truncated apex of length r_0 the ratio of closed to opened reed time intervals is r_0/L. As typically $r_0 \ll L$ the reed only closes during a short time interval, of the order of magnitude $2r_0/c_0$, during which the pipe inlet pressure p'_p drops suddenly. *Chaigne* and *Kergomard* [7.11] call this a Helmholtz motion in analogy to the velocity of the string in the violin. This yields a series of sharp downwards peaks, corresponding to a spectrum rich in higher harmonics. Therefore the sound of the oboe is much brighter than that of the clarinet. This is one of the reasons that oboe players provide the *A* note used in an orchestra to adjust the tuning of all instruments.

Other instruments, such as the bassoon, the saxophone and the chanter of a bagpipe, also display the typical behavior of conical instruments. Of course the actual geometry of the pipe is more complex than a simple truncated cone. This is necessary to adjust the tuning for all the notes played with the instrument [7.47].

7.4 The Harmonica

The harmonica is a very popular Western free-reed instrument. Like for the harmonium and the accordion the reed is a thin brass tongue. The fine tuning of tongues can be obtained by changing their mass (removing or adding material) or adjusting the reed stiffness in the area close to the clamped end of the tongue. A sometimes complex shape determines the initial opening of the reed (Fig. 7.11).

As shown in the sketch in Fig. 7.12, of the expected quasistatic flow characteristics, the reed behaves at low pressures as a closing reed (Sect. 7.1). Above a critical pressure the volume flow decreases with increasing $p_r = p_m - p_p$. At very high pressures the flow increases again as the reed tip has passed through the plate [7.8] and the reed behaves as an opening reed.

In the harmonica, harmonium and accordion there are no pipes attached to the reeds. It is therefore often assumed that free reeds can oscillate without acoustical feedback. The theoretical models of *St. Hilaire* et al. [7.48] and *Ricot* et al. [7.49] demonstrate that the inertia of the local flow around the reed at the upstream side can induce self-sustained oscillation. Unfortunately these models contain some arbitrariness in the coupling of a two-dimensional near-field flow model to a three-dimensional far-field model. Furthermore, no argument is given to justify neglecting the inertia of the flow at the downstream side of the reed. Furthermore, there is clear experimental evidence for the importance of the acoustical response of the pressure supply system [7.50, 51]. This is most spectacular for the harmonica where the musician modifies the geometry of his vocal tract, taking configurations corresponding to different vowels, in order to bend the pitch. The pitch remains close to the reed frequency f_{reed} and can only be lowered. For the harmonica, this appears to involve a coupling between two reeds attached to a single chamber. The first one oscillates upon blowing while the second is played upon drawing. Experiments

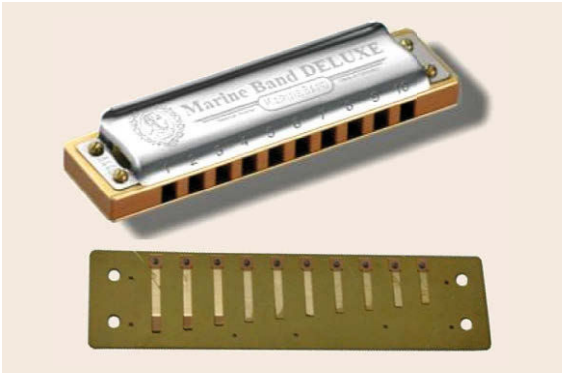


Fig. 7.11 Diatonic harmonica and a row of free reeds (courtesy of Hohner Co.)

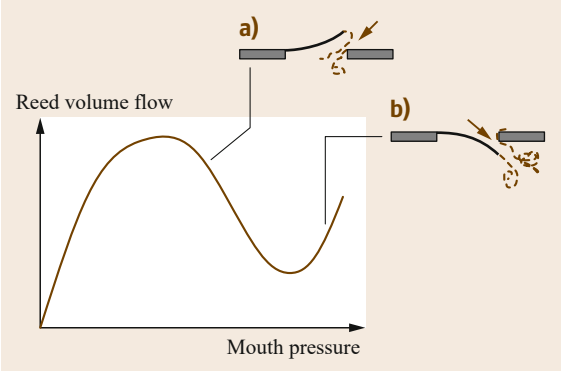


Fig. 7.12a,b Qualitative sketch of the supposed quasistatic volume flow-pressure characteristics of a harmonica reed. **(a)** For low pressures the reed behaves as a closing reed, like the clarinet reed. **(b)** When the reed deflection becomes larger than the plate thickness the reed starts behaving as an opening reed

on a single opening reed demonstrate that also the upstream volume (of the player's mouth) can strongly influence the pressure threshold for the onset of oscillation and the oscillation frequency of a free reed [7.8].

Paradoxically the free-reed instruments seem the most simple wind instruments, but the physical models for these instruments are yet to be confirmed.

7.5 The Trombone

The trombone, like the trumpet, has a long narrow cylindrical section attached to a conical section, which is terminated by a flaring bell (Fig. 7.13). As for the clarinet, the bell promotes the radiation at high frequencies, keeping the low-frequency waves within the pipe. Also the bell harmonizes the resonance frequencies of the pipe [7.7, 8, 11, 39]. A mouthpiece is fixed to the pipe inlet. The rim of the mouthpiece is wider than the pipe. When the player presses their lips against the rim a volume V_c is delimited between the lips and the cup of the mouthpiece. At the bottom of the cup a narrow channel of section S_c and length L_c forms a choke between the mouthpiece and the cylindrical section of the trombone.

The choke (tail pipe) acts as an acoustic mass, while the volume of the cup is a spring. The corresponding mass-spring resonance frequency is called the Helmholtz resonance frequency [7.11, 39, 52]

$$f_H \simeq \frac{c_0}{2\pi}\sqrt{\frac{S_c}{V_c L_c}}\,.$$

The radiated sound is enhanced in a so-called formant around this frequency, which is one of the parameters chosen by the player when selecting a mouthpiece. The amplitude of the acoustic impedance $|Z_p|$ of the instrument displays a series of peaks. One distinguishes the higher acoustic resonances that are close to a harmonic series $f_n \simeq nf_1$ $(n = 2, 3, 4, \ldots)$ from the lowest peak that typically has a frequency $f \simeq 0.7f_1$ (Fig. 7.14).

Fig. 7.13 **(a)** Trombone and **(b)** its mouthpiece (courtesy of Courtois Co.)

Hence when playing f_1 the musician uses a lip oscillation sustained by a series of higher harmonics rather than the fundamental. This particular note, called the *pedal note*, is difficult to play for a trumpet, but easy to play for a trombone.

As the trombone has a long cylindrical pipe, one would expect the impedance peaks to approach a series of uneven harmonics ($f = (2n-1)f_1$ with $n = 1, 2, 3, \ldots$). This was what occurred for the clarinet. The long cone followed by the large flaring bell are essential in making the series correspond to an almost full harmonic series ($f = nf_1$ with $n = 2, 3, 4, \ldots$) as we see in Fig. 7.14. Only the lowest impedance peak is not matching this harmonic series. In conical brass instruments, such as the French horn, the bell is not essential for the harmonicity of the resonance peaks. In some French horns the bell can be removed for transport convenience. Without the bell the instrument remains playable.

The frequency f_n is controlled by the length of the cylindrical section of the instrument. In order to play the full musical scale, the musician has to change this length. In a trombone this is done by using a slide. This allows a continuous change of the pitch providing the musician with a large flexibility, but limiting the

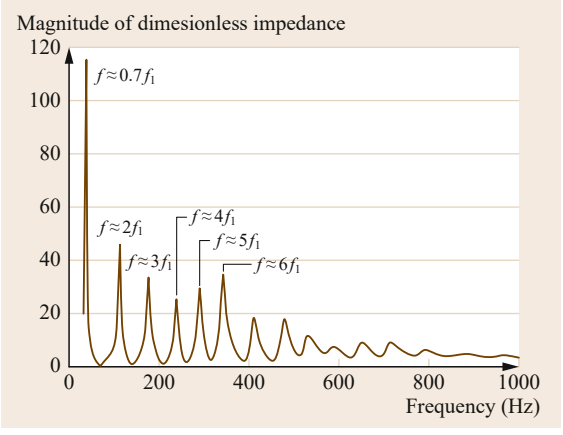


Fig. 7.14 Typical impedance of a trombone showing the series of harmonic resonances $f \simeq nf_1$ $(n = 2, 3, 4, \ldots)$ and the nonharmonic first peak $f \simeq 0.7f_1$. The lowest note f_1 is called the *pedal note*. In this example the Helmholtz resonance of the mouthpiece is around $f = 5f_1$

speed at which different notes can be played. In other instruments, such as the trumpet or the French horn, this problem is solved by using valves in order to modify the pipe length. In the French horn, the musician can place their hand within the bell to slightly affect the tuning, which is a common playing technique. Some baroque brass instruments (cornetto and serpent) have pipes with side holes [7.7].

When a musician plays a higher note for a given f_1, the pipe length can be much longer than the wavelength. A first consequence of this is that the player has to make their lips vibrate at the required pitch a few periods before they receive any acoustic feedback in order to sustain this oscillation. This is consistent with the observation that the reed (lip) frequency should be close to the pipe frequency $f_{\text{reed}}/f_{\text{pipe}} \simeq 1$ for brass instruments. The player can adjust f_{reed} by modifying the muscular tension (embouchure) in their lips. If the lips were simple opening (striking) reeds the acoustical feedback would be necessary in order to maintain an oscillation. The initial buzzing frequency of the lips would be strongly dependent on the configuration of the vocal tract, which would provide the acoustical feedback. The effect of the mouth impedance on trombone and trumpet playing is controversial [7.53, 54].

A more plausible explanation is that lips can display self-sustained (autonomous) oscillations due to the coupling of two mechanical modes of vibration with the flow [7.55–57]. This so-called flutter is the commonly accepted oscillation mechanism for the vocal folds in so-called voiced sound production. The energy transfer between the flow through the lip channel is the result of an asymmetry between the opening and closing phases of the lip movement. This is analogous to the asymmetry between the opening and closing phases of the swimmer's breast stroke, resulting in a net propulsion. During the opening phase, the lips form a converging flow channel that diverges abruptly at its exit. The neck of the channel is far downstream, just upstream of this abrupt exit. During the closing phase, the neck is more upstream. The lips form a gently diverging channel down to the abrupt exit. At the abrupt exit, flow separation occurs, forming a free jet in which the kinetic energy of the flow is dissipated with little pressure recovery. This dissipation provides the volume flow control by the opening of the reed as in any other reeds. The pressure in the jet is close to the pressure p_c in the cup of the mouthpiece. The volume flux $U_j S_j$ through the lips, with

$$U_j = \sqrt{2(p_m - p_{cup})/\rho_0}$$

the jet velocity in the cup and S_j the jet cross-section, is determined by the flow separation. Assuming a quasistatic incompressible one-dimensional flow between the lips, the local flow velocity $U = S_j U_j / S$ is inversely proportional to the local channel cross-sectional area S. The pressure p in a cross-section of the channel can then be deduced from the equation of Bernoulli $p = p_m - \rho_0 U^2/2$. Hence from the mouth to the flow separation point the pressure will be larger than the mouthpiece cup pressure $p \geq p_{cup}$ for the converging channel during the opening phase. Due to delayed flow separation in the gently diverging part of the flow channel, during the closing phase there will be a large region around the neck of the channel in which $p \leq p_{cup}$. Hence the lateral force due to the pressure in the flow will help opening the lips and help closing them. There is a net energy transfer from the flow to the lip oscillation, without need of an acoustical feedback. This process however involves the oscillation of two mechanical structural modes in order to obtain different geometries during the opening and closing phases of the oscillation. With rigid arms and legs (one degree of freedom) we could not swim (the breast stroke).

A consequence of the large length of the cylindrical section of a trombone is that, at high dynamical levels and especially at high pitches, the acoustic wave propagation is affected by nonlinear wave distortion effects. The high pressures occurring at high dynamical levels induce a local unsteady increase in speed of sound c in the compression part of the propagating pressure wave, corresponding to the adiabatic increase in temperature upon compression. The speed of sound in air is proportional to the square-root of the absolute temperature. For outgoing acoustic waves the increase in pressure also induces a particle velocity u' in the direction of the pressure wave propagation. Small perturbations propagate with the increased velocity $u' + c$ compared to the averaged speed of sound c_0. The expansion part of the wave will be slowed down by the same physical effects. As the expansion part of the wave is slow, it will tend to be overtaken by the fast compression part of the wave. This results into a steepening of the compression part of the wave. In contrast to water waves at the seashore where the top of the surface water wave overtakes the trough of the wave, the compression wave cannot actually overtake the expansion wave. This would correspond to a multiple-valued pressure distribution (three pressures in one point). The compression wave becomes eventually *infinitely* steep, which we call a shock wave. This occurs after a distance L_s, which we call the shock-wave formation distance. Neglecting the effect of reflection at the open pipe termination and friction along the pipe one finds

$$L_s \simeq \frac{2\gamma p_0 c_0}{(\gamma+1)\left(\frac{\partial p_p}{\partial t}\right)_{max}}, \tag{7.10}$$

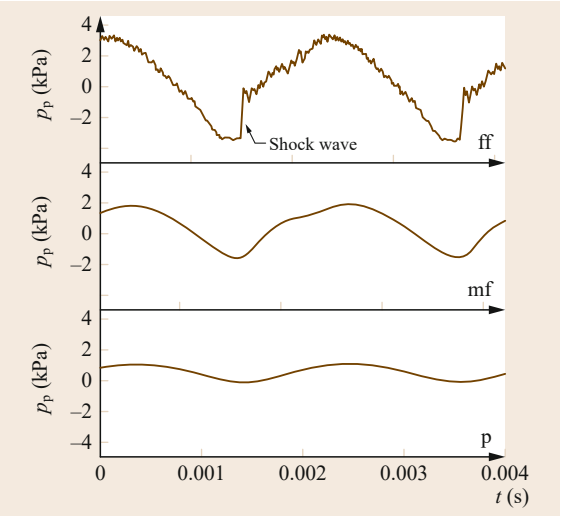


Fig. 7.15 Pressure oscillation at the end of the cylindrical section for a trombone played at various levels (*piano* p, *mezzoforte* mf, *fortissimo* ff). At *fortissimo* levels we observe a shock wave, which is a traveling pressure discontinuity (after [7.58])

where $\gamma = 1.4$ is the adiabatic exponent for air and p_0 is the atmospheric pressure. As demonstrated by experiments, shock waves can indeed be formed in brass instruments [7.58] (Fig. 7.15). When such a shock wave reaches the pipe exit, all the high frequencies in the wave are radiated away. The radiated pressure corresponds roughly to the time derivative of the incoming pressure signal in the pipe. The result is a periodic emission of sharp pressure peaks, corresponding to a spectrum with uniform harmonic amplitudes up to very high frequencies. This is recognized as brassy sound. Hence this is not related to the material of the instrument, but is sensitive to the length and shape of the bore, the geometry of the mouthpiece and the embouchure of the player. Note that before shock waves are reached the sound is already significantly affected by wave steepening with increasing dynamical levels [7.59]. In conical brass instruments this effect is reduced [7.60, 61]. One can therefore distinguish bright from mellow brass instruments depending on the bore shape.

It is interesting to note that earlier trombones have narrower bores than modern ones. Therefore in order to reach a certain dynamical level, i. e., radiate a certain acoustical power, one has to induce larger pressure fluctuations within the instrument in earlier trombones than in a modern one. This explains why these earlier trombones are brassier than modern instruments [7.60].

Note that due to heat transfer between the hot air blown by the musician and the walls of the instrument there is a temperature gradient along the pipe. Consequently the time-averaged speed of sound c_0 is not uniform in the pipe. This results in a modified effective acoustic length of the instrument. This is a linear effect and it is not related to the unsteady amplitude-dependent nonlinear variation in wave propagation speed $c + u'$, discussed above.

In contrast with other wind instruments, most brass instruments only radiate sound through the open pipe termination. This allows the use of a mute. The mute is usually designed to continue the horn shape, so that low frequencies taking care of the acoustical feedback are reflected as in the absence of mute. The mute does however reduce or modify the radiation. This is not only used for silent training but is extensively used for particular musical effects. The *wahwah* mute is an example. It is used for the trumpet in jazz performances.

7.6 The Flute

The modern transverse flute has a cylindrical pipe open at one side and is terminated by a mouthpiece at the other side (Fig. 7.16). The mouthpiece or head joint is closed at its end, but has a relatively large side opening called the blowhole or embouchure hole (called mouth in organ pipes). Sound is generated by blowing a thin grazing air jet across the blowhole towards the side called the labium. The labium of a modern transverse flute has a fairly acute angle (about 60°) and has a sharp edge. The labium of a recorder has an edge angle of about 20°. Rounding the tip of the labium strongly affects the sound quality and in extreme cases the playability of the instrument. In the transverse flute, the player blows in a direction normal to the pipe axis using their lips to form a slit from which the jet emerges. The player can modify the jet thickness (height), the direction of the jet and the length of the jet (distance from lips to edge of the labium). These parameters provide the player a control on the sound produced that is not available in the recorder or flue organ pipe. There the slit (flue) is incorporated in the instrument and has a fixed geometry. In the recorder and the flue organ pipe the jet flows are in the direction of the pipe axis.

In flue instruments, the acoustical oscillation in the pipe induces an oscillating transversal flow through the blowhole, which perturbs the jet at the flow separation point on the lips of the player (or flue exit in other in-

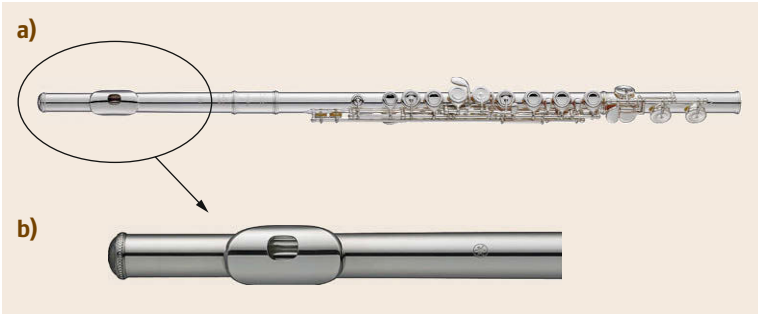


Fig. 7.16 **(a)** Boehm flute and **(b)** its head joint (mouthpiece) with the blowhole (courtesy of Yamaha Co.). Note that between the blowhole and the closed pipe termination of the head joint there is a cavity. The magnitude of this cavity can be adjusted with the cork. This cavity is essential for the harmonicity of the acoustic resonances of the flute

struments). At low amplitudes the jet perturbation can be described as a lateral displacement of the jet. The perturbation is convected by the flow at about half the center-line jet velocity $U_\mathrm{j} \simeq \sqrt{2p_\mathrm{m}/\rho_0}$, while it grows exponentially, in linear approximation. At the labium the jet is split into a volume flow Φ_in entering the pipe (below the labium) and a complementary volume flow Φ_out leaving the pipe (above the labium). For a total jet flow Φ_jet reaching the labium, one has $\Phi_\mathrm{jet} = \Phi_\mathrm{in} + \Phi_\mathrm{out}$. In terms of fluctuations, if we assume that Φ_jet is constant we have $\Phi'_\mathrm{in} = -\Phi'_\mathrm{out}$. Fluctuations are defined by $\Phi'_k = \Phi_k - \overline{\Phi}_k$ (k = in or out) where $\overline{\Phi}_k$ is the time-averaged flux.

Under normal playing conditions the travel time of jet perturbations across the blowhole is about a third of the oscillation period, hence this corresponds to a Strouhal number $\mathrm{Sr_W} = Wf/U_\mathrm{j} \simeq 0.15$, where W is the blowhole width, corresponding to the jet length. The transverse flute player adjusts the jet length W (by moving their lips or rotating the flute), the jet velocity U_j and the jet thickness in order to play in tune at different dynamic levels. In a recorder the jet length and jet thickness are fixed. Hence, the player has less control parameters. To reach a high pitch the recorder player has to blow much harder than for low notes.

At low blowing velocities, oscillations for $\mathrm{Sr_W} = O(1)$ can be observed, when the jet thickness h is sufficiently small compared to its length W ($W/h > 4$). In that case the jet flow breaks down into a vortex street; two rows of discrete vortices spinning in opposite direction (Fig. 7.17). We call these multiple vortex flow patterns ($\mathrm{Sr_W} \geq 1$) the higher-order hydrodynamic modes.

The sound generation mechanism in flue instruments has long been a subject of debate. *Helmholtz* [7.2] suggested that the volume flow entering the pipe Φ_in was acting as a monopole sound source. *Rayleigh* [7.3] argued that the sound source should be a fluctuating force because the sound source was located close to a pressure node of the acoustic standing wave, where a volume source would be rather ineffective. The work of *Powell* [7.62] and *Coltman* [7.63] has definitively established that the sound is produced by the reaction force $\boldsymbol{F}_\mathrm{h}$ of the walls to the unsteady hydrodynamic force due to the jet interaction with the walls. Using the so-called *aero-acoustical analogy* of Lighthill–Curle one can demonstrate mathematically from the mass conservation law and the momentum equation for a fluid that this force is a source of sound [7.9, 62]. In a lumped element model, it is most convenient to describe the sound source as a fluctuating discontinuity $\Delta p_\mathrm{source} = \boldsymbol{F}_\mathrm{h} \cdot \boldsymbol{n}/S_\mathrm{b}$ in the acoustical pressure across the blowhole of surface S_b with normal vector $\boldsymbol{n}$. At low Strouhal numbers the jet behavior can be described by a linear model: the so-called *jet drive*. In this model the jet flow is separated at the labium into a part entering the pipe and a part leaving the pipe. Acoustically this corresponds to two complementary volume sources $\Phi'_\mathrm{in} = -\Phi'_\mathrm{out}$ placed close to the edge of the labium and oscillating in opposite phase [7.64, 65]. In first-order approximation these sources form a so-called dipole. In terms of acoustic power we have

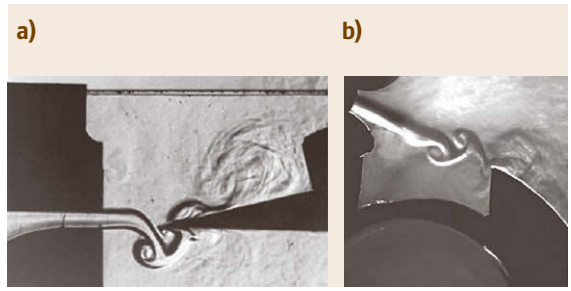


Fig. 7.17a,b Typical jet flow across the blowhole **(a)** of a recorder at $\mathrm{Sr_W} = 0.15$ and **(b)** of a transverse flute at $\mathrm{Sr_W} = 1$. Note the vortex shedding at the labium for low Strouhal numbers **(a)**. At the lower blowing velocities (higher Strouhal numbers) the jet breaks down into discrete vortices **(b)** before reaching the labium

$$\overline{P}_\mathrm{ac} = \frac{\omega}{2\pi} \int_0^{\frac{2\pi}{\omega}} \Delta p_\mathrm{source} u'_\mathrm{b} S_\mathrm{b} \mathrm{d}t \,, \tag{7.11}$$

where $u'_b S_b$ is the acoustic volume flux through the blowhole. Hence the source needs a large acoustical velocity to produce sound. This occurs at maxima of the amplitude (modulus) of the pipe admittance $1/|Z_p|$ [7.66]. When at higher Strouhal numbers the jet breaks down into discrete vortices (higher hydrodynamic modes), the so-called vortex-sound theory [7.67] provides an order of magnitude estimate for the sound source [7.65]. It would be however simplistic and misleading to think that sound is produced by the vortices. It is the interaction of the unsteady flow with the wall that produces sound. The motion of the vortices is associated with the hydrodynamic force on the wall. This vortex-sound theory allows the deduction of the force on the walls from an observation of the evolution of these vortices. The reaction force of the wall on the air is the source of sound. An accurate description of the sound production in flue instruments can only be reached by a solution of the equations describing the fluid motion (Navier–Stokes equations). Due to the nonlinearity of the equations, this is still a nontrivial numerical problem. Furthermore such a solution is actually a numerical experiment, which provides detailed information but little global insight. The jet-drive model and the discrete vortex model do provide some global understanding. Moreover, these caricatures can be used in lumped-element physical models for sound synthesis [7.68].

The amplitude of the transversal acoustic velocity u'_b through the blowhole reaches values as large as 25% of U_j [7.64, 65]. A consequence of this is that this transversal flow induces vortex shedding at the labium [7.69], which results in a nonlinear dissipation of the acoustic energy. This vortex shedding at the labium is a major amplitude limiting mechanism in flue instruments. It is interesting to note that when developing the modern flute Boehm increased the blowhole open area [7.16]. By doing this he increased the power. He thought that this was due to an enhanced radiation of sound. Actually this increased the jet flow velocity (at constant Sr_W) and reduced the nonlinear losses, for a given pressure amplitude in the pipe. The larger tone holes also improved the acoustics of the transverse flute [7.66].

An essential point is that the sound production in the flue instruments is due to the unsteadiness of the jet flow and primarily does not involve wall vibration. Consequently the material from which the walls are made is not critical and a golden flute is not necessarily better than a silver flute [7.70]. Of course the instrument maker can spend more time when building an expensive golden flute, so that such an instrument can be excellent.

While the sound production mechanism described above is common to all flue instruments, the acoustical response of the modern transverse flute is very peculiar and quite sophisticated. First of all, there is within the head joint a small cavity between the blowhole of the instrument and the end of the pipe closed by a *cork* (Fig. 7.16). This cavity is essential for the acoustics of the resonator. The depth of this cavity can be tuned to approach the response of a perfect open-open pipe with harmonic acoustical resonances $f_n = n\ f_1$ [7.7, 8, 39]. The cavity corrects for the fact that the pipe blowhole is narrow compared to the pipe cross-section ($S_b/S_p < 1$) inducing a locally higher acoustic velocity through the blowhole than in the pipe just upstream of the blowhole. This locally higher acoustic velocity induces an additional inertia of the acoustic flow through the blowhole. The fundamental frequency of the instrument corresponds to that of an open-open pipe with the actual pipe length L extended by a length correction δ_b: $f_1 = c_0/(L + \delta_b)$. The cavity provides the compliance corresponding to this additional pipe length. Because the cavity is closed (has a uniform pressure), while the virtual additional pipe segment is open (pressure increasing linearly with the distance from the pressure node at the open end), the cavity depth should be $\delta_b/2$ about half the virtual pipe segment length δ_b. Note that the pipe length L should also include another length correction δ_p to take into account the inertia of the acoustic flow around the open pipe termination. This so-called end correction is of the order of $\delta_p \simeq 0.7\sqrt{S_p/\pi}$. The exact value depends strongly on the ratio of pipe wall thickness to diameter and is strongly affected by the position of the player's lips. The closed chimneys of the tone holes also induce a difference between the effective acoustic length of the pipe and its actual length [7.6].

One of the subtleties of the Boehm flute is that the head joint is not cylindrical, as it was in baroque flutes. The head joint has a small convergence from 19 mm in the pipe towards 17 mm at the cork (closed-end). This is to make the correction for the harmonicity of the modes needed at higher pitches. For a stationary flow the pitch is constant, so that the sound source produces a signal consisting of pure harmonics of the fundamental frequency. We have a line spectrum. When the acoustic resonances of the pipe are also harmonically related, the sound is efficiently radiated. This makes the modern transverse flute sound rich in harmonics even though the labium is not as acute as in a recorder. In addition to this the musician can vary the distance between their lips and the labium by rotating the instrument around the pipe axis or modifying their lip's position. This allows for a correction of the pitch when changing the blowing pressure to achieve different dynamical levels. Hence the instrument can be played in tune for a much larger dynamical range than a recorder.

As the jet velocity is increased the jet becomes turbulent (Fig. 7.18) when the Reynolds number $\mathrm{Re}_h = U_j h/\nu$, with ν the kinematic viscosity of air, exceeds 10^3. Blowing with a turbulent jet against a sharp edge produces broad-band noise. Blowing with closed teeth produces such broad-band noise. This can be considered as musically interesting and is exploited in the Japanese flute (shakuhachi) and South American traditional flutes. The shakuhachi has a very sharp and acute labium, with edge angle of about 30°, which enhances the broad-band sound (noise) due to turbulence. In the Boehm flute the labium angle is about 60°, reducing the broad-band noise production. It is further interesting to note that turbulence tends to excite standing acoustic waves transversal to the pipe axis. The first of these transversal resonances has a *cutoff frequency* f_c corresponding roughly to a wavelength of twice the pipe width H. For square pipe cross-sections $f_c = 0.5c_0/H$ and $f_c \simeq 0.6c_0/D$ for cylindrical pipes of diameter D. For a piccolo flute $D \simeq 1$ cm corresponding to $f_c \simeq 20$ kHz, which is above the audio range. For a normal flute one finds $f_c \simeq 10$ kHz, which is high in the audio range. For a bass flute or a large organ pipe one finds $f_c \leq 5$ kHz, which is in the most sensitive frequency range of our hearing. Also the jet flow in large organ pipes will certainly be turbulent, while it remains laminar in a recorder. This partially explains the stronger broad-band sound in larger flue organ pipes compared to recorders [7.71].

Flue instruments have relatively long attack transients. While reed instruments, such as the clarinet and the oboe, reach a steady limit cycle within a few oscillation periods, the attack transient of a flue instrument can be several tens of oscillation periods $1/f_1$. This attack transient depends strongly on the initial perturbation in the instrument and the steepness of the increase in blowing pressure. When impulsively closing the holes in the pipe of an instrument, one can make the pipe oscillate. The instrument can be used as a percussion instrument, in the so-called *pizzicato* technique. For normal playing the initial oscillation obtained by abruptly closing the tone holes can significantly shorten the attack transient. In large (bass) recorders this technique is essential, because the instrument does not speak easily. When the pressure is increased without such an initial percussive drive, the instrument tends to start oscillating at a higher frequency than its fundamental pitch. This depends strongly on the ratio W/h of blowhole width W to jet thickness h. For $W/h \geq 10$ as in flue organ pipes one can observe an initial oscillation at about three times the pitch. This is because at small oscillation amplitudes (linear response) the low frequency amplification of perturbations convected by the jet is about a factor $\exp(2\pi) \simeq 530$ for each hydrodynamic wavelength $U_j/(2f)$ along the jet. Hence the third hydrodynamic jet mode is amplified about a factor 2.5×10^5 more than the first mode. This huge linear amplification will however soon induce nonlinear saturation by the formation of discrete vortices as shown in Fig. 7.17. Then the second hydrodynamic mode gets a chance to compete, and becomes dominant. Finally, the first mode slowly takes over, because it has lower radiation and viscous/thermal losses than the higher pipe modes. This complex attack transient can be reduced by using a short blowhole width $W/h \simeq 4$, such as in the recorder. For hydrodynamic wavelength shorter than about three times the jet thickness $fh/U_j > 0.2$ the jet does not amplify hydrodynamic perturbations. This implies that the attack transient of a recorder involves the second pipe mode (acoustic resonance), but not a third mode (Fig. 7.19).

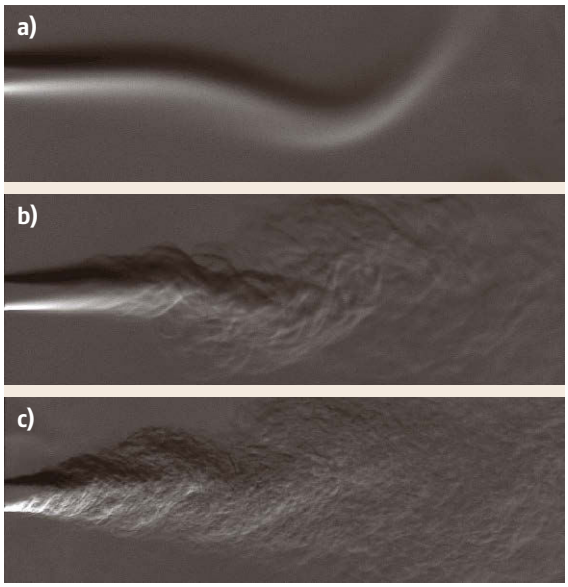


Fig. 7.18a–c Difference between laminar and turbulent jets: $\mathrm{Re}_h = hU_j/\nu = 200$ **(a)**, 500 **(b)**, and 3000 **(c)**. The jet is placed between two loudspeakers, inducing a transversal acoustic flow. This drives a jet oscillation as found in the blowhole of a flute

The higher hydrodynamic modes do not only appear during the initial phase of the attack transient. When blowing very softly they can produce sustained oscillations, which are called *whistle tones*. They are not only very weak, they are also less stable than the louder sounds obtained under normal playing conditions.

Under normal playing conditions, the acoustic oscillation of the pipe dominates the feedback loop, imposing a pitch close to a pipe resonance frequency. Flue instruments can however display, at low blowing pressures, oscillations dominated by local hydrody-

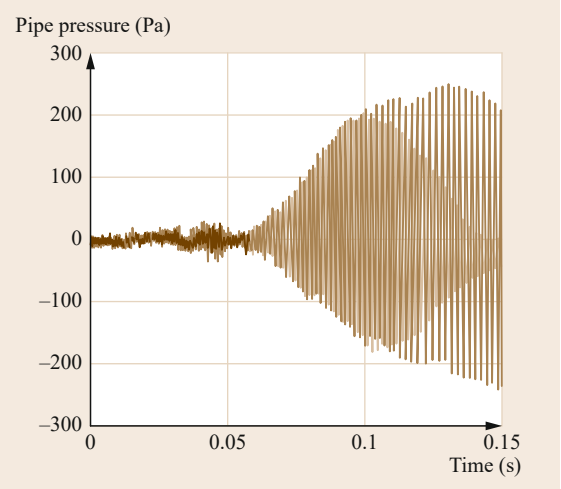


Fig. 7.19 Attack transient of a recorder (Ganassi flute of P. Bolton). Pressure in the pipe close to the labium. One clearly observes a dominance of the second acoustic pipe mode during the transient. In the initial phase of the transient even higher frequencies are shortly excited

namic feedback. This corresponds to so-called *edge tones* [7.62]. Edge tones can be recognized by the fact that the pitch increases linearly with the jet flow velocity. Such oscillations are called *mouth sounds* in organ pipes and are perceptively important [7.37, 38]. The perception of the sound quality is not only strongly affected by the attack transient, but also small variations in control parameters during the *limit cycle* oscillation strongly contribute. The flute player can vary the jet flow velocity U_j, the jet length W (distance between lips and labium edge), the jet thickness h and the offset y_{off} of the jet at the labium. By adjusting the Strouhal number $Sr_W = Wf/U_j$ the player can optimize the produced sound power or adjust the oscillation frequency f. The player can also vary the sound power by increasing h or U_j. When increasing the jet velocity U_j, the player has to change the jet length W in order to keep the pitch constant. This is commonly obtained by rotating the instrument around the pipe axis and/or moving the lips towards the labium. This also affects the passive acoustic response of the pipe, because it changes the end correction of the mouth of the pipe (the acoustic flow through the blowhole). The variation of h does not only affect the produced power, it also affects the attack transient. Indeed the ratio W/h determines the importance of higher hydrodynamic modes in the initial phase of the attack. It also influences the steepness of the attack. Also the offset y_{off} of the jet at the labium is crucial for the attack and the sound quality. When the jet is symmetrical with respect to the labium $y_{off} = 0$ the second harmonic in the sound is reduced. In the pipe, the amplitude of the second harmonic is typically reduced by an order of magnitude compared to the case $y_{off}/h = 0.4$, which is typical for a recorder. This will not only affect the sound quality, but also the pressure threshold for overblowing to a higher octave. The offset is not only determined by the flow direction at the outlet of the lip opening (flue). Due to the Coanda effect (pressure gradient induced by viscous flow entrainment), the jet tends to deflect towards the pipe wall. This effect is stronger for turbulent jet flow because turbulence enhances flow entrainment. This effect is so strong at the labium that a turbulent jet tends to *glue* to the labium, which reduces the offset at the labium. Hence upon laminar-to-turbulent transition one observes a drastic reduction of the second harmonic in the internal spectrum of the flute. In flue organ pipes this attachment of the turbulent jet to the labium is avoided by the use of very large mouth height ($W/h > 10$). A consequence of this is that higher hydrodynamic modes appear during the transients.

Small modulation of the sound, *vibrato*, during the limit cycle oscillation is often used to increase the liveliness of the sound. In extreme cases, this can be obtained by the player by singing into the flute (phonation).

This short discussion indicates that a virtuoso player has to control a lot of parameters to obtain a certain sound quality. This also makes modern transverse flute playing quite difficult. Indeed, this corresponds to an *expert* skill for which a typical intensive training period of ten years [7.72] is required before playing at professional level. As a result of this training, players produce highly repeatable and personal control on the instrument, as can be observed for instance from the dynamic shaping of the blowing pressure.

References

7.1 A. Baines: *Woodwind Instruments and Their History* (Dover, New York 1991)

7.2 H. Helmholtz: *On the Sensation of Tone* (Dover, New York 1954)

7.3 J.W. Strutt (Lord Rayleigh): The Theory of Sound (Dover, New York 1945)

7.4 H. Bouasse: Instruments a Vent (Librairie Delagrave, Paris 1929/30)

7.5 J. Backus: *The Acoustical Foundation of Music* (Norton, New York 1969)

7.6 C.J. Nederveen: *Acoustical Aspects of Woodwind Instruments* (Northern Illinois Univ. Press, DeKalb 1998)

7.7 M. Campbell, C. Greated: *The Musician's Guide to Acoustics* (Schirmer Book, New York 1987)
7.8 N.H. Fletcher, T. Rossing: *The Physics of Musical Instruments*, 2nd edn. (Springer, New York 1998)
7.9 A. Hirschberg, J. Kergomard, G. Weinreich: *Mechanics of Musical Instruments* (Springer, Wien 1995)
7.10 L. Henrique: *Acustica Musical*, 2nd edn. (Fundacao Calouste Gulbenkian, Lisboa 2007)
7.11 A. Chaigne, J. Kergomard: *Acoustics of Musical Instruments* (Springer, New York 2016)
7.12 B. Fabre, J. Gilbert, A. Hirschberg, X. Pelorson: Aeroacoustics of musical instruments, Ann. Rev. Fluid Mech. **44**, 1–25 (2012)
7.13 P. Taillard, J. Kergomard, F. Laloe: Iterated maps for clarinet-like systems, Nonlinear Dyn **62**, 253–271 (2010)
7.14 R. Bader: *Nonlinearities and Synchronization in Musical Acoustics and Music Psychology* (Springer, Berlin, Heidelberg 2013)
7.15 M. Campbell: Brass instruments as we know them today, Acta Acust. United Acust. **90**(4), 600–610 (2004)
7.16 T. Boehm: *The Flute and Flute-Playing* (Dover, New York 1964)
7.17 J.W. Coltman: Resonance and sounding frequencies of the flute, J. Acoust. Soc. Am. **40**, 99–107 (1966)
7.18 J.P. Dalmont, C.J. Nederveen, V. Dubos, S. Olivier, V. Méserette, E. te Sligte: Experimental determination of the equivalent circuit of an open side hole: Linear and non-linear behavior, Acta Acust. United Acust. **88**, 567–575 (2002)
7.19 J.P. Dalmont, J. Gilbert, S. Olivier: Non-linear characteristics of single reed instruments: Quasi-static volume flow and reed opening measurements, J. Acoust. Soc. Am. **114**, 2253–2262 (2003)
7.20 A. da Silva, G. Scavone, M. van Walstijn: Numerical simulations of fluid-structure interaction in single-reed mouthpieces, J. Acoust. Soc. Am. **122**, 1798–1810 (2007)
7.21 V. Lorenzoni, D. Ragni: Experimental investigation of the flow inside a saxophone mouthpiece by particle image velocimetry, J. Acoust. Soc. Am. **131**, 716–721 (2012)
7.22 M. Deverge, X. Pelorson, C. Vilain, P.Y. Lagrée, F. Chentouf, J. Willems, A. Hirschberg: Influence of collision on the flow through in-vitro rigid models of the vocal folds, J. Acoust. Soc. Am. **114**, 3354–3362 (2003)
7.23 P. Guillemain: Some roles of the vocal tract in clarinet breath attacks: Natural sounds analysis and model-based synthesis, J. Acoust. Soc. Am. **121**, 2396–2406 (2007)
7.24 G.P. Scavone, A. Lefebvre, A.R. da Silva: Measurement of vocal-tract influence during saxophone performance, J. Acoust. Soc. Am. **123**, 2391–2400 (2008)
7.25 J. Chen, J. Smith, J. Wolfe: Pitch bending and glissandi on the clarinet: Roles of the vocal tract and partial tone hole closure, J. Acoust. Soc. Am. **126**, 1511–1520 (2009)
7.26 J. Kergomard, X. Meynial: Systèmes micro-intervalles pour les instruments de musique à vent a trous lateraux, J. Acoust. **1**, 255–270 (1988)
7.27 J. Gilbert, J. Kergomard, E. Ngoya: Calculation of the steady-state oscillation of a clarinet using the harmonic balance technique, J. Acoust. Soc. Am. **86**, 35–41 (1989)
7.28 J. Kergomard, S. Olivier, J. Gilbert: Calculation of the spectrum of the self-sustained oscillators using a variable truncation method: Application to cylindrical reed instruments, Acustica **86**, 685–703 (2000)
7.29 M.E. McIntyre, R.T. Schumacher, J. Woodhouse: On the oscillations of musical instruments, J. Acoust. Soc. Am. **74**, 1325–1345 (1983)
7.30 E. Ducasse: Modélisation d'instruments de musique pour la synthèse sonore: Application aux instruments à vent, Sup. J. Phys. Colloq. Phys. **51-C2**, 837–840 (1990)
7.31 J.O. Smith III: Physical modeling synthesis update, Comput. Music J. **20**, 44–56 (1996)
7.32 V. Välimäki: *Discrete-time modeling of acoustic tubes using fractional delay filters*, Ph.D. Thesis (Helsinki University of Technology, Helsinki 1995)
7.33 C. Vergez, P. Tisserand: The BRASS project, from physical models to virtual musical instruments. In: *CMMR Third Int. Symp. Play. Issues (Computer Music Modelling and Retrivial)* (2005) pp. 1–10
7.34 P. Guillemain, J. Kergomard, T. Voinier: Real-time synthesis of wind instruments using nonlinear physical models, J. Acoust. Soc. Am. **105**, 444–455 (2005)
7.35 E. Mandaras, V. Gibiat, C. Besnainou, N. Grand: Caractérisation mécanique des anches simples d'instruments à vent, Suppl. J. Phys. III **4-C5**, 633–636 (1994)
7.36 T.D. Rossing, F.R. Moore, P.A. Wheeler: *The Science of Sound*, 3rd edn. (Person, Harlow 2001)
7.37 M. Castellengo: Acoustical analysis of initial transients in flute like instruments, Acta Acust. United Acust. **85**, 387–400 (1999)
7.38 A. Miklos, J. Angster: Properties of the sound of flue organ pipes, Acta Acust. United Acust. **86**, 611–622 (2000)
7.39 A.H. Benade: *Fundamentals of Musical Acoustics* (Oxford University Press, Oxford 1976)
7.40 E. Moers, J. Kergomard: On the cutoff frequency of clarinet-like instruments. Geometrical versus acoustical regularity, Acta Acust. United Acust. **97**, 984–996 (2011)
7.41 U. Ingard, H. Ising: Acoustic nonlinearity of an orifice, J. Acoust. Soc. Am. **42**, 6–17 (1967)
7.42 J. Buick, M. Atig, D. Skulina, M. Campbell, J.P. Dalmont, J. Gilbert: Investigation of Non-Linear Acoustic Losses at the Open End of a Tube, J. Acoust. Soc. Am. **129**, 1261–1272 (2011)
7.43 D. Noreland: An experimental study of temperature variations inside a clarinet. In: *In: Proc. Stockh. Music Acoust. Conf* (KTH, Stockholm 2013) pp. 446–450
7.44 T. Grothe: *Experimental Investigation of Bassoon Acoustics*, Ph.D. Thesis (Technische Universität Dresden, Dresden 2014)
7.45 A. Almeida, C. Vergez, R. Causse: Quasi-static nonlinear characteristics of double-reed instruments, J. Acoust. Soc. Am. **121**, 536–546 (2007)
7.46 N. Grand, J. Gilbert, F. Laloe: Oscillation threshold of woodwind instruments, Acustica **83**, 137–151 (1997)

7.47 J.P. Dalmont, B. Gazengel, J. Gilbert, J. Kergomard: Some aspects of tuning and clean intonation in reed instruments, Appl. Acoust. **46**, 19–60 (1995)

7.48 A.O. St Hilaire, T.A. Wilson, G.A. Beavers: Aerodynamic excitation of the harmonium reed, J. Fluid Mech. **49**, 803–816 (1971)

7.49 D. Ricot, R. Caussé, N. Misdrariis: Aerodynamic excitation and sound production of blown-closed free reeds without acoustic coupling: The example of the accordion reed, J. Acoust. Soc. Am. **117**, 826–841 (2005)

7.50 A.Z. Tarnopolsky, N.H. Fletcher, J.C.S. Lai: Oscillating reed valves: An experimental study, J. Acoust. Soc. Am. **108**, 400–406 (2000)

7.51 L. Millot, C. Baumann: A proposal for a minimalModel of free reeds, Acta Acust. United Acust. **93**, 122–144 (2007)

7.52 R. Causse, J. Kergomard, X. Lurton: Input impedance of brass musical instruments – Comparison between experiment and numerical models, J. Acoust. Soc. Am. **75**, 241–254 (1984)

7.53 J.M. Chen, J. Smith, J. Wolfe: Do trumpet players tune resonances of the vocal tract?, J. Acoust. Soc. Am. **131**, 722–727 (2012)

7.54 V. Freour, G.P. Scavone: Acoustical interaction between vibrating lips, downstream air column, and upstream airways in trombone performance, J. Acoust. Soc. Am. **134**, 3887–3898 (2013)

7.55 K. Ishizaka, M. Matsudaira: *Fluid Mechanical Considerations of Vocal Cord Vibration* (Speech Commun. Res. Lab., Santa Barbara 1972)

7.56 J. Cullen, J. Gilbert, M. Campbell: Brass instruments linear stability analysis and experiments with an artificial mouth, Acta Acust. United Acust. **86**, 704–724 (2000)

7.57 M. Newton, D.M. Campbell, J. Gilbert: Mechanical response measurements of real and artificial brass players lips, J. Acoust. Soc. Am. **123**, EL14–EL20 (2008)

7.58 A. Hirschberg, J. Gilbert, R. Msallam, A.P.J. Wijnands: Shock waves in trombones, J. Acoust. Soc. Am. **99**, 1754–1758 (1996)

7.59 J.W. Beauchamp: Analysis of simultaneous mouthpiece and output waveforms. In: *66th AES Conf., Los Angeles* (1980)

7.60 A. Myers, R.W. Pyle Jr., J. Gilbert, D.M. Campbell, J.P. Chick, S. Logie: Effects of nonlinear sound propagation on the characteritic timbres of brass insruments, J. Acoust. Soc. Am. **131**, 678–688 (2012)

7.61 L. Norman, J. Chick, D.M. Campbell, A. Myers, J. Gilbert: Player control of 'brassiness' at intermediate dynamic levels in brass instruments, Acta Acust. United Acust. **96**, 614–738 (2010)

7.62 A. Powell: On the edge tone, J. Acoust. Soc. Am. **33**, 395–409 (1961)

7.63 J.W. Coltman: Sound radiation from the mouth of an organ pipe, J. Acoust. Soc. Am. **46**, 477 (1969)

7.64 M.P. Verge, A. Hirschberg, R. Caussé: Sound production in recorderlike instruments. II A simulation model, J. Acoust. Soc. Am. **101**, 2925–2939 (1997)

7.65 S. Dequand, J.F.H. Willems, M. Leroux, R. Vullings, M. van Weert, C. Thieulot: Simplified models of flute instruments: Influence of mouth geometry on the sound source, J. Acoust. Soc. Am. **113**, 1724–1735 (2003)

7.66 J. Wolfe, J. Smith, J. Tann, N.H. Fletcher: Acoustic impedance spectra of classical and modern flutes, J. Sound Vib. **243**, 127–144 (2001)

7.67 M.S. Howe: Contributions to the theory of aerodynamic sound, with application to excess jet noise and the theory of the flute, J. Fluid Mech. **71**, 625–673 (1975)

7.68 R. Auvray, A. Emoult, B. Fabre, P.Y. Lagrée: Time-domain simulation of flute-like instruments: Comparison of jet-drive and discrete-vortex models, J. Acoust. Soc. Am. **136**, 389–400 (2014)

7.69 B. Fabre, A. Hirschberg, A.P.J. Wijnands: Vortex shedding in steady oscillation of a flue organ pipe, Acta Acust. United Acust. **82**, 863–877 (1996)

7.70 J.W. Coltman: Effect of material on flute tone quality, J. Acoust. Soc. Am. **49**, 520–523 (1971)

7.71 G. Paal, J. Angster, W. Garen, A. Miklos: A combined LDA and flow-vizualization on flue organ pipes, Exp. Fluids **40**, 825–835 (2006)

7.72 R. Chaffin, A. Lemieux: Musical excellence strategies and techniques to enhance performance. In: *General Perceptives on Achieving Musical Excellence*, ed. by A. Williamon (Oxford University Press, Oxford 2004) pp. 19–39

8. Properties of the Sound of Flue Organ Pipes

Judit Angster, András Miklós

This chapter is an overview of the characteristic sound properties of flue organ pipes. The characteristic properties of the stationary spectrum and attack transient have been surveyed and assigned to properties of the physical systems (air column as acoustic resonator, air jet as hydrodynamic oscillator, and pipe wall as mechanical resonator) involved in the sound generation process. The measurements presented underline the primary role of the acoustic resonator in the stationary sound and of the edge tone in the attack.

The pipe organ produces a majestic sound which differs from all other musical instruments. The richness and variety of timbre cannot be matched, because of the almost uncountable possibilities for mixing the sounds of different stops. The different stops of the organ also show a large variety in the richness of timbre; this is achieved by the proper selection of pipe shape, dimensions, wall material, shape and size of the opening(s) etc. The flue pipe ranks are divided into three groups according to their characteristic sound. The flutes produce a fundamental sound with fast attack, the diapason family can be characterized by their strong second partial especially in the attack, while the string family has a very bright sound with more than 20 harmonic partials, but with a slow attack.

Although the main features of the sound of flue organ pipes have been investigated extensively [8.1–6], the understanding of the connection between sound character and pipe shape and dimensions is still incomplete. Conversely, the sound character of the different ranks is unambiguously associated with pipe shape and measurements in the tradition of organ building [8.7–9]. Although the timbre, especially the attack, may be changed significantly by voicing adjustments, the main characteristics of the sound are quite stable for a given stop, and depend mostly on the form and scaling of the pipes. Nevertheless, the organ-building tradition cannot provide a proper explanation of the great variability of the sound character among the different stops on one hand and an explanation of the high stability of the sound character within one stop on the other hand. It is of interest to scientists that only a limited number of all the possible dimensions (diameter, wall thickness, cut-up height, flue width, etc.) and materials, which could be varied are actually used for organ pipes. Some of these limitations are technological in nature, but most of them have no basis in science. Consequently, several researchers have shared the view that the limitations of organ building need not be taken into account in experiments. This is the critical point of many of the scientific investigations carried out to date. Many scientific experiments published in the acoustic literature have been performed on models which differ considerably from real organ pipes. Consequently, the dimensions, shapes, and materials often did not fall within the narrow ranges used in organ building. According to the opinions of organ experts such experiments cannot model the operation of real organ pipes and, therefore, the results are useless for organ building. This view expressed by most organ builders, has frequently resulted in a rejection of attempts at mutual cooperation.

According to the opinion of the authors the right way lies somewhere in the middle. When basic phenomena are studied, experiments on models are ac-

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_8

ceptable. But the fine details of pipe sounds should be investigated on real organ pipes. Moreover, all physical effects contributing to the production of sound should be taken into account; an investigation limited to a specific parameter may lead to misinterpretation of the results. During the last 15 years the authors have carried out numerous experimental and theoretical investigations concerning the acoustic properties of flue organ pipes and the possible mechanisms of their sound generation. Although several questions are still unanswered, a better understanding of the processes involved and their interactions has been achieved.

8.1 Experimental Methodology

Although flue pipes offer a very wide variety of sound, the measured properties of these sounds contain several common elements which can be used to characterize them. In order to determine such characteristics, reproducible measurement methods have to be applied. The fact that open flue pipes radiate sound from both openings [8.10] should be taken into account when planning measurements. Since a sounding pipe produces an interference field even in an anechoic room, the measurement is extremely sensitive to the location of the microphone. The spectral envelope especially can change drastically with location, because changes in the distances to the openings mean different phase shifts for the different partials. Therefore, the measured spectrum may be more characteristic of the microphone location than of the properties of the pipe.

If the microphone is placed close to one opening of the pipe, the signal radiated from the other opening is negligible because of the much larger propagation distance. Although the magnitudes of the measured spectral components will decrease rapidly with the microphone-opening distance, their phases will remain almost constant, because the source-microphone distance is much smaller than their wavelengths. The sound pressure level of every spectral component will change by the same decibel value when the distance changes. Thus, the relative strengths of the spectral components will be quite insensitive to the location of the microphone. Therefore, the simplest method for obtaining reproducible spectral envelopes is the measurement in the vicinity of one of the openings.

The research was carried out mostly by laboratory experiments, although in situ measurements were also performed. A small slider chest was placed in the anechoic room; the wind was supplied by a centrifugal fan through a bellows. The fan and the bellows were located outside the anechoic room. A flexible tube with 80 mm diameter was used for connecting the bellows to the slider chest. The pallet in the slider chest was operated by an electromagnet, or it was manually fixed in an open position for measuring stationary sound.

For the characterization of the pipe sound usually three measurements were used: the stationary spectra at the mouth and the open end, and the attack transient at the mouth.

- *Stationary spectra* were measured and analyzed by a two-channel spectrum analyzer (HP35670A). In order to avoid interference the microphones usually were placed close ($\approx 3-5$ cm) to the openings of the pipes.
- *Attack transients* were directly converted to digital data and were evaluated using a special computer program [8.11] developed by the authors. The time evolution of the first 18 harmonic partials can be followed with this program. The results can be displayed in two groups (1–9th or 10–18th partials).

8.2 Steady-Sound Characteristics

In the last 15 years several hundred flue organ pipes have been investigated in our laboratory. The detailed evaluation and comparison of the measured spectra have revealed several features, which seem to be characteristic properties of the spectra of flue organ pipes.

The stationary spectra of a Diapason pipe can be seen in Fig. 8.1. These spectra show the most important properties of the sound of flue pipes. Characteristic features of the sound spectrum of a flue organ pipe can be listed as follows:

1. A series of harmonic partials
2. A second series of smaller and wider peaks, which are not harmonically related, but slightly stretched in frequency
3. A frequency-dependent base line

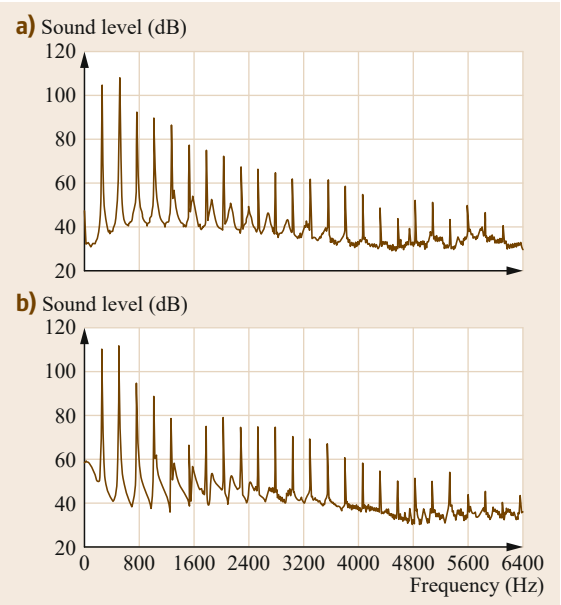


Fig. 8.1a,b Typical stationary spectrum of a flue organ pipe **(a)** at the open end, **(b)** at the labium

4. Envelope of the harmonic partials
5. Different spectral envelopes at the mouth and open end
6. Irregularities in the high-frequency part of the spectrum
7. Sharp, nonharmonic lines in the high-frequency part of the spectrum
8. Weak subharmonic components halfway between the partials.

Features 1–6 can be recognized in Fig. 8.1. Spectra containing features 7–8 will be shown later. In the following, the above features will be demonstrated with several measured spectra.

8.2.1 Physical Phenomena Related to the Observed Characteristic Features

The features of the spectrum are related to the physical properties of the flue pipe. The assumed physical reasons are summarized in Table 8.1. More detailed description of the phenomena can be found in the references given in the last column of Table 8.1. Short explanations for all items found in the second column of Table 8.1 follow:

1. It is well known from the elements of Fourier theory [8.12] that a periodic signal can be regarded as the convolution of only one period with an infinite series of Dirac-delta pulses separated by the period T. The Fourier transform (FT) of such a convolution will be the product of the FTs of the components. Thus, the FT of a periodic signal contains a series of harmonic partials (FT of the Dirac-delta pulse train) multiplied by an envelope determined by the FT of a single period. Therefore, any periodic signal with complex time waveform must have a lined spectrum with several partials and a complicated spectral envelope.
2. The small, broad peaks shown in the spectrum demonstrate the presence of acoustic eigenmodes of the pipe. This statement can be tested experimentally by using external acoustic excitation. Such demonstrations are regularly presented on the organ acoustic courses of the authors [8.13]. If a pipe is placed in the sound field generated by a loudspeaker, it will amplify the frequency components which correspond to the eigenresonances. Placing a small microphone in the pipe and using white noise excitation the eigenresonance spectrum can be determined. Such a spectrum is shown in Fig. 8.2 for a cylindrical tube. The 60 cm long 31 mm diameter tube was made of a lead-tin alloy (organ metal). The eigenresonances are slightly stretched; the frequencies are a bit higher than the harmonics of the first resonance. The stretching of the eigenfrequencies is much more pronounced in open organ pipes. In the spectrum of a Diapason pipe (Fig. 8.1a) the 9th eigenresonance lies about halfway between the 9th and 10th harmonic partials. The stretching becomes larger for larger diameter/length ratio and for smaller openings at the pipe ends. Conversely, conical pipes have smaller stretching, which decreases with increasing angle. At a certain angle the

Table 8.1 Characteristic features of the sound of flue organ pipes

Feature	Reasons	References
1	The sound generation is periodic	[8.12]
2	Eigenmodes of the acoustic resonator	[8.13]
3	Broadband noise produced by the jet	[8.14–16]
4	a) Losses of the acoustic resonator b) Radiation from the opening c) Excitation by the air jet d) Position of the eigenmodes	[8.17] [8.10, 18, 19] [8.20, 21] [8.13, 18]
5	Asymmetry of the standing wave in the pipe	[8.18]
6	a) Transversal resonances of the air column b) Wall resonances	[8.10] [8.22–25]
7	Simultaneous excitation of a higher eigenmode by the edge tone	[8.13]
8	Nonlinear coupling between air column and pipe wall	[8.23, 26]

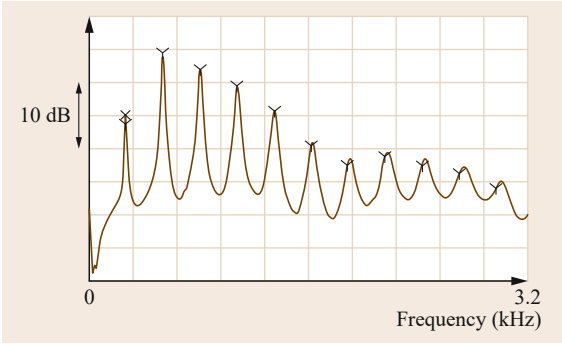


Fig. 8.2 Eigenresonances of a 60 cm long and 31 mm diameter tube. The harmonic partials are marked by *v-shaped cursors*

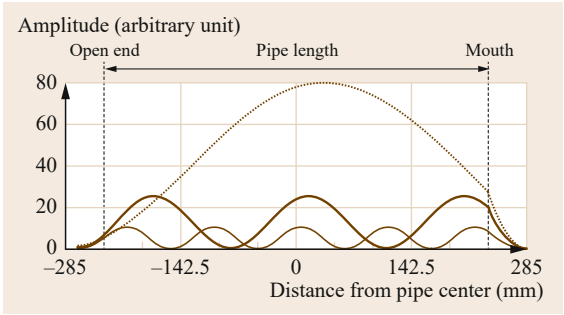


Fig. 8.3 Standing waves in an organ pipe. The sound pressure distributions of the 1st, 3rd, and 5th eigenmodes in a Nachthorn pipe

eigenresonances are harmonically related; above this angle they are even compressed.

These experimental facts can be understood by taking into account the physical properties of the organ pipe as an acoustic resonator. The air column in the pipe has several eigenmodes (standing wave patterns) with characteristic resonance frequencies (eigenfrequencies). Since the diameter is always much smaller than the length of the air column, the pure longitudinal eigenmodes appear first. Their frequencies are not harmonically related because of the *end correction* [8.10], which decreases with the frequency. As the *end correction* is proportional to the pipe diameter the stretching of the eigenfrequencies is larger for wide pipes than for narrow ones. Moreover, the *end correction* for a small opening (mouth) is larger than that of the larger open end. Therefore, the eigenfrequency stretching of an organ pipe is larger than that of a tube with the same length and diameter. Because of the different end corrections at the openings the standing wave is located asymmetrically inside the organ pipe [8.18]; it is shifted towards the smaller opening.

The measured asymmetry of the 1st, 3rd, and 5th eigenmodes in a Nachthorn (fairly wide) pipe are shown in Fig. 8.3. Besides the asymmetry it can also be observed that the phases of the measured sound pressure signals at the ends are fairly different from the multiples of π, because of the end corrections.

The pipe resonator can be unambiguously characterized by its eigenmodes. Each eigenmode is represented by three quantities: the resonance frequency, the quality (Q)-factor, and the position of the standing wave inside the pipe.

The pipe resonator, as all resonant physical systems, can collect and store energy in its eigenmodes. In case of a continuous excitation an equilibrium will be reached when the energy gain taken from the excitation compensates for the energy losses of the eigenmodes. The equilibrium ratio of the energy stored in an eigenmode to the energy lost in a period defines the Q-factor of the eigenmode. The numerical value of the Q-factor can be easily calculated by dividing the resonance frequency by the FWHM (full width at half maximum) value of the resonance curve. Since a resonator can absorb and emit energy only in small amounts, several periods are needed to reach equilibrium during the excitation or to lose the energy after the excitation ceased. The number of periods necessary to gain or lose 99.8% of the stored energy is given by 2Q. Therefore, the Q factor limits the speed of the attack significantly. The speech of a high-Q pipe must be slow, while a low-Q pipe can have a prompt onset. Conversely, the Q-factor gives the value of the peak amplification of the resonator. Thus, the sound pressure inside the pipe is usually much higher for a high-Q pipe than for a low-Q pipe.

3. The base line of the spectrum (Fig. 8.1) is determined by the broadband noise at the mouth of the pipe. This noise is produced at the flue and the upper labium [8.14, 27]. Quasirandom changes of the jet direction caused by turbulence in the foot [8.15] of the pipe also influences the noise. Moreover, tonal components (edge tone, see later) are also present in this background. Since the resonator amplifies this noise around the eigenresonances, the amplified noise may dominate the sound of the pipe in the high-frequency range, where the partials of the fundamental are usually weak. The high-frequency noise content can be very effectively reduced by nicking [8.16]. This method can increase the ratio of the harmonic partials to the base line very significantly.

4a. The form of the envelope depends on the total losses in the pipe; these consist of the sum of the volume losses in the air, the surface losses at the pipe

	Viscous loss	Radiation loss		Total loss	Quality factor Q	Spectrum
		Open end	Labium			
Narrow pipe (–12 HT)	9.7	0.8	4	14.5	70	
Normal pipe	5.7	2.4	12	20.1	50	
Broad pipe (+12 HT)	3.5	6.7	34	44.2	23	

Fig. 8.4 Calculated losses of open organ pipe resonators for the fundamental in ‰ ±12 HT: pipe diameter corresponds to the diameter of normal Diapason pipes having ± 12 semitone higher/lower pitch

wall due to viscosity and heat conduction, the radiation losses at the openings, and the energy loss due to the coupling of sound to wall vibrations. For organ pipes surface and radiation losses are much larger than the other two effects. The surface losses are theoretically [8.17] proportional to the factor $1/R\sqrt{f}$, while the radiation losses scale with R^2f^2, where R and f are the pipe radius and frequency, respectively. At the same frequency surface losses are relatively larger and radiation losses relatively smaller for narrow pipes than for wide pipes. Since surface losses decrease and radiation losses increase with the frequency a loss minimum occurs at a certain frequency. Indeed, such a loss minimum can be observed in narrow pipes, i. e., the largest amplitude occurs not for the fundamental, but for a higher partial. The theoretical frequency dependence of the envelope, which is inversely proportional to the total loss, is shown in Fig. 8.4 for narrow, normal, and wide organ pipe resonators. Measured spectra of a narrow and a wide pipe are shown in Fig. 8.5. In the case of the wide pipe resonator there are less higher partials, the amplitudes decrease faster with frequency and the stretching of the eigenmodes is bigger than by the narrow pipe resonator.

4b. Organ pipes radiate sound mostly at their openings. Although the pipe body can also radiate sound, this contribution can be neglected (see 6b in this list). Since the openings are much smaller than the wavelength of the sound, both of them can be regarded as simple sources (monopoles) with the source strength given by the acoustic flow at the openings [8.10, 18]. The envelope of the harmonics of the radiated spectrum shows a formant-like structure with a minimum (around the 6th partial in Fig. 8.1b). Although this phenomenon has been observed for sounding pipes, it can be explained by taking into account only the properties of the acoustic resonator. It is well known that the half-wavelength of the sound is longer than the length of the pipe (the *end correction* problem). Thus, the sound pressures at the open ends are nonzero. Assuming an end correction ΔL_m at the mouth the

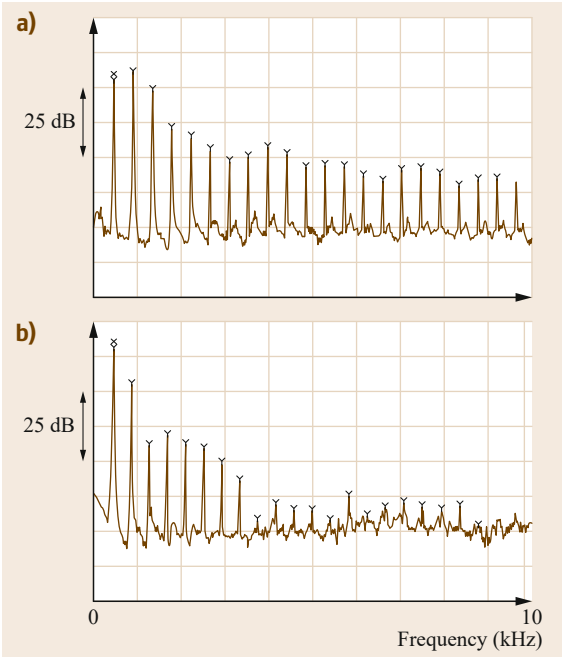


Fig. 8.5a,b Measured spectra of a **(a)** narrow (Salicional) and a **(b)** wide (Kornett) pipe

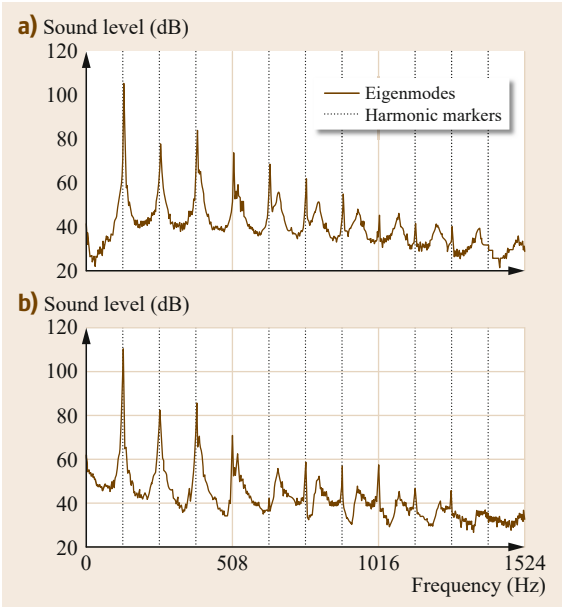


Fig. 8.6a,b Steady spectra of a Diapason pipe **(a)** at the open end, **(b)** at the labium

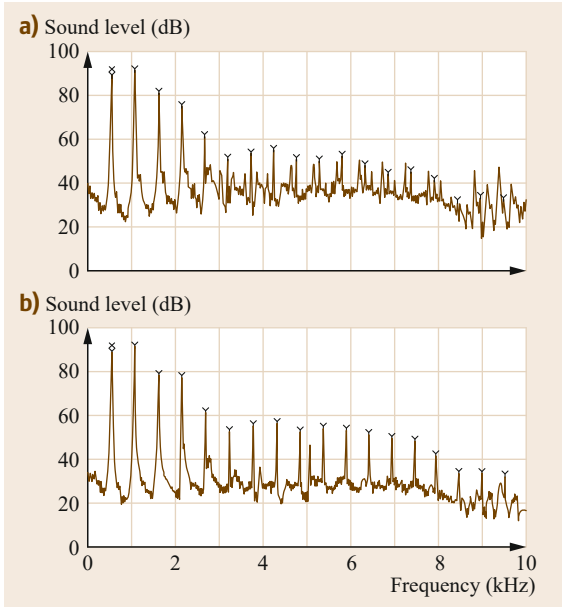


Fig. 8.7a,b Spectra of a 2 feet c (130.8 Hz) Diapason pipe **(a)** without nicks, **(b)** with nicks. The harmonic partials are marked by *v-shaped cursors*

sound pressure can be written as $p = P\sin\{k_1(x + \Delta L_m)\}$, where $k_1 = 2\pi/\lambda$ is the wavenumber. The spatial phase (the argument of the sine function) at the mouth ($x = 0$) then equals $\varphi_1 = k_1 \Delta L_m$. That is, the spatial phase of the standing pressure wave is larger than zero at the mouth [8.18]. The deviation φ_1 can be regarded as a phase correction for the mouth.

The spatial phase for the harmonic partials at the mouth can be determined simply by substituting the wavenumber of the fundamental k_1 with the corresponding wavenumber of the N-th partial, $k_N = Nk_1$. The spatial phase for the N-th partial at the mouth can then be written as $\varphi_N = N\varphi_1$. Since the acoustic flow depends on the cosine of the spatial phase [8.18], the sound pressure radiated from the mouth will have a minimum when $N\varphi 1$ approaches $\pi/2$. The first minimum of the envelope is determined by the diameter of the pipe and the size of the mouth opening. For normal Diapason pipes (mouth width $\approx 1/4$ circumference, mouth height $\approx 1/4$ mouth width) the first minimum lies in the range of the 5–6th partial (see Fig. 8.1b). For wide pipes the minimum may occur already around the 3rd partial (Fig. 8.5b), for narrow pipes it is shifted up to the 7–8th partials (Fig. 8.5a).

4c. As was mentioned in 1a above, the spectral envelope is determined by the spectrum of a single period of the sound. The shape of the temporal signal depends also on the excitation by the air jet. It was shown [8.20] that by changing the relative position of the jet and the upper lip the harmonic content of the sound can be influenced. When the jet moves symmetrically around the upper lip, no even harmonics are excited [8.10]. Probably such a symmetry effect is responsible for the weakness of the 2nd partial in Fig. 8.6. Nevertheless, this effect is not very important for real organ pipes, because the normal motion of the jet is quite asymmetric [8.21].

4d. The relative position of a harmonic partial and the neighboring eigenmode influences the amplitude of that partial. If the harmonic frequency is close to the eigenfrequency, the partial will be amplified by the eigenresonance. A harmonic partial lying midway between two eigenmodes will not be amplified. Therefore, the stretching of the eigenmodes also influences the envelope of the partials. In the case of small stretching (narrow pipes) several partials will be amplified, while for wide pipes the 3–4th partials are located in the valley between the eigenmodes. Nevertheless, as higher partials approach the peak of the one-index-lower eigenmode, they will also be amplified. This effect can be seen in Fig. 8.1b.

5. It has been shown [8.19] that the radiated acoustic field corresponds to that of two simple sources located at the openings of the pipe. The simple sources radiate in-phase for the odd partials and out-of-phase for the even ones. The strengths are

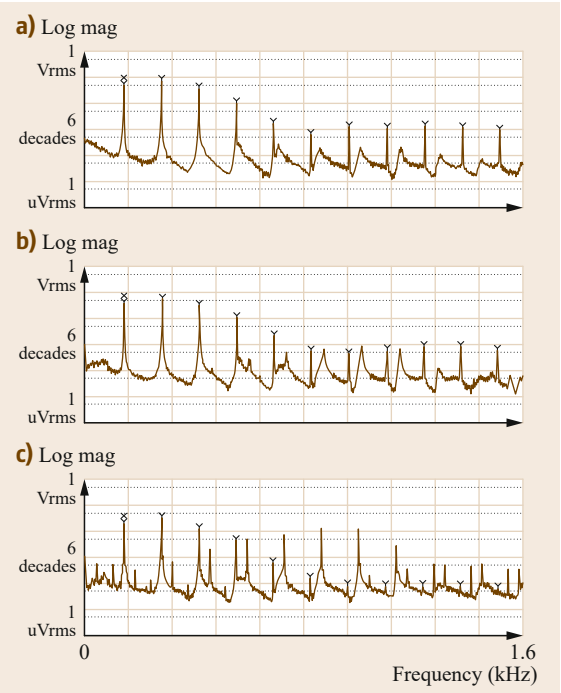


Fig. 8.8a–c Formation of the second sound of the Salicional pipe for decreasing wind pressure

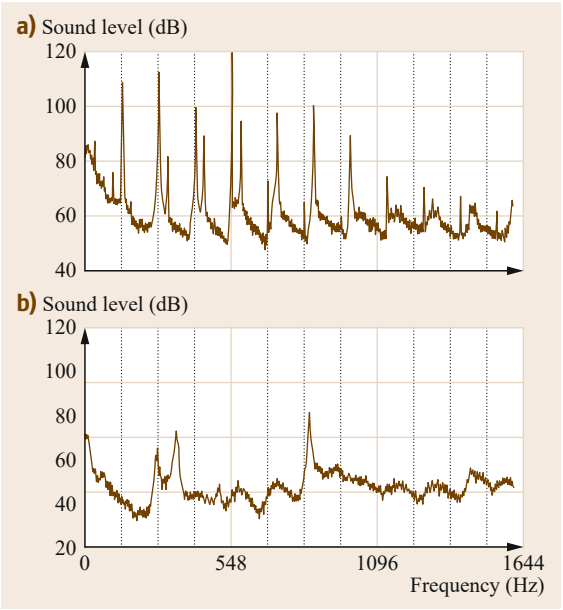


Fig. 8.9 (a) Sound spectrum of the Salicional pipe. **(b)** Mouth tone of the same pipe (the pipe sound was stopped by an absorbing termination at the open end)

different for both sources, and the two openings radiate different spectra [8.18]. The spectra of the sound radiated at the openings are different, because the standing waves in the pipe are asymmetrically located (see Fig. 8.3). Since the end correction is inversely proportional to the area of the opening [8.10, 18], the envelope minimum occurs for a smaller serial number of partial at the mouth than at the open end, in accordance with the experimental observations. That is, the spectral envelopes at the mouth and open end are always different. It may accidentally happen that the effective opening at the open end has about the same size as the mouth (for example when a tuning slot is applied). In such cases the symmetry may be restored, or the standing wave may even be shifted towards the tuning slot. Slight changes in the geometry of the openings can thus result in significant alterations of the radiated spectra. Another important parameter is the overlapping of the envelope minimum with a certain partial. It can be seen by comparing Figs. 8.1b and 8.6b that a very deep envelope minimum exists in Fig. 8.6b at the 5th partial (almost perfect overlapping), and only a slight decrease of the 6th partial in Fig. 8.1b (bad overlapping). Since the envelope minimum is shifted by certain voicing steps (for example by the change of the cutup), the voicer can influence the spectral envelope of the steady sound.

6a. Irregularities in the range of higher harmonics can be caused by the excitation of transverse resonances of the pipe. It is well known [8.10] that above the cut-off frequency of the pipe ($f > 200/d$, where d is the pipe diameter in meters) radial, azimuthal and mixed transverse modes exist. Certain pipe ranks may have harmonic partials well above the cut-off, therefore, transverse resonances can appear in the spectrum between the harmonic partials (Fig. 8.1, first transverse resonance ≈ 4.6 kHz). These resonances are excited by the high-frequency noise or edge tone at the upper lip.

6b. It has been shown [8.22, 23] that wall vibration cannot radiate sound directly. Conversely, a linear coupling exists [8.24] between the air column and the pipe wall for rectangular pipes, and also for cylindrical pipes if the pipe cross section is not a perfect circle, but slightly elliptical. In this case, wall vibrations can influence the sound radiated at the openings [8.23]. If a sharp vibration mode is close to an eigenmode or harmonic partial of the pipe sound, both modes will be coupled, which leads to a slight detuning of the corresponding sound component [8.25]. Nevertheless, such a coincidence is quite rare in practice.

7. It is very characteristic, especially for metal pipes in the $1'$ and $2'$ stops, that nonharmonic sharp

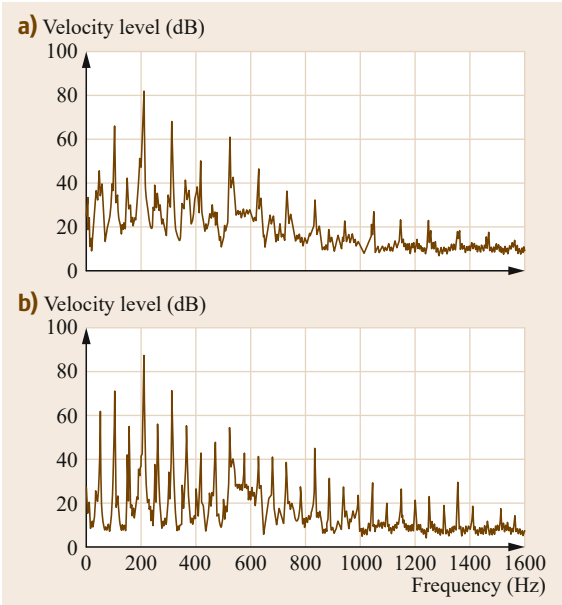


Fig. 8.10a,b Wall velocity spectrum of the Salicional pipe detected 5 mm under the wrapped plastic adhesive tape. **(a)** Pipe with tape. **(b)** Pipe without tape

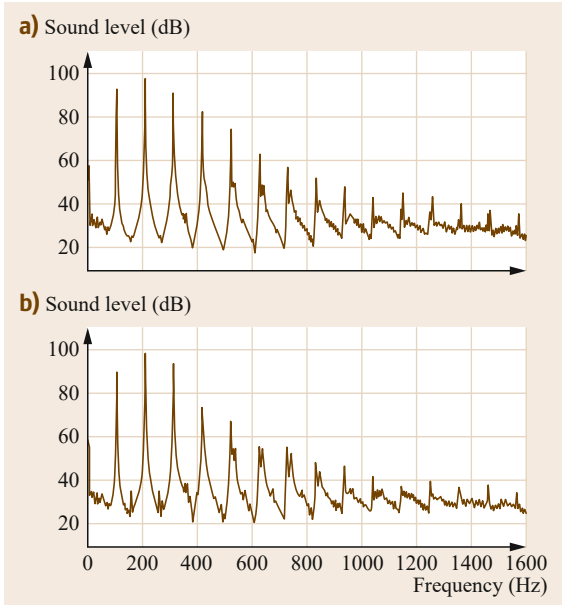


Fig. 8.11a,b Sound spectrum of the Salicional pipe detected 3 cm from the pipe mouth. **(a)** Pipe with tape. **(b)** Pipe without tape

peaks occur on the sides of high-frequency partials (Fig. 8.7a). These components give a roughness and a metallic character to the timbre. It was assumed [8.23] that these additional peaks can be assigned to wall vibrations, but recent investigations revealed that this phenomenon is probably caused by the excitation of higher eigenmodes by a higher component of the edge tone. This effect can be avoided by properly adjusting the edge tone, or by reducing the levels of the higher eigenmodes by nicking (Fig. 8.7b). Similar nonharmonic peaks were produced for example by reducing the wind pressure gradually. In this case the acoustic properties of the resonator do not change, while the edge tone frequencies shift downwards. It can be seen (Fig. 8.8) that certain eigenmodes (6th and 7th) start to grow. On further reducing the pressure the pipe produces two pitches, one corresponding to the fundamental and a second one corresponding to the virtual pitch of the second series of peaks. Stopping the pipe sound by an absorbing termination the so-called mouth tone (Chap. 3) shows a sharp and strong component around the 6th eigenmode (Fig. 8.9). This component may be responsible for the excitation of the second sound of the pipe.

8. Subharmonic peaks exactly halfway between the harmonic partials can be produced by strong nonlinear coupling between sound and wall vibration. The physical mechanism of such an interaction can be found in [8.26]. The phenomenon can be understood as follows: The pipe wall, if vibrating, undergoes an elliptical deformation of the originally circular cross section. As the area of an ellipse is always smaller than that of a circle with the same circumference, the cross-sectional area of the pipe will be modulated by a double frequency. The modulation of the cross section modulates the acoustic flow in the pipe, thus a coupling occurs between the sound and the wall vibration with about half of the sound frequency. Conversely, the oscillating sound pressure generates a periodic stiffening of the pipe wall; this effect can be regarded as a time-dependent elastic constant. The effect leads to a parametric instability of the wall. Because of this nonlinear coupling the wall and the air column can oscillate at half of the sound frequency. Such an effect was observed in the acoustic laboratory of the Fraunhofer-Institute in Stuttgart [8.23] with a very thin-walled pipe. This pipe is used in a church organ, but the nonlinear vibrations are damped by means of plastic adhesive tape wrapped around the pipe at about one third length from the mouth. Removing the tape causes the nonlinear vibration to increase tremendously. The comparison of Figs. 8.10a and 8.10b clearly show this effect. However, hardly any change can be seen in the corresponding sound spectra (Fig. 8.11), although some changes are evidenced in the levels of the

higher harmonics. Subharmonic components can be seen between the fundamental and the 2nd and between the 2nd and 3rd partials.

The explanations given above show that several different mechanisms influence the steady sound of the pipe. Nevertheless, even the most complicated features could be assigned to the measurable and understandable physical properties of flue organ pipes. Certainly, this is just a synopsis; more detailed analysis instead of the short descriptions given above will be published in the near future.

8.3 Edge and Mouth Tones

Since the edge tone seems to play an important role in the attack, the characteristic properties of the edge and mouth tones will be surveyed preceding the discussion of attack transients.

The air jet emerging from the flue and hitting the upper lip produces a very characteristic sound (jet sound), which is called *edge tone* in this chapter, because the edge tone is the dominant component of the jet sound. In order to study the edge tone without the influence of the acoustic resonator so-called *foot models* were used [8.28]. A foot model is essentially a flue organ pipe without acoustic resonator; it consists of a foot, a flue, and an upper lip only. The primary excitation of the organ pipes can be studied also by reducing the sound reflection (terminating the pipe with an absorber) from the other end below the threshold of oscillation. The sound heard in this case is called *mouth tone* and its role in the attack of the flue pipes was recently extensively investigated [8.29]. In the following the most important results of the research of the authors on edge and mouth tones will be surveyed.

8.3.1 Edge Tone of a Foot Model

The sound produced by the foot model has a very interesting character. It is noisy, the amplitude fluctuates, but it does have a definite pitch. The spectrum of the edge tone clearly shows the mentioned features (Fig. 8.12). It contains at least two or three peaks, which are not harmonically related, and a broadband base line, which decreases with the frequency. In the high-frequency range ($f > 10$ kHz) other very broad maximums may occur. The presence of the peaks shows that several hydrodynamical modes can coexist in the edge tone. Several theoretical and experimental works have been published about edge tones [8.30–33], but they cannot be used directly for organ pipes. Organ pipes utilize only a very limited range of wind pressure (usually 500–900 Pa), therefore the jet velocity is also limited. However, very detailed data are needed within this limited range. Therefore, extensive investigations have been carried out in the laboratory [8.15, 28, 34] in order to determine the main parameters of the edge tone and their dependence on the mouth dimensions and pressure. The main results can be summarized as follows:

- The jet emerging from the flue is not symmetric to the vertical plane containing the flue but bends slightly towards the lower lip. The angle of deviation from the vertical direction depends on the relative position of the lower lip and the edge of the languid.

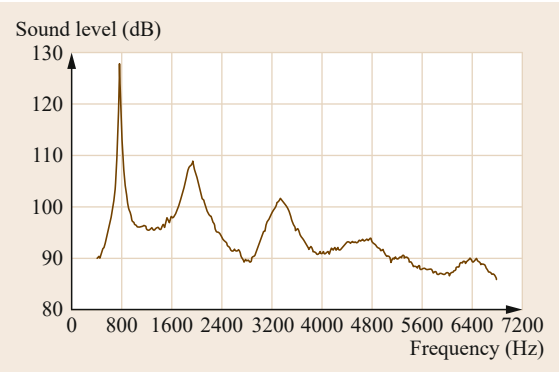


Fig. 8.12 Typical spectrum of a foot model (edge tone)

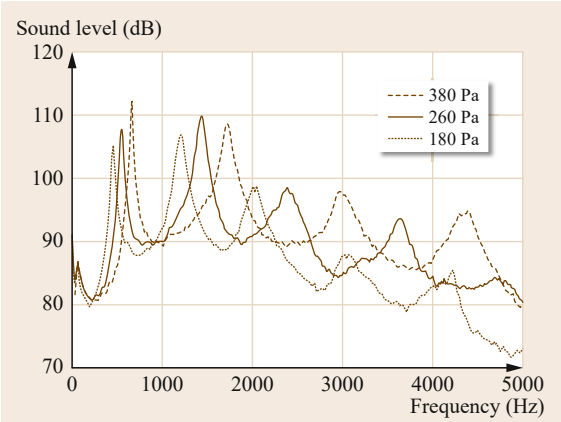


Fig. 8.13 Comparison of spectra of a foot model (edge tone) as a function of wind pressure

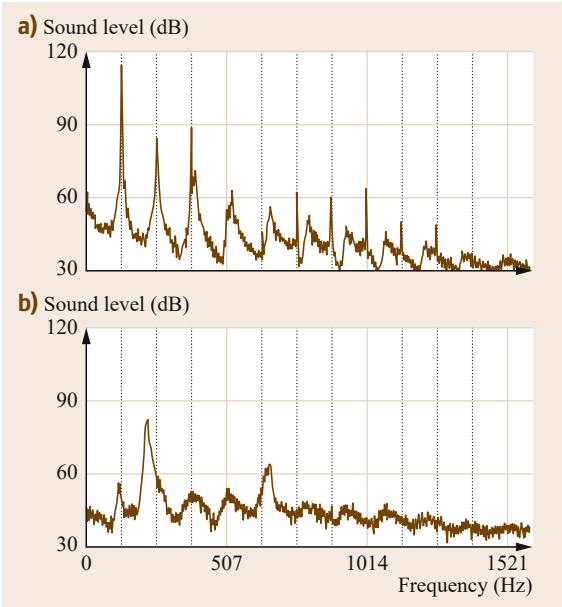


Fig. 8.14 **(a)** Sound spectrum of the Diapason pipe shown also in Fig. 8.4. **(b)** Mouth tone of the same pipe (the pipe sound was stopped by an absorbing termination at the open end)

- The jet broadens with the distance from the flue, but the angle is quite small (only 3–5°).
- The source of the edge tone is an acoustic dipole located slightly above the edge of the upper lip. The dipole character can be observed in the spatial distribution of the edge tone.
- The peaks of the edge tone spectrum move downwards in frequency when the jet velocity at the upper lip decreases (pressure decreases or the distance between the flue and upper lip increases) (Fig. 8.13).
- The peaks will be compressed with decreasing and stretched with increasing jet velocity (see Fig. 8.13). When moving downwards, more and more peaks appear, and the maximal amplitude shifts from the lowest peak to the second one.
- No jump or hysteresis of the edge tone frequency can be observed in the pressure range of the investigations.
- The build-up of the edge tone is very fast compared to that of the sound of a pipe having corresponding dimensions.
- The amplitude fluctuations are produced mostly by pressure fluctuations in the foot. Two stable flow states were observed in the foot [8.15, 28], and it is assumed that the flow randomly oscillates between them, producing quite large pressure changes at the flue. Indeed, the edge tone became more stable and the peaks in the spectrum became sharper when the flow was forced to remain in the selected flow state.

8.3.2 Mouth Tone of a Damped Pipe

When the jet hits the upper labium, an edge tone is generated. The rise time of the edge tone is very short, therefore it is already present at the mouth area well before the air column (the acoustic resonator) can enter into the game. As mentioned, the source of the edge tone is an acoustic dipole located slightly above the upper lip [8.14], thus the edge tone is immediately radiated outwards from the upper lip area. Since a dipole can be regarded as two simple sources with opposite phase, the same signal with opposite phase is radiated into the pipe simultaneously. This component, emphasized by the response of the resonator, is re-radiated by the mouth. Both components together form the mouth tone.

Mouth tones are also regularly studied in the laboratory. Their measurement is unfortunately quite difficult, because the effect of a termination of the open end with an absorber is twofold; not only will the amplitude of the reflected signal be reduced, but also the phase shift will be influenced. Reliable results can be obtained only if small phase shifts are produced by the absorber. From this point of view a low density rockwool absorber proved to be the best [8.34].

The mouth tone has a very similar spectrum to the edge tone. The only difference is that the broadband base line will be emphasized slightly by the (damped) eigenmodes of the pipe (Fig. 8.14). The subjective impression is also very similar; mouth and edge tones have the same character. The build-up time of the mouth tone is also very fast; it is comparable to that of the edge tone.

8.4 Characteristics of the Attack Transients

In the attack of the flue organ pipes three phases can be subjectively distinguished. These parts cannot be entirely separated in time, because they overlap quite broadly. Perhaps it is better to speak about three components, which start almost simultaneously, but with different rates of development. These three components can be characterized as follows:

1. *Forerunner*: This is the sound heard first. It is very difficult to describe. It may have a pitch, but sometimes no pitch can be assigned to it. Several different terms are used for this component (chiff, ping, hiss, cough, etc.). A detailed description of this phenomenon [8.29] was published recently.
2. *Appearance of a pitch*: The second component in the attack usually has a pitch close to the pitch of a higher harmonic partial. This component is very important for certain stops. For example, for several Diapason stops the second or the third harmonic can be heard preceding the fundamental, for Quintadena stops the third component is always prominent, while for narrow string stops this component often overlaps with the domain of higher (6–7th) harmonics.
3. *Onset of the fundamental*: The third parameter of the attack is the rise time of the fundamental. For stops of the flute family this rise time is very short, for stops of the string family it is very slow. As the fundamental grows certain components of the attack simultaneously become weaker.

The presence of the first two components is not compulsory in the attack. Moreover, the voicer can influence the attack very much, thus, according to his taste, he can produce faster or slower attack, more or less forerunner, brighter or more fundamental sound, etc. It is worth mentioning that sometimes one or more partials are quite strong in the beginning of the attack, but later they become weaker. Such behavior is often displayed by the fifth harmonic of the Diapason pipes.

The measurements show that the perception of the attack can be assigned to measurable properties. The attack transient of a Diapason pipe is shown in Fig. 8.15. This measurement was performed with the help of a computer program developed by the authors [8.11]. The sound was measured at the mouth of the pipe and the signal was converted to digital data by a 16-bit AD converter. A running window was shifted along the time recording, and a 256-point FFT was calculated at every step, then the levels of the harmonic partials were determined. About 250 FFT calculations were performed, then the levels of the first nine partials were displayed as functions of time. Since the frequency resolution is quite low, the method is not sensitive to minor changes in frequency. The three parts of the attack can be clearly distinguished in Fig. 8.15. The forerunner appears in every partial, implying its broadband nature (chiff). Then the partials start to grow; the fastest component is the 6th one. After a while the second will be the strongest; it dominates the attack in the 35–40 ms domain. The fundamental slowly overtakes the 2nd, which becomes slightly weaker as the fundamental rises.

The presented characteristics of the attack of flue pipes are very general, they can be found in almost all pipe sounds having been measured during the last 15 years. It can be assumed that these characteristics are related to the basic physical properties of the pipes. These relations will be investigated in the next section.

8.4.1 Physical Phenomena Related to the Observed Features of the Attack

The features of the attack can also be explained by the physical properties of the flue pipes. The reasons for the different features are summarized in Table 8.2. A more detailed description of the phenomena can be found in the references given in the last column of Table 8.2. A short explanation is given later for all items found in the second column of the Table.

1a. In order to understand the onset of the sound the events should be followed in the time domain. When the corresponding valve is opened (the key is pressed), the pressure in the pipe foot suddenly increases and, as the pressure difference drives the air through the flue, a jet is formed. As the jet

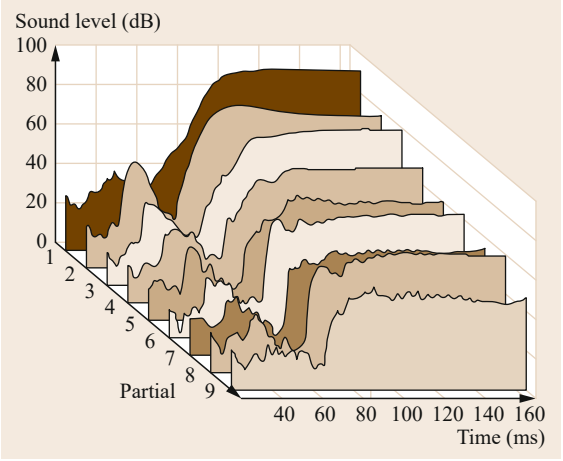


Fig. 8.15 Analyzed attack transient of a Diapason (311.1 Hz) pipe (Mühleisen)

Table 8.2 Characteristic features of the attack of flue organ pipes

Feature	Reasons	References
1	a) Jet noise b) Mouth tone	[8.14] [8.29, 34]
2	a) Excitation of a higher eigenmode by the mouth tone b) Mouth tone	This chapter [8.35, 36]
3	a) Locking of the jet to the fundamental b) Shift of the frequency to the corresponding harmonic partial	[8.37] This chapter

emerges from the flue, turbulent noise is produced by the flow. First, a pressure pulse is produced, then broadband noise appears due to the turbulence of the jet [8.14]. Both sound components are broadband, no pitch can be assigned to them. This sound is radiated directly from the mouth area without any contribution from the acoustic resonator.

1b. When the jet hits the upper labium, an edge tone, which will be transformed within a very short time to a mouth tone, is generated. Since the jet velocity increases within the first ≈ 10 ms of the attack from zero to the final value [8.37], the mouth tone cannot develop a certain pitch, its frequency and spectrum change continuously according to the jet velocity change. This sliding signal produces the special hearing sensation in the forerunner. If the rise time of the jet velocity is very fast compared to the period of the edge tone, only a short sound pulse can be observed.

2a. The eigenresonances of the acoustic resonator need about ten times longer than the nearby mouth tone component for the sound to build up. Since the effect of the weak sound components on the transversal motion of the jet is small at the beginning, the already developed mouth tone provides a larger contribution to the excitation than the still weak feedback of the acoustic resonator. Therefore, the tonal components of the mouth tone will excite the nearby eigenmodes of the air column. Since the mouth tone has several spectral components more eigenmodes could be excited simultaneously. Among these eigenmodes the highest will grow the fastest, because the rise time is smaller for higher eigenmodes than for the lower ones. This effect can be clearly observed in Fig. 8.15. It can be seen that the first excitation occurred around the 5–6th eigenmode, followed by the excitation of the 2nd eigenmode by the strongest component of the mouth tone. A well-trained ear can hear these components as a fast changing pitch in the speech of the pipe.

2b. For narrow pipes the onset of the fundamental is usually very slow. Since the higher eigenresonances have large Q-values, these higher components also need a long time to build up. Therefore, the mouth tone can be heard (and observed) between the forerunner and the onset of the sound of the resonator. A good voicer can adjust the edge tone so that a tone reminiscent of a harmonic component of the steady sound can be heard. Such voicing produces very pleasant pipe speech.

3a. The fundamental tone starts to grow also, because the reflections from the ends of the pipe amplify this component most effectively. As the signal grows in the resonator, the sound pressure and the acoustic flow at the mouth become so large that they can influence the movement of the jet. After a while the transverse movement of the jet will be locked to the fundamental [8.29], and a self-sustaining oscillation will be developed [8.10, 14]. The time needed for achieving the fully developed steady state depends mostly on the Q-factor of the fundamental eigenmode of the pipe. Since the Q is small for wide pipes (flute ranks), medium for Diapasons and high for the string stops, the duration of the attack scales accordingly (prompt attack for flutes, slow speech for the string family).

3b. The frequencies of the higher components of the sound and the pressure distribution in the pipe change during the stabilization process until the final steady state is achieved. During this process the higher components shift from the eigenmodes to the partials of the fundamental. These changes could be quite large, because the eigenmodes have their own spatial distribution which depends only on the boundary conditions at the openings (see Fig. 8.6), while the spatial distribution of the partials depends only on the fundamental standing wave. The change of the spatial phase at the openings influences the sound radiation, therefore the magnitude of the partial closest to the envelope minimum of the stationary spectrum will decrease to the steady level. In the case of Diapason pipes the minimum usually lies around the 5–6th partial [8.18], thus these components may decrease significantly during the third phase of the attack. This process may last quite long; in the middle frequency range ($\approx$ 130–260 Hz, the tenor octave) 100–120 ms stabilization times were observed.

Since the voicing adjustment can change the attack tremendously [8.16, 35, 36], the assignment of the features of the attack to measurable and understandable physical properties of the flue organ pipes is much more complicated than in the case of steady sound. A more detailed description of the research, summarized in this chapter, will be published in the near future.

8.4.2 Acoustic Effects of the Voicing Adjustment Steps

As mentioned, the voicing adjustment can influence both the steady spectrum and the attack transient very significantly. The voicing steps influence mostly the velocity of the jet at the upper labium and the volume flow through the flue. Certain steps change the end correction at the mouth and open end. Nicking (making thin cuts into the edge of the languid) decreases the broadband noise very effectively.

Since only a few papers have been published [8.16, 20] about the effect of voicing adjustments a short summary about the acoustic effect of the most important voicing steps is given as follows:

1. *Change of the diameter of the foothole/of the width of the flue*: The pressure in the foot is determined not only by the applied wind pressure but also by the ratio of the open areas at the foothole and the flue. Both the volume flow and the jet velocity at the flue can be influenced by these voicing steps. They are used mostly for adjusting the loudness of the steady sound and the speed of the attack.
2. *Knocking the languid up and down*: This step influences the direction of the jet. If the languid is knocked up/down, the jet moves outwards/inwards. Accordingly, the velocity of the jet when hitting the upper labium and amount of the air flowing into the pipe decreases/increases. Thus, this voicing step can influence the loudness of the steady sound, the speed of the attack and the steady spectrum (because of the changing asymmetry of jet motion at the upper labium).
3. *Cutting up the mouth*: This step changes the end correction at the mouth. The frequency increases and the envelope minimum of the steady spectrum is shifted upwards. Conversely, the velocity of the jet at the upper lip and the amount of air flowing into the pipe decreases. Thus, the edge tone spectrum and the pipe resonances are shifted in opposite direction. This voicing step influences the steady spectrum and modifies the order of the appearance of the partials in the attack. Therefore, this step is the most important one in adjusting the sound character of the pipe.
4. *Nicking*: The small cuts in the edge of the languid produce vertical vortex lines, which stabilize the jet and reduce the turbulence. The broadband noise in the steady spectrum decreases in the range of the higher partials. Conversely, the attack will be somewhat slower.
5. *Tuning at the open end*: Depending on the applied tuning method either the length of the pipe is changed, or the end correction is modified by this step. When a small opening (tuning slot) is used, the tuning may influence (because of the change of the location of the standing wave in the pipe) the steady spectra at the open end and also at the mouth.

Several other voicing steps are known, but the above-mentioned steps are the most important ones. Steps 1–3 obviously influence the frequency, the spectrum, and the amplitude of the edge tone.

8.5 Discussion and Outlook

Since the scope of this chapter is to give an overview of the connection between the measurable properties of the sound and the physical properties of the systems involved in the sound generation (acoustic resonator, air jet, wall vibrations), no detailed discussion has been given of particular phenomena; rather the process of sound generation in flue pipes as a whole was discussed. Nevertheless, certain facts, which are not well known or which contradict published results, have emerged.

The sound generation may be regarded as a complicated physical interaction among all of the involved systems, aerodynamic (air jet), acoustical (air column in the pipe), and mechanical (pipe wall). The result of this interaction is the sound of the pipe. This approach was to select, demonstrate, and explain the most important connections between the sound and the above three physical systems. Since more systems are usually involved and since several properties may contribute to a single observable quantity, this approach inevitably leads to an oversimplification of the problem. Still, it is believed that the approach presented in this chapter can help the reader to reach a better understanding of the sound of flue organ pipes.

Although the main features of the steady state sound of different pipes and stops are quite well known, the difference between the radiated spectra at the openings

was completely overlooked in the literature. As was shown, the size (and shape) of the opening influences not only the eigenfrequencies and the Q-factors of the pipe, but also determines the stretching of the eigenmodes and the shape of the spectrum envelope. The latter two properties are much more important for the steady state spectrum than the first two.

The appearance of nonharmonic peaks in the range of higher partials is a very characteristic property of metal pipes. Still, this phenomenon was not discussed in the scientific literature. The first information on this phenomenon was published recently [8.29], but it was regarded as a special effect of a single pipe. Moreover, the appearance of the peaks was described as difference tones of the mouth tone and fundamental. The explanation based on the simultaneous excitation of a higher eigenmode assumes that the higher components of the edge (mouth) tone are present in the steady sound. The presence of the mouth tone in the steady sound was indirectly proved in another experiment [8.38]. Applying a telescopic extension on the pipe the nominal length (198 cm) was changed in several (48) 1 cm steps in the 174–222 cm range. The driving pressure and the mouth parameters were not changed during the experiment. In this case only the acoustic resonator changes, the jet remains the same. Accepting the regenerative excitation mechanism [8.10, 27] used for describing the sound generation in flue pipes, the pipe should produce a sound with smoothly decreasing frequency. The amplitude and the spectrum should also change smoothly. The measurements have shown, that the frequency changed smoothly indeed, but in a certain range the sound became very unstable, a beating occurred, sometimes a second sound could be heard for a very short time. Leaving this domain the sound became stable again. The attack showed similar phenomena: it became very slow in the above-mentioned domain and strong beats were observed.

Another indirect proof of the presence of the mouth tone in the steady sound was presented recently [8.39]. The harmonic partials were digitally removed from the pipe sound and the remaining sound was used for listening tests. The character of the remaining sound was very similar to the mouth tone of a damped pipe.

The short explanations given for the observed properties of the attack show that several different mechanisms are involved in the attack transient of the pipe. Certain phenomena can be explained by the properties of the acoustic resonator i. e., the observed differences in the rise time of the partials are mostly determined by the differences of the Q-factors, while the gradual withdrawal of certain partials can be assigned to the shift of their frequencies from the corresponding eigenfrequency to the harmonic frequency. It is also clear, that the very beginning of the speech is dominated by the noise and edge tone produced by the jet. However, the role of the edge tone in the excitation of the pipe sound is not understood completely yet.

The effect of wall vibrations on the attack was not taken into account, because this phenomenon is extremely complicated, but plays only a secondary role in the attack. It was shown [8.13] that pipes made from very different material can be voiced to almost identical speech.

The presented observations underline that the role of the edge tone in sound generation is not yet properly understood. Detailed investigations of edge tone–pipe sound interaction will be carried out in the near future.

Acknowledgments. The contents of this chapter have in large parts been published before [8.40].

The research surveyed in this paper was supported by several foundations in Hungary (Soros Foundation, Foundation of the Hungarian Credit Bank and István Széchenyi Foundation) and in Germany (Deutscher Akademischer Ausländerdienst (DAAD), Katholischer Akademischer Ausländerdienst (KAAD)). The research could not have progressed without the generous help of organ builders and experts, including the participants of the short courses. The authors are especially grateful for the friendly support of K. Mühleisen (Organ Builder Company Mühleisen, Leonberg, Germany) and F. Frasch (Technical College for Building of Musical Instruments, Ludwigsburg, Germany)

References

8.1 A.W. Nolle: Sinuous instability of a planar air jet: Propagation parameters and acoustic excitation, J. Acoust. Soc. Am. **103**, 3690–3705 (1998)
8.2 S.A. Elder: The mechanism of sound production in organ pipes and cavity resonators, J. Acoust. Soc. Jpn. (E) **13**, 11–23 (1992)
8.3 J.W. Coltmann: Jet drive mechanisms in edge tones and organ pipes, J. Acoust. Soc. Am. **60**, 724–733 (1976)
8.4 N.H. Fletcher: Sound production by organ flue pipes, J. Acoust. Soc. Am. **60**, 926–936 (1976)
8.5 S. Yoshikawa, J. Saneyoshi: Feedback excitation mechanism in organ pipes, J. Acoust. Soc. Jpn. (E) **1**, 175–191 (1980)
8.6 M.P. Verge, B. Fabre, W.E. Mahu, A. Hirschberg: Feedback excitation mechanism in organ pipes, J. Acoust.

Soc. Am. **95**, 1119–1132 (1994)
8.7 C. Mahrenholz: *Berechnung der Mensuren* (Orgelbau-Fachverlag Rensch, Lauffen/Neckar 1987) pp. 35–125
8.8 W. Ellerhorst: *Handbuch der Orgelkunde* (Benzinger, Einsiedeln 1936) pp. 17–19
8.9 M.A. Cavaillé-Coll: *Sämtliche theoretischen Arbeiten* (Jochum, Dornbirn 1982) pp. 126–137
8.10 N.H. Fletcher, T.D. Rossing: *The Physics of Musical Instruments* (Springer, New York 1991)
8.11 J. Angster, A. Miklós: Documentation of the sound of a historical pipe organ, Appl. Acoust. **46**, 61–82 (1995)
8.12 G.A. Korn, T.M. Korn: *Mathematical Handbook for Scientists and Engineers* (McGraw-Hill, New York 1975)
8.13 J. Angster, A. Miklós: Intensive courses of organ and church acoustics. Organised at the Fraunhofer Institute of Building Physics in Stuttgart, Germany, https://www.ibp.fraunhofer.de/en/Expertise/Acoustics/Musical-Acoustics.html
8.14 B. Fabre, A. Hirschberg, A.P.J. Wijnands: Vortex shedding in steady oscillation of a flue organ pipe, Acust.-Acta Acust. **82**, 863–877 (1996)
8.15 S. Pitsch, J. Angster, M. Strunz, A. Miklós: *Spectral properties of the edge tone of a flue organ pipe, ISMA '97, Edinburgh* 1997) pp. 339–344
8.16 J. Angster, G. Paál, W. Garen, A. Miklós: Effect of voicing steps on the stationary spectrum and attack transient of a flue organ pipe. In: *ISMA '97, Edinburgh* (1997) pp. 285–294
8.17 P.M. Morse, K.U. Ingard: *Theoretical Acoustics* (McGraw-Hill, New York 1968)
8.18 J. Angster, A. Miklós: Sound radiation of open labial organ pipes; the effect of the size of the openings on the formant structure. In: *Int. Symp. on Music. Acoust. (ISMA '98, Leavenworth), Acoust. Soc. Amer and Catgut Ac. Soc., Leavenworth* (1998) pp. 267–272
8.19 J. Kümmel: *Raumakustische Probleme bei der Aufstellung von Orgelpfeifen*, Diplomarbeit (Universität Stuttgart, Stuttgart 1994)
8.20 A.W. Nolle: Theoretical Acoustics. Flue organ pipes: Adjustments affecting steady waveform, J. Acoust. Soc. Am. **73**, 1821–1832 (1983)
8.21 G. Paál, J. Angster, W. Garen, A. Miklós: Sound and flow in the mouth of flue organ pipes. Part I: Fully developed state. In: *ISMA '97, Edinburgh* (1997) pp. 295–301
8.22 J. Backus, T.C. Hundley: Wall vibrations in flue organ pipes and their effect on tone, J. Acoust. Soc. Am. **39**, 936–945 (1965)
8.23 J. Angster, G. Paál, W. Garen, A. Miklós: The effect of wall vibrations on the timbre of organ pipes. In: *16th Int. Congr. Acoust. and 135th Meet. Acoust. Soc. Amer., Seattle*, Vol. 2 (1998) pp. 753–754
8.24 J. Angster, Z. Dubovski, S. Pitsch, A. Miklós: Impact of the material on the sound of flue organ pipes (acoustic and vibration investigations with modern measuring techniques). In: *Analysis and Description of Music Instruments Using Engineering Methods*, ed. by C. Birnbaum (Stiftung Händel-Haus, Halle (Saale) 2011) pp. 34–41
8.25 J. Angster, I. Bork, A. Miklós, K. Wogram: The investigation of the vibrations of an open cylindrical organ flue pipe. In: *9th FASE Symp. and 10th Hung. Conf. Acoust., Balatonfüred* (1991)
8.26 M.A. Mironov: Parametric instability of a circular shell propagating a Korteweg wave, Acoust. Phys. **41**, 707–711 (1995)
8.27 M.P. Verge: *Aeroacoustics of Confined Jets with Applications to the Physical Modelling of Recorder-Like Instruments*, Ph.D. Thesis (University of Eindhoven, Eindhoven 1995)
8.28 S. Pitsch: *Schneidentonuntersuchungen an einem Orgelpfeifen-Fußmodell mittels Wasserkanal- und akustischen Messungen*, Diplomarbeit (Universität Stuttgart, Stuttgart 1996)
8.29 M. Castellengo: Acoustical analysis of initial transients in flute like instruments, Acustica **85**(3), 387–400 (1999)
8.30 A. Powell: On the edgetone, J. Acoust. Soc. Am. **33**, 395–409 (1961)
8.31 D.G. Crighton: The edgetone feedback cycle; linear theory for the operating stages, J. Fluid. Mech. **234**, 361–391 (1992)
8.32 W.K. Blake, A. Powell: The development of a contemporary view of flow tone generation. In: *Recent Advances in Aeroacoustics*, ed. by A. Krothapalli, C.A. Smith (Springer, New York 1983)
8.33 M.S. Howe: The role of displacement thickness fluctuations in hydroacoustics and the jet-drive mechanism of the flue organ pipe, Proc. R. Soc. Lond. A **374**, 543–568 (1981)
8.34 N. Zagyva: *Computer Modelling of the Onset of the Sound of Flue Organ Pipes*, MSc Thesis (ELTE University Budapest, Budapest 1993), in Hungarian
8.35 J. Angster, J. Angster, A. Miklós: Über die Messungen während des Intonationsprozesses Lippenpfeifen der Orgel, Instrumentenbau-Zeitschrift **45**, 71–76 (1991)
8.36 J. Angster, S. Pitsch, A. Miklós: Vergleich der subjektiven und objektiven Beurteilungen des Orgelpfeifenklangs. In: *Fortschritte der Akustik – DAGA'97*, ed. by P. Wille (DEGA, Kiel 1997) pp. 303–304
8.37 J. Angster, A. Miklós: End-correction of open flue organ pipes. In: *Fortschritte der Akustik DAGA'92, Berlin* (1992) pp. 260–263
8.38 J. Angster, A. Miklos: Transient sound spectra of a variable length organ pipe. In: *Int. Symp. on Music. Acoust., Tokyo* (1992) pp. 159–162
8.39 V. Rioux, D. Västfjäll, M.Z. Yokota, M. Kleiner: Noise quality of transient sounds: Perception of "hiss" and "cough" in a flue organ pipe, Acust.-Acta Acust. **85**, 76 (1999)
8.40 A. Miklós, J. Angster: Properties of the Sound of Flue Organ Pipes, Acta Acust. united Acust. **86**, 611–622 (2000)

9. Percussion Musical Instruments

Andrew C. Morrison, Thomas D. Rossing

Percussion instruments are an important part of every musical culture. Although they are probably our oldest musical instruments (with the exception of the human voice), there has been less research on the acoustics of percussion instruments, as compared to wind or string instruments. Quite a number of scientists, however, continue to study these instruments.

Over the years we have written several review articles on the acoustics of percussion instruments [9.1, 2] as well as a book [9.3]. They are also the subject of chapters in most books on musical acoustics and on musical instruments [9.4–8].

9.1 Drums

Drums generally have membranes of animal skin or synthetic material stretched over some type of air enclosure. Nowadays synthetic materials, such as Mylar (polyethylene terephthalate), are more common, although some percussionists still prefer animal skin (leather). Some type of tensioning device is nearly always included. The speed of waves on the membrane (and thus the frequency of the various modes) depends upon the tension, the thickness, and the density of the membrane. Some drums (e.g., timpani, tabla, boobams) sound a definite pitch; others convey almost no sense of pitch at all. Some drums have a single membrane (drumhead), while others include two membranes coupled together by vibrations of the drum shell and the enclosed air. The first 12 modes of vibration of a circular membrane are shown in Fig. 9.1. Above each sketch are given the values of m (the number of nodal diameters) and n (the number of nodal circles), and below it the frequency of vibration for that mode divided by the frequency of the lowest (01) mode. Mathematically, the mode frequencies of an ideal membrane are proportional to those of the mn Bessel function.

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_9

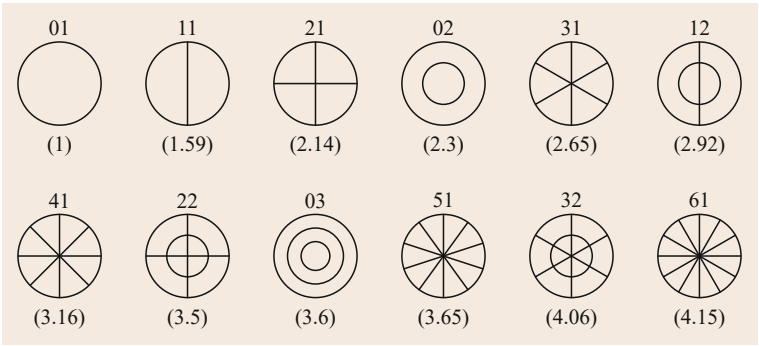


Fig. 9.1 Modes of vibration of a circular membrane

9.1.1 Timpani

The timpani or kettledrums are the most important drums in the orchestra, with one member of the percussion section usually devoting attention exclusively to them. Most modern timpani have a pedal-operated tensioning mechanism in addition to six or eight tensioning screws around the rim of the kettle. Although the modes of vibration of an ideal membrane are not harmonic, a carefully tuned kettledrum will sound a strong principal note plus two or more nearly harmonic overtones. *Rayleigh* [9.9] recognized the principal note as coming from the (11) mode and identified overtones about a perfect fifth ($f : f_1 = 3 : 2$), a major seventh ($15 : 8$), and an octave ($2 : 1$) above the principal tone. The inharmonic modes of an ideal membrane are shifted into a nearly harmonic series mainly by the effect of air loading [9.10]. Mode frequencies of a kettledrum, with and without the kettle, are given in Table 9.1.

Normal striking technique produces prominent partials with frequencies in the ratios $0.85 : 1 : 1.5 : 1.99 : 2.44 : 2.89$. If we ignore the heavily damped fundamental, the others are nearly in the ratios $1 : 1.5 : 2 : 2.5$, a harmonic series built on an octave below the principal note. Measurements on timpani of other sizes give similar results [9.11].

9.1.2 Snare Drums

The snare drum is a two-headed instrument about 33–38 cm in diameter and 13–20 cm deep. The shell is made from wood, metal, or Mylar. Strands of wire or gut are stretched across the lower (snare) head. When the upper (batter) head is struck, the snare head vibrates against the snares. The coupling between the snares and the snare head depends upon the mass and the tension of the snares. At a sufficiently large amplitude of the snare head, properly adjusted snares will leave the head at some point during the vibration cycle and then return to strike it, thus giving the snare drum its characteristic sound. The greater the tension on the snares, the larger the amplitude needed for this to take place [9.12]. Vibrational modes of a snare drum shell, with and without the drumheads, are shown in [9.12].

Table 9.1 Mode frequencies and ratios of kettledrum membranes with and without the kettle

Mode	Kettledrum		Drumhead alone		Ideal membrane
	f (Hz)	f/f_{11}	f (Hz)	f/f_{11}	f/f_{11}
01	127	0.85	82	0.53	0.63
11	150	1.00	155	1.00	1.00
21	227	1.51	229	1.48	1.34
02	252	1.68	241	1.55	1.44
31	298	1.99	297	1.92	1.66
12	314	2.09	323	2.08	1.83
41	366	2.44	366	2.36	1.98
22	401	2.67	402	2.59	2.20
03	418	2.79	407	2.63	2.26
51	434	2.89	431	2.78	2.29
32	448	2.99	479	3.09	2.55
61	462	3.08	484	3.12	2.61
13	478	3.19	497	3.21	2.66
42			515	3.32	2.89

9.1.3 Bass Drums

The bass drum is capable of radiating up to 20 W of peak acoustical power, probably the most of any instrument in the orchestra. A concert bass drum usually has a diameter of 80–100 cm, although smaller drums (50–75 cm) are popular in marching bands. Most bass drums have two heads, set at different tensions, although single-headed gong drums are used when a more defined pitch is desired. Mylar heads with a thickness of 0.25 mm are widely used, although calfskin heads are preferred by some drummers for large concert bass drums. Generally the batter or beating head is tuned to a greater tension than the carry or resonating head.

9.1.4 Tom-Toms

Tom-toms range from 20 to 45 cm in diameter, and they may have either one or two heads. Although often characterized as untuned drums, tom-toms do convey an identifiable pitch, especially the single-headed type. When a tom-tom is struck a hard blow, the deflection of the drumhead may be great enough to cause a significant change in the tension, which momentarily raises the frequencies of all modes of vibration and thus the apparent pitch. The fundamental frequency in a 33 cm tom-tom, for example, was found to rise about eight percent (slightly more than a semitone) during the first 0.2 s after the strike [9.13], resulting in a perceptible pitch glide. The pitch glide can be enhanced by loading the outer portion of the drumhead with a Mylar ring.

9.1.5 Indian Drums

Foremost among the drums of India are the tabla (north India) and mrdanga (south India). The overtones of both these drums are tuned harmonically by loading the drumhead with a paste of starch, gum, iron oxide, charcoal, or other materials. The tabla has a rather thick head made from three layers of animal skin (calf, sheep, goat, or buffalo skins are apparently used in different regions). The innermost and outermost layers are annular, and the layers are braided together at their outer edge and fastened to a leather hoop. Tension is applied to the head by means of a long leather thong that weaves back and forth between the top and bottom of the drum. The tabla is usually played together with a larger drum, called the banya or left-handed tabla. The head of the larger drum is also loaded, but slightly off center. The mrdanga is a two-headed drum that functions, in many respects, as a tabla and banya combined into one. The smaller head, like that of the tabla, is loaded with a patch of dried paste, while the larger head is normally loaded with a paste of wheat and water shortly before playing. A tabla and a mrdanga are shown in Fig. 9.2.

The acoustical properties of these drums have been studied by a succession of Indian scientists, including Nobel laureate C.V. Raman. *Raman* and his colleagues recognized that the first four overtones of the tabla are harmonics of the fundamental, and they identified these five harmonics as coming from nine normal modes [9.14]. For example, Fig. 9.3 shows how combinations of the (0,2) and (2,1) modes produce the third harmonic partial.

9.1.6 Japanese Drums

Drums have been used for centuries in Japanese temples. In Buddhist temples, it has been said that the sound of the drum is the voice of Buddha. In Shinto temples it is said that drums have a spirit (kumi) and that with a drum one can talk to the spirits of animals, water, and fire. Drums were often used to motivate warriors into battle and to entertain in town festivals and weddings [9.15]. The Japanese taiko (drum) has broken out of its traditional setting, and today's taiko bands have given new life to this old tradition. Japan's famous taiko band, the Kodo drummers, have performed in many countries of the world. Taiko bands exist in many Western countries. The o-daiko is a large drum consisting of two cowhide membranes stretched tightly across the ends of a wooden cylinder 50–100 cm in diameter and about 1 m in length. The drum, which hangs in a wooden frame, is struck with large felt-padded beaters. It is often used in religious functions at shrines, where its deep sound adds solemnity to the occasion. *Obata* and *Tesima* [9.15] found modes of vibration in the o-daiko to be somewhat similar to those in the bass drum. The tsudzumi

Fig. 9.2a,b The tabla (**a**) and mrdanga (**b**)

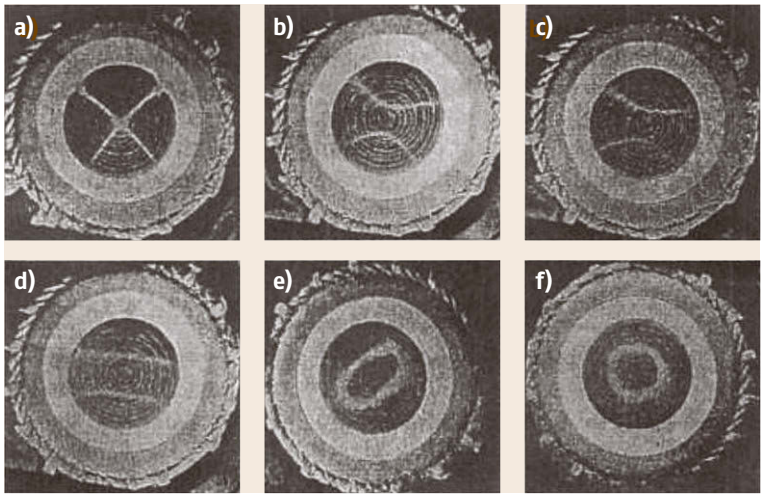


Fig. 9.3a–f Modes of vibration of the tabla (after [9.14]). **(a)** The (0,2) normal mode. **(b–e)** Combination of (0,2) and (2,1) normal modes. **(f)** The (2,1) normal mode

(or tsuzumi) is a braced drum whose body has cup-shaped ends and leather heads on both ends. A few sheets of paper wet with saliva cover an area of about $1\,\mathrm{cm}^2$ at the center of the bottom head, which tunes the modes of this head into a nearly harmonic relationship [9.16].

9.2 Mallet Percussion Instruments

9.2.1 Vibrating Bars

Bars or rods can vibrate either longitudinally or transversely. The most important vibrations in percussion instruments are the transverse bending vibrations in which internal elastic forces supply the necessary restoring force. When a bar is bent, the outer part is stretched and the inner part is compressed. Somewhere in between is a neutral axis whose length remains unchanged, as shown in Fig. 9.4.

A filament located at a distance z below the neutral axis is compressed by an amount $z\mathrm{d}\phi$. The strain is $z\mathrm{d}\phi/\mathrm{d}x$, and the amount of force required to produce the strain is $E\mathrm{d}Sz\mathrm{d}\phi/\mathrm{d}x$, where $\mathrm{d}S$ is the cross-sectional area of the filament and E is Young's modulus. This leads to a fourth-order differential equation whose solution can be found in [9.4, Sect. 2.15]. The solution leads to different modal frequencies, depending upon whether the ends of the bar or rod are free, clamped, or simply supported (hinged). The most commonly used bars in percussion instruments are bars that are free at both ends, whose relative frequencies are given by

$$f_n = \frac{\pi K}{8L^2}\sqrt{\frac{E}{\rho}}\left[3.011^2, 5^2, 7^2, ..., (2n+1)^2\right] .$$

The frequencies and nodal positions for the first four bending vibrational modes in a thin bar with free ends are given in Table 9.2.

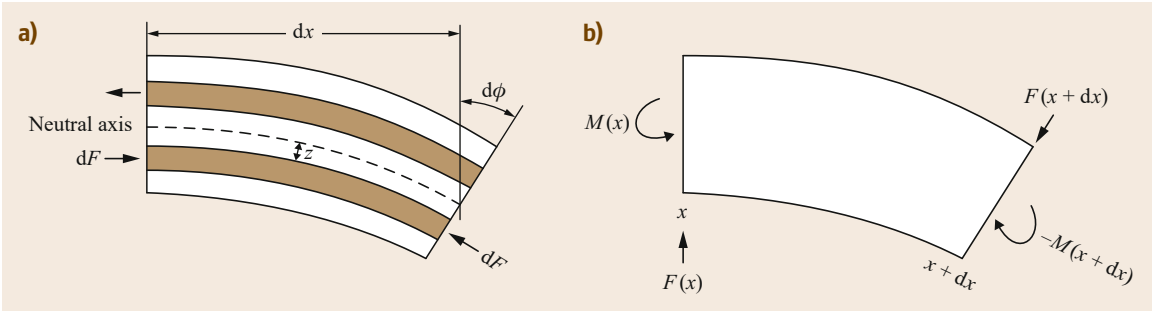


Fig. 9.4 **(a)** Bending strains in a bar. **(b)** Bending moments and shear forces in a bar

Table 9.2 Properties of transverse vibrations in a bar free at both ends

Frequency (Hz)	Wavelength (m)	Nodal positions (m from end of $1-m$ bar)
$f_1 = 3.5607\frac{K}{L}\sqrt{\frac{E}{\rho}}$	$1.330\,L$	0.224, 0.776
$2.756\,f_1$	$0.800\,L$	0.132, 0.500, 0.868
$5.404\,f_1$	$0.572\,L$	0.094, 0.356, 0.644, 0.906
$8.933\,f_1$	$0.445\,L$	0.073, 0.277, 0.500, 0.723, 0.927

9.2.2 Marimbas

In most of the world, the term marimba denotes a deep-toned instrument with tuned bars and resonator tubes that evolved from the early Latin American instrument. The marimba typically includes three to four-and-a-half octaves of tuned bars of rosewood or synthetic material with a deep arch cut to tune the overtones. The first overtone, which is radiated by the second bending mode, is normally tuned to the fourth harmonic of the fundamental in the first two to three-and-a-half octaves, after which the interval decreases [9.17]. Details of bar shapes for harmonic tuning are given by *Bork* [9.18]. Below each marimba bar is a cylindrical resonator pipe tuned to the fundamental mode of the bar. A pipe with one closed end and one open end resonates when its acoustical length is one-fourth of a wavelength of the sound. The tubular resonators emphasize the fundamental and also increase the loudness, which is done at the expense of shortening the decay time of the sound. The statement is sometimes made that the resonator prolongs the sound but that is incorrect. That impression may be conveyed when it is played with other instruments in an ensemble since the sound decay curve begins higher and may cross the background sound at a slightly later time. Some companies now make large five-octave concert marimbas that cover the range C2 to C7. In such instruments, generally the second bending mode is accurately tuned to the fourth harmonic in the first three-and-a-half octaves. The third bending mode is tuned to the tenth harmonic in the first two octaves, after which the interval decreases. The fourth mode varies from the 20th harmonic in the lowest bars to about the sixth harmonic in the highest bars [9.19]. Relative frequencies of the first four bending modes in a Malletech marimba are shown in Fig. 9.5a, while those of several torsional modes in the same marimba are shown in Fig. 9.5b. The first torsional mode frequency ranges from about 1.9 times the nominal frequency (largest bars) to about 1.2 times the nominal frequency (smallest bars).

In normal playing, the bars are struck near their centers, where the torsional (twisting) modes have nodes, and thus they will not be excited to any great extent. On the other hand, if the bars are struck away from the center, deliberately or not, the torsional modes may contribute to the timbre. Applying finite element methods to marimba and xylophone bars showed that a small curvature in the bars has very little effect on the relative frequencies of the vibrational modes. Henrique and Antunes have used finite element methods both to optimize the shape of marimba and xylophone bars and to model the sound. They employ a physical modeling approach that addresses the spatial aspects of the problem and is suitable for both dispersive and nondispersive systems [9.20]. The sound field radiated by a simulated marimba bar has been calculated by assuming the vibrating bar to be equivalent to a linear array of oscillating spheres. This sound pressure excites a monodimensional lossy tube of finite length terminated by a radiation impedance at its open end, which represents the tubular resonator. The amount of frequency decrease as the resonator is moved closer to the bar is then calculated [9.21].

9.2.3 Xylophones

Xylophones also use bars of wood or synthetic material, but the arch is not cut as deep as that of a marimba. The first overtone is tuned to the third rather than the fourth harmonic of the fundamental. The closed-tube resonators placed below the bars reinforce the third harmonic as well as the fundamental, thus producing a brighter sound than the marimba. This is further enhanced by using hard mallets.

9.2.4 Vibes

Vibraphones or vibraharps, as they are called by different manufacturers, have aluminum bars deeply arched (as in marimbas) so that the first overtone has a frequency four times that of the fundamental. The aluminum bars in vibes have much longer decay times than the wood or synthetic bars of the marimba, and so vibes are equipped with pedal-operated dampers. The most distinctive feature of vibes, however, is the vibrato introduced by motor-driven discs at the top of the resonators, which alternately open and close the tubes. The vibrato produced by these rotating discs of pulsators produces a vibrato (hence the name). The speed of rotation of the discs may be adjusted to produce a slow

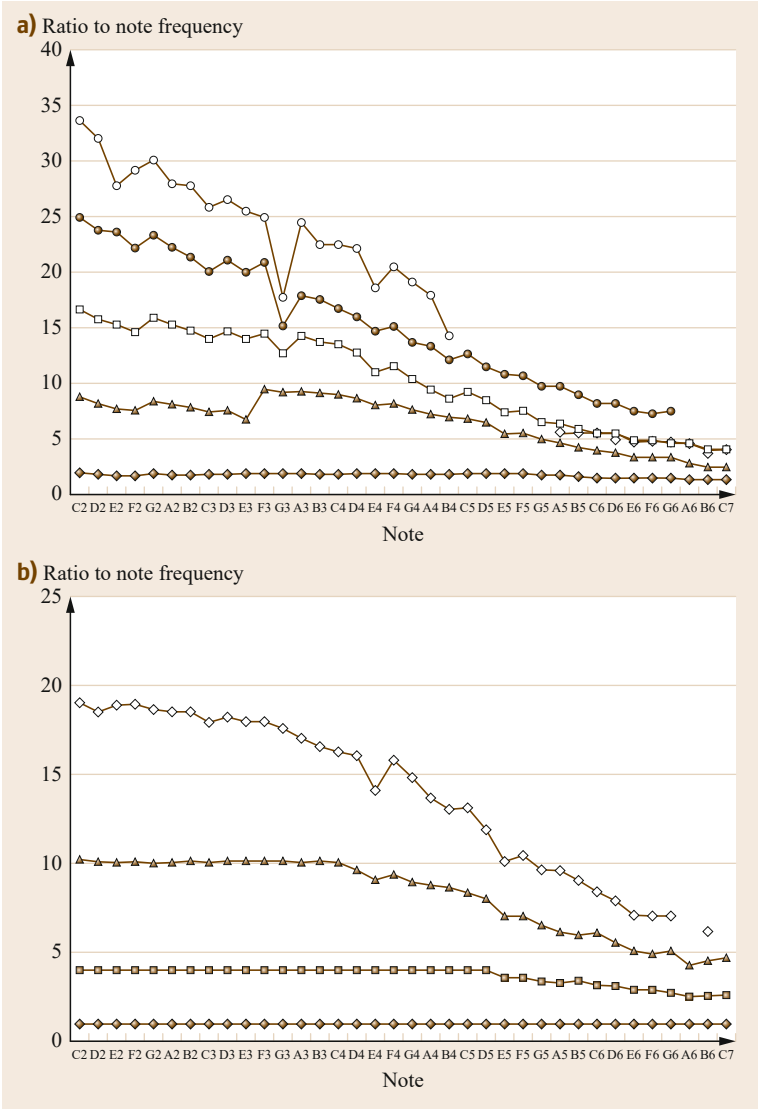


Fig. 9.5a,b Torsional modes (a) and bending modes (b) in a Malletech five-octave marimba

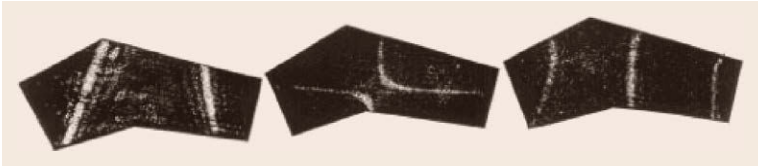


Fig. 9.6 Holographic interferogram showing vibrational modes of a jade chime stone

vibe or a fast vibe. Sometimes vibes are played without vibrato by switching off the motor. Vibes are generally played with soft mallets that produce a mellow tone.

9.2.5 Glockenspiel

The glockenspiel, or orchestra bells, uses rectangular steel bars 2.5–3.2 cm wide and 8–9 cm thick. Its range is customarily from G5 to C8, although it is scored two octaves lower than it sounds. The glockenspiel is usually played with brass or hard plastic mallets. The bell lyra is a portable version, popular in marching bands, that uses aluminum bars. Because the high overtones die out quickly, no effort is made to tune the overtones harmonically, as in the marimba, xylophone, and vibes.

Fig. 9.7 Set of 16 pyeon-gyoung (stone chimes) from the Chosun Dynasty in Korea

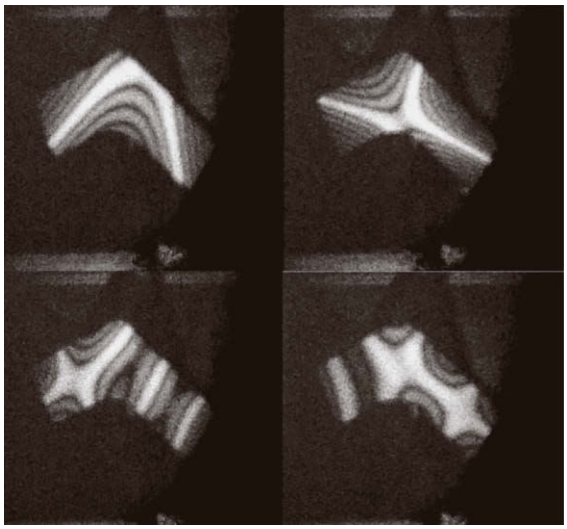


Fig. 9.8 Interferograms showing vibrational modes of a Korean pyeon-gyoung stone

9.2.6 Chimes

Chimes or tubular bells are generally fabricated from lengths of brass tubing 32–38 mm in diameter. The upper end of each tube is partially or completely closed by a brass plug with a protruding rim. The rim forms a convenient and durable striking point. The modes of transverse vibration in a pipe are essentially those of a thin bar. One of the most interesting characteristics of chimes is that there is no mode of vibration with a frequency at the pitch of the strike tone one hears. Modes four, five and six, which are near the ratios 2 : 3 : 4 in a beam or tube, appear to determine the strike tone, which is heard one octave below the fourth mode [9.1].

9.2.7 Lithophones

Lithophones are stones that vibrate and produce sound. The ancient Chinese were fond of stone chimes, many of which have been found in ancient Chinese tombs. A typical stone chime was shaped to have arms of different lengths joined at an obtuse angle. The stones were generally struck on their longer arm with

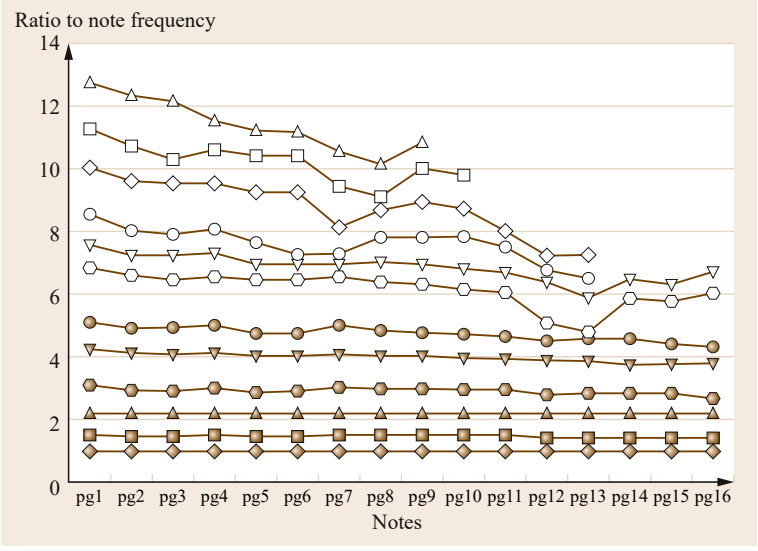


Fig. 9.9 Relative frequencies of the pyeon-gyoung stone

a wooden mallet. Sometimes the stones were richly ornamented. A lithophone of 32 stone chimes found in the tomb of the Marquis Yi (which also contained a magnificent set of 65 bells) was scaled in size, although the dimensions of the chimes do not appear to follow a strict scaling law [9.22]. In later times, the Chinese made stone chimes of jade. Holographic interferograms showing some of the modes of vibration of a small jade chime are shown in Fig. 9.6. Korean chime stones, called pyeon-gyoung, were originally brought from China to Korea in the 12th century.

A set of 16 stone chimes from the Chosun Dynasty is shown in Fig. 9.7. Unlike the Chinese stone chimes, these stones all have the same size but differ from each other only in thickness. The fundamental frequency is essentially proportional to the thickness, just as in a rectangular bar such as a marimba bar. The second mode in each stone is approximately 1.5 times the fundamental, while the third mode is about 2.3 times the nominal frequency. The fourth mode is about three times the nominal frequency up to the 12th stone, after which the ratio drops to about 2.7 [9.23]. Holographic interferograms of several modes of vibration in a pyeon-gyoung stone tuned to D♯6 are shown in Fig. 9.8. Relative frequencies of the modes in the pyeon-gyoung are shown in Fig. 9.9.

9.3 Cymbals, Gongs, and Plates

The vibrations of plates have fascinated scientists, as well as musicians, for many years. Nearly 200 years ago, *E.F.F. Chladni* published a book describing his well-known method of sprinkling sand on vibrating plates to made the nodal lines visible [9.24]. Chladni's lectures throughout Europe attracted any famous persons, including Napoleon. The nodal lines in the vibrational modes of a circular plate are not too different from those in a circular membrane, shown in Fig. 9.1. The modal frequencies are very different, however, because the stiffness of the plate contributes a substantial amount of elastic restoring force. In fact, a plate will vibrate without externally applied tension. The modes of a circular plate are often given the labels m and n, like those of a membrane, to designate the numbers of nodal diameters and nodal circles. Chladni observed that the frequencies of the various modes of a circular plate are nearly proportional to $(m+2n)^2$, a relationship that has been called Chladni's law [9.25].

9.3.1 Cymbals

Cymbals are very old instruments and have had both religious and military use in a number of cultures. The Turkish cymbals generally used in orchestras and bands are saucer-shaped with a small dome in the center, in contrast to Chinese cymbals, which have a turned-up edge. Orchestral cymbals are often designated as French, Viennese, and Germanic in order of increasing thickness. Jazz drummers use cymbals designated by such onomatopoeic terms as *crash*, *ride*, *swish*, *splash*, *ping*, and *pang*. Cymbals range from 20 to 75 cm in diameter. The strong aftersound that gives cymbal sound its characteristic shimmer is known to involve nonlinear processes [9.17]. There is considerable evidence that the vibrations exhibit chaotic behavior. A mathematical analysis of cymbal vibrations using nonlinear signal processing methods reveals that there are between three and seven active degrees of freedom, and that physical modeling will require a like number of equations [9.26]. One procedure is to calculate Lyapunov exponents from experimental time series, so that the complete spectrum of exponents can be obtained. The chaotic regime can be quantified in terms of the largest Lyapunov exponent [9.27].

9.3.2 Gongs

Gongs of many different sizes and shapes are popular in both Eastern and Western music. They are usually cast of bronze with a deep rim and a protruding dome. Tamtams are similar to gongs and are often confused with them. The main differences between the two are that tamtams do not have the dome of the gong, their rim is not as deep, and the metal is thinner. Tamtams generally sound a less definite pitch than do gongs. In fact, the sound of a tamtam may be described as somewhere between the sounds of a gong and a cymbal. The sound of a large tamtam develops slowly, changing from a sound of low pitch at strike to a collection of high-frequency vibrations, which are described as shimmer. These high-frequency modes fail to develop if the tamtam is not hit hard enough, indicating that the conversion of energy takes place through a nonlinear process [9.28].

9.3.3 Chinese Gongs

Among the many gongs in Chinese music are a pair of gongs used in Chinese opera orchestras, shown in Fig. 9.10. These gongs show a pronounced nonlinear behavior. The pitch of the larger gong glides downward

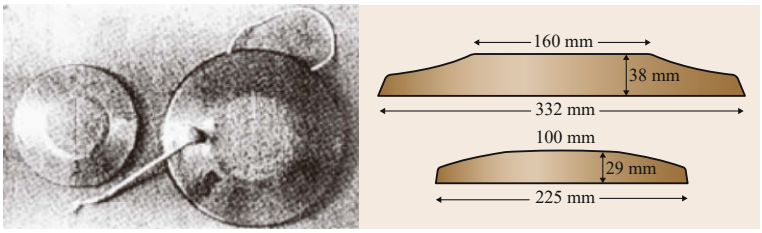


Fig. 9.10 Examples of gongs used in Chinese opera

as much as three semitones after striking, whereas that of the smaller gong glides upward by about two semitones [9.28]. Several vibrational modes of the larger gong are shown in Fig. 9.11.

In some of the modes, vibrations are confined pretty much to the flat inner portion of the gong, some to the sloping shoulders, and some involve considerable motion in both parts. When the gong is hit near the center, the central modes (178, 362, 504, 546 Hz) dominate the sound. When the gong is hit lightly on the shoulder, the lowest mode at 118 Hz is heard. The vibrations of a large tamtam were studied by *Chaigne* et al. [9.29]. They found that the nonlinear phenomena have the character of quadratic nonlinearity. Forced excitation at sufficiently large amplitude at a frequency close to one mode leads to a bifurcation with the appearance of lower frequencies corresponding to other modes. Varying the excitation frequency at constant force yielded subharmonics that were not observed at constant excitation frequency. This is quite similar to the nonlinear behavior of cymbals [9.17] [9.26].

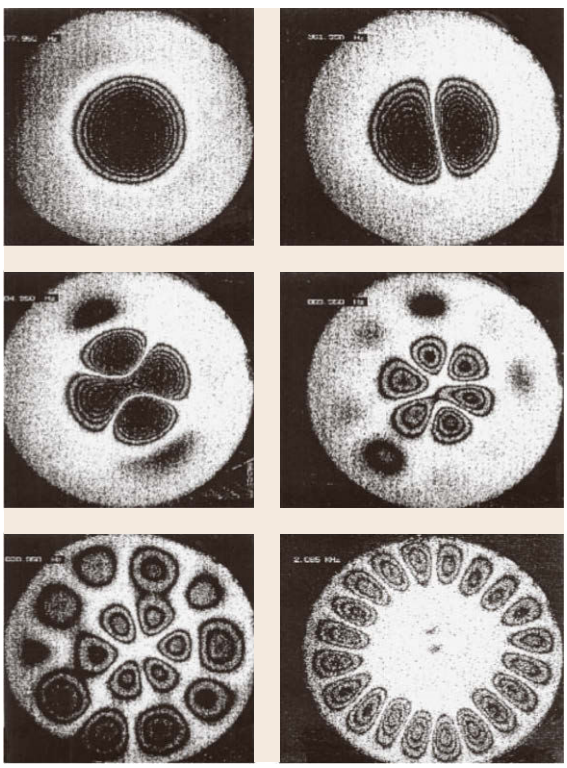


Fig. 9.11 Holographic interferograms of the modes of vibration of the larger gong shown in Fig. 9.10

9.3.4 The Caribbean Steelpan

The Caribbean steelpan is one of the most widely used acoustical instruments developed in the last 70 years. The instrument was developed on the islands of Trinidad and Tobago when local craftsman discovered methods of transforming surplus 50-gallon oil barrels into tuned drums. The Caribbean steelpan is an object of considerable acoustical study, both in its home country of Trinidad and Tobago and in the United States. Modern steel bands include a variety of instruments, such as tenor, double second, double tenor, guitar, cello, quadraphonic, and bass. Our earlier review paper [9.17] included holographic interferograms of several instruments showing how individual notes vibrate, how the entire instrument vibrates, and how the skirts of the instruments vibrate. Another piece of the puzzle, so to speak, is to understand how the vibrating components radiate sound. An effective aid to understanding sound radiation is to map the sound intensity field around the instrument. Since sound intensity is the product of sound pressure (a scalar quantity) and the acoustic fluid velocity (a vector), a two-microphone system is used. The acoustic fluid velocity can be readily calculated from the difference in sound pressure at the two accurately spaced microphones. Both the active intensity and the reactive intensity can be obtained at the desired points in the sound field. The active intensity represents the outward flow of energy, while the reactive intensity represents energy that is stored in the sound field near the instrument. While the active intensity is the most significant field in a concert hall, both active and reactive intensity fields have to be considered in recording a steelpan. Figure 9.12 shows a map of active and reactive sound intensity in a plane that bisects a double second steelpan when a single note ($F\sharp 3$) is excited at its fundamental frequency [9.30].

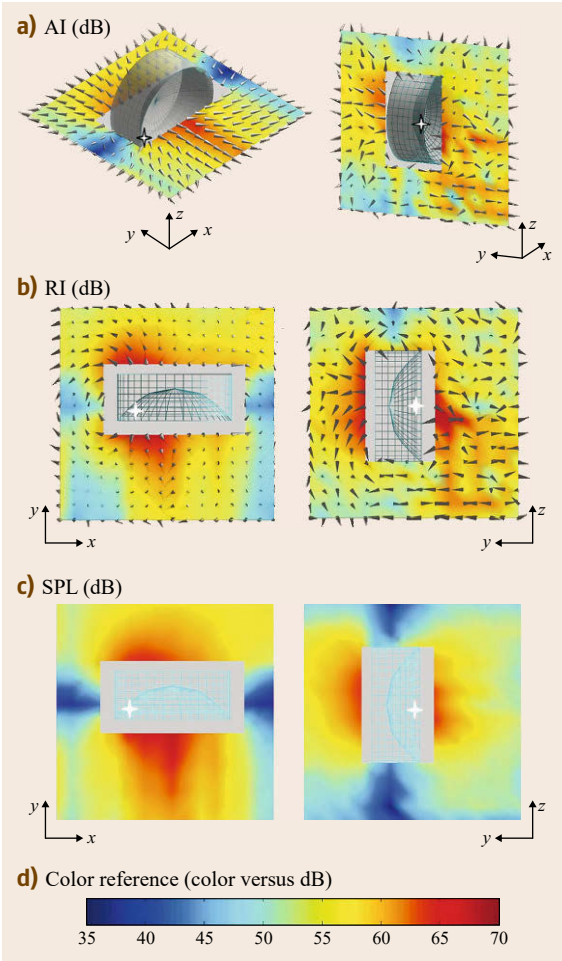


Fig. 9.12a–d Active intensity (AI) (a) and reactive intensity (RI) (b) of a Caribbean steelpan. (c) Sound pressure level (SPL), (d) color reference

9.3.5 The Hang

The Hang is a new steel percussion instrument, consisting of two spherical shells of steel, suitable for playing with the hands. Seven to nine notes are harmonically tuned around a central deep note, which is formed by the Helmholtz (cavity) resonance of the instrument body. The Hang shown in Fig. 9.13 has eight notes that can be tuned in any tonal systems between A3 and G5, including 30 tonal systems suggested by the tuners. The central note is usually tuned a fifth or fourth below the lowest note of the scale. Although it is a new instrument, many units have been shipped all over the world by PanArt, its creators. Holographic interferograms in Fig. 9.14 show the first five vibrational modes in the G3 note area of the Hang. The second and third modes are tuned to the second and third harmonics of the fundamental mode respectively [9.31].

Fig. 9.13 The Hang (image credit: Michael Paschko)

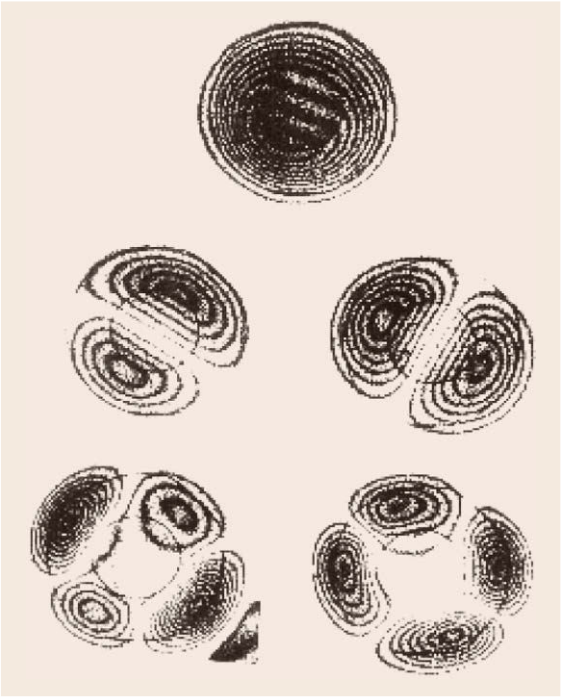


Fig. 9.14 The first five modes of vibration of the G3 note of the Hang

Figure 9.15 shows the active sound intensity in a plane 8 cm above the E4 note. The arrowheads show the direction of the sound intensity at each point in the plane, while the gray scale shows the sound pressure level. Note the sound level is greatest at the fundamental frequency, and the sound intensity is strongly upward from the note, while at the frequency of the second and third modes, considerable sound is radiated laterally.

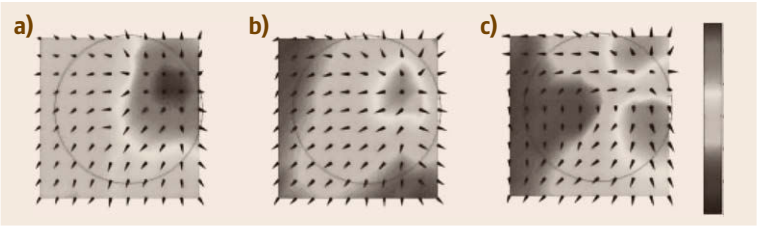


Fig. 9.15a–c Active intensity of the Hang when the E4 note is excited in **(a)** its fundamental mode **(b)** its second harmonic mode **(c)** its third mode

9.3.6 Bells

Bells have been a part of nearly every culture in history. Bells existed in the Near East before 1000 BCE, and a number of Chinese bells from the time of the Shang dynasty (1600–1100 BCE) can be found in museums around the world. In 1978 set of tuned bells from the fifth century BCE was discovered in the Chinese province of Hubei [9.32]. Bells developed as Western musical instruments in the seventeenth century when bell founders discovered how to tune their partials harmonically. The founders in the Low Countries, especially the Hemony brothers (François and Pieter) and Jacob van Eyck, took the lead in tuning bells, and many of their fine bells are found in carillons today. When struck by its clapper, a bell vibrates in a complex way. In principle, its vibrational motion can be described in terms of a linear combination of the normal modes of vibration whose initial amplitudes are determined by the distortion of the bell when struck. In practice, such a description becomes quite complex because of the large number of normal modes of diverse character that contribute to the motion. The first five modes of a church bell or carillon bell are shown in Fig. 9.16. Lines show the locations of the nodal lines. The numbers at the top denote the numbers of complete nodal meridians extending over the top of the bell and the number of nodal circles respectively. Note that there are two modes with $m = 3$ and $n = 1$, one with a circular node at the waist and one with a node near the sound bow. Thus we denote the one as (3,1♯) in Fig. 9.16. The ratio of each modal frequency to that of the prime is given at the bottom of each diagram.

When a large church bell or carillon bell is struck by its clapper, one first hears the sharp sound of metal on metal. This sound quickly gives way to a strike note that is dominated by the prominent partials of the bell. Most observers identify the metallic strike note as having a pitch at or near the frequency of the second partial. Finally, as the sound of the bell ebbs, the slowly decaying hum tone (an octave below the prime) lingers on. A new type of carillon bell, that has the dominating minor-third partial (Fig. 9.16) replaced by a partial tuned a major-third above the prime, has been developed at the Royal Eijsbouts bell foundry in The Netherlands [9.32]. The new bell design evolved partly from the use of a technique for structural optimization using finite element methods [9.33]. This technique allows a designer to make changes in the profile of an existing structure, and then to compute the resulting changes in the vibrational modes. Based on the results of the structural optimization procedure, *André Lehr* and his colleagues have designed two different bells, each having a major-third partial [9.34].

9.3.7 Handbells

Although handbells date back to at least several millennia BCE, handbells developed as Western musical instruments in the 18th century. One early use was to provide tower bell-ringers with a convenient means to practice change ringing. In more recent years, handbell choirs have become popular in schools and churches; some 40 000 choirs are reported in the United States alone. Handbells have modes of vibration somewhat similar to those of church bells or carillon bells. Hologram interferograms of a number of modes in a C5 handbell are shown in Fig. 9.17. Nodes show as bright lines, and the bullseyes locate the antinodes. In a well-tuned handbell, the (3,0) mode with three nodal meridians is tuned to a frequency three times that of the fundamental (2,0) mode.

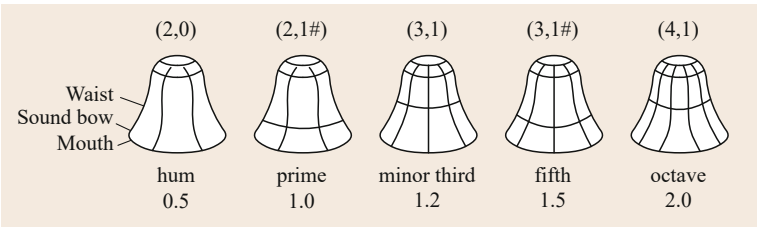


Fig. 9.16 The first five modes of a church or carillon bell

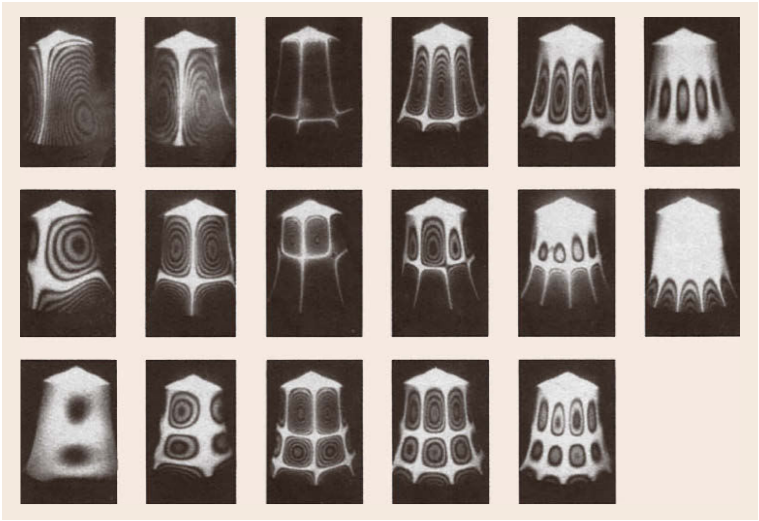


Fig. 9.17 Holographic interferograms of the vibrational modes of a C5 handbell

9.4 Methods for Studying the Acoustics of Percussion Instruments

Recent studies of the acoustics of percussion instruments have included:

1. Theoretical studies of modes of vibration
2. Experimental studies of modes of vibration
3. Sound radiation studies
4. Physical modeling
5. Studies of nonlinear behavior.

9.4.1 Finite Element and Boundary Element Methods

For all but the simplest vibrator shapes, it is difficult to calculate vibrational modes analytically. Fortunately, there are powerful numerical methods that can be carried out quite nicely by use of digital computers. These are generally described as finite element methods or boundary element methods.

9.4.2 Experimental Studies of Modes of Vibration

When a percussion instrument is excited by striking (or bowing or plucking), it vibrates in a rather complicated way. The motion can be conveniently described in terms of normal modes of vibration. A normal mode of vibration represents the motion of a linear system at a normal frequency (eigenfrequency). It should be possible to excite a normal mode of vibration at any point in a structure that is not a node and to observe motion at any other point that is not a node. It is a characteristic only of the structure itself, independent of the way it is excited or observed. In practice, however, it is difficult to avoid small distortions of the normal modes due to interaction with the exciter, the sensor, and especially the supports. Normal modes shapes are unique for a structure, whereas the deflection of a structure at a particular frequency, called its operating deflection shape (ODS), may result from the excitation of more than one normal mode [9.35].

Normal mode testing has traditionally been done using sinusoidal excitation, either mechanical or acoustical. Detection of motion may be accomplished by attaching small accelerometers, although optical and acoustical methods are less obtrusive. Modal testing with impact excitation, which became popular in the 1970s, offers a fast, convenient way to determine the normal modes of a structure. In this technique, an accelerometer is generally attached to one point on the structure, and a hammer with a load cell is used to impact the structure at carefully determined positions. Estimates of modal parameters are obtained by applying some type of curve-fitting program. Experimentally, all modal testing is done by measuring operating deflection shapes and then interpreting them in a specific manner to define mode shapes [9.35]. Strictly speaking, some type of curve-fitting program should be used to determine the normal modes from the observed ODSs, even when an instrument is excited at a single frequency. In practice, however, if the mode overlap is small, the single-frequency ODSs provide a pretty good approximation to the normal modes.

9.4.3 Scanning with a Microphone or an Accelerometer

Probably the simplest method for determining ODSs (and hence normal modes) is to excite the structure at single frequency with either a sinusoidal force or a sinusoidal sound field, and to scan the structure with an accelerometer or else to scan the near-field sound with a small microphone [9.36]. With practice, it is possible to determine mode shapes rather accurately by this method.

9.4.4 Holographic Interferometry

Holographic interferometry offers by far the best spatial resolution of operating deflection shapes (and hence of normal modes). Whereas experimental modal testing and various procedures for mechanical, acoustical, or optical scanning may look at the motion at hundreds (or even thousands) of points, optical holography looks at an almost unlimited number of points. Recording holograms on photographic plates or film (as in the holographic interferograms shown in Fig. 9.17) tends to be rather time consuming since each mode of vibration must be recorded and viewed separately. TV holography, on the other hand, is a fast, convenient way to record ODSs and to determine the normal modes. An optical system for TV holography is shown in Fig. 9.18.

A beam splitter (BS) divides the laser light to produce a reference beam and an object beam. The reference beam reaches the charge-coupled device (CCD) camera via an optical fiber, while the object beam is reflected by phase modulated (PM) mirror so that it illuminates the object to be studied. Reflected light from the object reaches the CCD camera, where it interferes with the reference beam to produce the holographic image. The speckle-averaging mechanism (SAM) alters the illumination angle in small steps in order to reduce laser speckle noise in the interferograms. Generally holographic interferograms show only variations in amplitude. It is possible, however, to recover phase information by modulating the phase of the reference beam by moving PM mirror at the driving frequency. This is a useful technique for observing motion of very small amplitude or resolving normal modes of vibration that are very close in frequency.

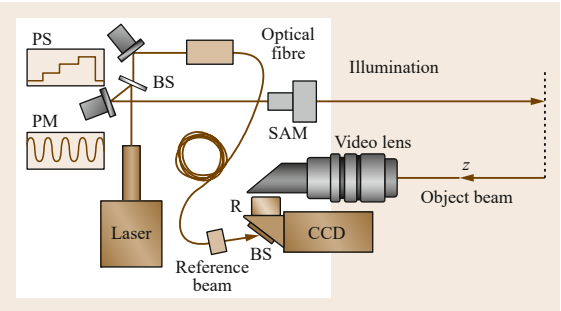


Fig. 9.18 Optical layout for a TV holography system

9.4.5 Experimental Modal Testing

Modal testing may be done with sinusoidal, random, pseudorandom, or impulsive excitation. In the case of sinusoidal excitation, the force may be applied at a single point or at several locations. The response may be measured mechanically (with accelerometers or velocity sensors), optically, or indirectly by observing the radiated sound field. In modal testing with impact excitation, an accelerometer is typically attached to a force transducer (load cell). Each force and acceleration waveform is Fourier transformed and a transfer function H_{ij} is calculated. Several different algorithms may be used to extract the mode shape and modal parameters from the measured transfer functions [9.35].

9.4.6 Radiated Sound Field

The best way to describe sound radiation from complex sources such as percussion instruments is by mapping the sound intensity field. Sound intensity is the rate at which sound energy flows outward from various points on the instrument. The sound intensity field represents the direction and the magnitude of the sound intensity at every point in the space around the source. A single microphone measures the sound pressure at a point, but not the direction of the sound energy flow. In order to determine the sound intensity it is necessary to compare the signals from two identical microphones spaced a small distance apart. The resulting pressure gradient can be used to determine sound intensity. When this is done at a large number of locations, a map of the sound intensity field results [9.30, 37, 38].

9.4.7 Physical Modeling

Synthesizing sounds by physical modeling has attracted a great deal of interest in recent years. The basic notion of physical modeling is to write equations that describe how particular sets of physical objects vibrate and then to solve those equations in order to synthesize the resulting sound. Percussion instruments have proven particularly difficult to model completely enough to be able to synthesize their sounds entirely based on a physical model. Physical modeling is complicated by their nonlinear behavior and by the strong role that transients play in their sound.

References

9.1 T.D. Rossing: Acoustics of percussion instruments part I, Phys. Teach. **14**, 546–556 (1976)
9.2 T.D. Rossing: Acoustics of percussion instruments part II, Phys. Teach. **15**, 278–288 (1977)
9.3 T.D. Rossing: *Science of Percussion Instruments* (World Scientific, Singapore 2000)
9.4 N.H. Fletcher, T.D. Rossing: *The Physics of Musical Instruments*, 2nd edn. (Springer, New York 1998)
9.5 T.D. Rossing, F.R. Moore, P.A. Wheeler: *The Science of Sound*, 3rd edn. (Addison Wesley, San Francisco 2002)
9.6 D.E. Hall: *Musical Acoustics* (Brooks/Cole, Pacific Grove 2002)
9.7 M. Campbell, C. Greated, A. Myers: *Musical Instruments* (Oxford Univ. Press, Oxford 2004)
9.8 J. Meyer: *Acoustics and the Performance of Music* (Springer, New York 2009), transl. by U. Hansen
9.9 Lord Rayleigh (J.W. Strutt): *The Theory of Sound*, Vol. 1, 2nd edn. (Macmillan, London 1894), reprinted by Dover, New York 1945
9.10 T.D. Rossing: The physics of kettledrums, Sci. Am. **247**(5), 172–178 (1982)
9.11 T.D. Rossing, G. Kvistad: Acoustics of timpani: preliminary studies, The Percussionist **13**, 90–98 (1976)
9.12 T.D. Rossing, I. Bork, H. Zhao, D. Fystrom: Acoustics of snare drums, J. Acoust. Soc. Am. (1992), https://doi.org/10.1121/1.404080
9.13 C.D. Rose: *A New Drumhead Design: An Analysis of the Nonlinear Behavior of a Compound Membrane*, M.S. Thesis (Northern Illinois University, DeKalb 1978)
9.14 C.V. Raman: The Indian musical drum, Proc. Indian Acad. Sci. A1 (1934), https://doi.org/10.1007/BF03035705
9.15 J. Obata, T. Tesima: Experimental studies on the sound and vibration of drum, J. Acoust. Soc. Am. **6**(4), 267–274 (1935)
9.16 S. Ando: Acoustical studies of Japanese traditional drums. In: *Jt. Meet. Acoust. Soc. Am/Acoust. Soc. Jn, Honolulu* (1996), Paper 4aMUb3
9.17 T.D. Rossing: Acoustics of percussion instruments: Recent progress, Acoust. Sci. Technol. **22**, 177–188 (2001)
9.18 I. Bork: Practical tuning of xylophone bars and resonators foobar, Appl. Acoust. **46**, 103–127 (1995)
9.19 J. Yoo, T.D. Rossing, B. Larkin: Vibrational modes of five-octave concert marimbas. In: *Proc. Stockholm Music Acoust. Conf. (SMAC03), Stockholm* (2003) pp. 355–357
9.20 L. Henrique, J. Antunes: Optimal design and physical modelling of mallet percussion instruments, Acta Acustica/Acustica **89**, 948–963 (2003)
9.21 V. Doutaut, A. Chaigne, G. Bedrane: Time-domain simulation of the sound pressure radiated by mallet percussion instruments. In: *Proc. ISMA, Dourdan* (1995) pp. 519–524
9.22 A. Lehr: Designing chimes and carillons in history, Acustica/Acta Acustica **83**, 320–336 (1997)
9.23 J. Yoo, T.D. Rossing: Vibrational modes of pyen-gyoung, Korean chime stone. In: *Proc. ISMA 2004, Nara* (2004) pp. 312–315
9.24 E.F.F. Chladni: *Entdeckungen über die Theorie des Klanges* (Breitkopf Härtel, Leipzig 1787), translated excerpts in R.B. Lindsay: *Acoustics: Historical and Philosophical Development* (Dowden Hutchinson Ross, Stroudsburg 1973) pp. 156–165
9.25 T.D. Rossing: Chladni's law for vibrating plates, Am. J. Phys. **50**, 271–274 (1982)
9.26 C. Touze, A. Chaigne, T. Rossing, S. Schedin: Analysis of cymbal vibrations and sound using nonlinear signal processing methods. In: *Proc. ISMA98* (1998)
9.27 C. Touzé, A. Chaigne: Lyapunov exponents from experimental time series: application to cymbal vibrations, Acta Acustica/Acustica **86**(3), 557–567 (2000)
9.28 T.D. Rossing, N.H. Fletcher: Acoustics of a tamtam, Bull. Australian Acoust. Soc. **10**(1), 21–26 (1982)
9.29 A. Chaigne, C. Touzé, O. Thomas: Nonlinear axisymmetric vibrations of gongs. In: *Proc. ISMA 2001, Perugia* (2001) pp. 147–152
9.30 B. Copeland, A. Morrison, T.D. Rossing: Sound radiation from Caribbean steelpans, J. Acoust Soc. Am. **117**, 375–383 (2005)
9.31 T.D. Rossing, U.J. Hansen, F. Rohner, S. Schärer: The HANG: A hand-played steel drum. In: *Proc. SMAC 93, Stockholm* (2003) pp. 351–354
9.32 T.D. Rossing: Acoustics of Eastern and Western bells, old and new, J. Acoust. Soc. Jpn. (E) **10**, 241–252 (1989)
9.33 B. Schoofs, F. van Asperen, P. Maas, A. Lehr: Computation of bell profiles using structural optimization, Music Percept **4**, 245–254 (1985)
9.34 A. Lehr: *The Designing of Swinging Bells and Carillon Bells in the Past and Present* (Athanasius Kircher Foundation, Asten 1987)
9.35 M.H. Richardson: Is it a mode shape, or an operating deflection shape?, Sound Vibr. **31**(1), 54–61 (1997)
9.36 T.D. Rossing, D.A. Russell: Laboratory observation of elastic waves in solids, Am. J. Phys. **58**, 1153–1162 (1990)
9.37 F.J. Fahy: *Sound Intensity*, 2nd edn. (E. F.N. Spon, London 1995)
9.38 A. Morrison: *Acoustical Studies of the Steelpan and HANG: Phase-Sensitive Holography and Sound Intensity Measurements*, Ph.D. Thesis (Northern Illinois University, DeKalb 2005)

10. Musical Instruments as Synchronized Systems

Rolf Bader

Most musical instrument families have nearly perfect harmonic overtone series, for example plucked, bowed, or wind instruments. However, when considering the complex geometry and nonlinear driving mechanisms many of these instruments have we would expect them to have very inharmonic overtone series. So to make musical instruments play notes that we accept as harmonic sounds, synchronization needs to occur to arrive at the perfect harmonic overtone series the instruments actually produce. The reasons for this synchronization are different in the singing voice, organs, saxophones or clarinets, violin bowing or in plucked stringed instruments. However, when examining the mechanisms of synchronization further, we find general rules and suitable algorithms to understand the basic behavior of these instruments.

10.1 Added versus Intrinsic Synchronization

Nonlinearities contribute new aspects to the sound of musical instruments which are often considered to be the most interesting. Generally, brightness is increased, attacks are sharpened, or fluctuations are intensified when additional nonlinearities in the vibrating geometries of musical instruments appear compared to the linear case. Most listeners find an increased brightness more attractive and a sharpened attack helpful when following a melody. Fluctuations are often associated with liveliness in the tone and therefore are often preferred compared to a completely stable tone. These additional attributes of tone color are caused by a variety of effects which will be discussed below.

Conversely, musical instruments may have nonlinearities within their actual driving mechanisms. Then these nonlinearities are no longer add-ons to a linear system but are the very core of the instrument or the driving mechanism. In these cases, the sounds pro-

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_10

duced show a very complex nature with many different ways to produce the sounds. So, for example, the saxophone may be played with normal pressure producing a harmonic tone. However, it may also be played with a pressure too low to reach the stable state, thus only producing noise. Additionally, sophisticated players are also able to produce fancy sounds like multiphonics, where two, three, or even five tones are played simultaneously on the instrument, which are mostly in a complex mathematical ratio one to another. All these behaviors are caused by the same single driving mechanism which is nonlinear in nature and therefore able to produce these very different sounds. In these cases, the normal tone is often produced by synchronizing the different vibrating systems one to another, where one system tells the other in which frequencies to vibrate. Therefore, these instruments can be treated with methods known from self-organization and synergetics known from many other physical, biological, or neural systems.

When discussing linear and nonlinear behavior, taking the notions mathematically it is clear that there can only be one linear case – there is only one straight line with some slope – but there are endless possibilities of nonlinear cases – a curve may show an endless variety of different turns and twists. Linearity is present when the rule holds that if one unit is increased by a fixed amount another unit is also in- or decreased by another also fixed amount, no matter where we set the starting point. This is a simple property of a straight line. Conversely, for a nonlinear system this no longer holds. So if one adds a certain amount to one unit, another unit in- or decreases to a certain amount, too, but this amount now depends on the starting point. So an exponential function is a nonlinear curve, but also a logarithm is one. Such functions appear with large displacements of a string or with damping in musical instruments. There may be sudden jumps in the curve like the change from sticking to slipping with the violin bow attached or sliding over a string. There may also be a Gauss-shaped curve as appears with blown instruments when comparing blowing pressure and flow into the mouthpiece.

Because of this large variety of nonlinear cases only a few frameworks have been suggested to be able to explain musical instruments in general taking the nonlinearities into consideration. These ideas are discussed at the end of this chapter, focusing on different questions like mode coupling, determination of stability, or transient behavior. In the following sections, the most important examples of such nonlinear or synchronizing behavior are discussed. First, however, we provide an overview of the different instrument families and their synchronization relations.

10.1.1 Generator/Resonator Synchronization

Musical instruments are most often described as coupled systems of a generator and a resonator. In wind instruments, the labium or the reed are the generator and the tube is the resonator. With string instruments, the string generates the sound while the wooden box is the resonator. Still it is interesting to note that sometimes the generator determines the played pitch, sometimes it is the resonator. With wind instruments, generally the resonator determines the pitch so by using certain fingerings the musician can play different notes. With stringed instruments, it is the generator determining the played pitch. But not only does the role of these parts differ between instrument families, other conditions also have an effect. The overblowing of trumpets or horns is performed using a combination of generator and resonator. Reed organ pipes can be tuned by adjusting the reed but also by adjusting the tube length. When examined in greater detail, most instruments show very complex behavior when played in extreme regions as often performed in free jazz, free improvised music, or contemporary classical music using extended techniques. For example, saxophones or clarinets may play several notes at the same time using multiphonics. With the violin string, one may also play subharmonics, where the pitch is no longer determined by the string length but by the bowing force; also this mechanism may lead to bifurcations and play scratchy sounds easily.

Table 10.1 gives a rough overview of the generator/resonator behavior for different instrument families and which determines the pitch. We will discuss in more detail below the reasons for this behavior which is caused by nonlinearities in the generator and a complex coupling to the resonator.

10.1.2 Synchronizing Conditions

In synergetics several reasons are known why one oscillator is taking over – or is synchronizing – the other oscillator, and so is forcing the second oscillator into the frequencies of the first one [10.1, 2]. The simplest is a difference in system damping. If two oscillators f and g are driven by themselves with damping α and β, respectively, and both are coupled with coupling constants c_1 and c_2 like

$$\dot{f} = -\alpha f + c_1 f\, g\,, \tag{10.1}$$

$$\dot{g} = -\beta g + c_2 f^2\,, \tag{10.2}$$

then if the damping of g is much stronger than that of f like

$$\beta \ll \alpha\,, \tag{10.3}$$

Table 10.1 Generator/resonator model of musical instruments. With some instrument families the generator determines the played pitches while with others it is the resonator

Instrument	Generator	Resonator	Part determining instruments pitch
Guitar	Strings	Body	Generator
Violin	Strings	Body	Generator
Singing voice	Vocal folds	Vocal tract	Generator
Saxophone	Reed	Air column	Resonator
Trumpet	Lips	Air column	Resonator
Percussion	Mallet	Body	Resonator
Organ	Labium	Air column	Generator/resonator
Violin bowing	Bow	String	Generator/resonator
Vocal folds	Lung pressure	Vocal folds	Generator/resonator

the coupled oscillator equations simplify to

$$\dot{f} = -\alpha f + c_1 f\, g\,,$$
$$\dot{g} = c_2 f^2\,.$$

So g depends only on f and no longer upon itself and therefore g is synchronized to f, where f determines the temporal development of g.

This general finding was already discussed by *Aschoff* [10.3] for the saxophone. He finds that the saxophone tube must take over the frequency of the reed because the reed is much heavier damped than is the tube. In the following discussion of several instruments in terms of their driving mechanism, the reason for synchronization of the oscillators will become clearer and in the end we will collect the different reasons.

10.2 Models of the Singing Voice

The singing voice consists of a nonlinear driving mechanism, the vocal folds and a vocal tract filtering the impulse train produced by the larynx to vowels or transient sounds. The vocal tract is basically passive and linear, while the vocal folds are highly nonlinear. Because of this nonlinearity they are able to produce periodic sounds with a harmonic overtone structure at normal vocal fold tension and air pressure produced by the lungs. However, when leaving this normal range the vocal folds may produce bifurcations, sounds with more than one harmonic overtone series which is known as rough voice and is often desired in singing styles like Blues or Rock. The folds may also produce subharmonics used in throat singing of ethnic groups for example those in Tuva, Mongolia, or Tibet. With very low lung air pressure the folds produce noise which can still be shaped to arrive at a whispering voice.

Describing the voice in general is beyond the scope of this chapter where we want to focus on the nonlinear behavior of the vocal tract and examine how such a nonlinear system can arrive at producing a perfect harmonic overtone series. For more detailed descriptions of the singing voice see [10.4–6]. Many papers experimentally examining the vocal folds using glottography [10.7, 8] discuss them with respect to their nonlinear nature. Also fractal dimensions have been calculated to estimate the chaoticity of vocal fold vibrations, for example in [10.9]. The role of vocal fold nonlinear expressiveness is discussed in [10.10] for contemporary vocal music.

Understanding vocal fold vibration is a matter of ongoing debate. Still, over the decades of research many models have been proposed that each have pros and cons.

10.2.1 Bernoulli Effect

The first approximation used in understanding the vocal folds is the Bernoulli effect. Here a pressure gradient appears when two flows of air next to each other have different traveling speed. An example is air traveling over and under an air plane wing. The air takes longer to travel over the wing than under it. Therefore, the air above the wing is stretched or decompressed because of the increased speed, while the air below it is more compressed because of its lower speed. So below the wing there is more pressure than above the wing. As pressure is a scalar acting in all directions both pressures act on the wing. As the pressure below the wing is higher than the one above it the wing is pushed upwards and so is the plane.

The same principle holds for the vocal folds. If an air flow produced by the lungs is traveling through the folds the air stream becomes faster as the folds have a smaller area for the air to flow through. Therefore, there is higher pressure at the position of the folds than

above or below them. This causes the folds to open releasing the lung pressure. As the cycle proceeds the upper part of the folds are wide open while the lower part is narrower. Again according to the Bernoulli effect of over- and underpressure acting on the folds, the air now at the upper part of the folds have an underpressure while the air at its lower part have an overpressure. Therefore, the upper part of the lips are forced inwards to close the upper lip part again. After closing of the upper lip part the fold is again in its initial position and one cycle of oscillation has occurred. The next cycle will be started in the same way as the previous one and therefore the folds are in steady oscillation.

This very basic understanding of vocal fold vibration is still found in the more sophisticated models discussed below. Improved models are needed as this first approximation does not explain many more complex vocal fold phenomena. The most striking one is the sudden onset of a periodic motion when increasing lung pressure or lip tension. Also, a sudden change into falsetto singing is not explained by this model alone. Additionally, rough voice, subharmonics, or other bifurcations can not be explained by the Bernoulli effect alone.

10.2.2 Two-Mass Model

On the basis of the assumption of understanding the vocal folds only when splitting them into an upper and a lower part, the most prominent model in the literature today is the two-mass model initially proposed by [10.11], which has been implemented in many variations [10.12–14]. Figure 10.1 shows the basic idea. Both sides of the vocal folds are modeled with two masses, one upper and one lower one. The upper is supraglottal while the lower is subglottal. This is in accordance with the Bernoulli effect discussed above, where an over- or underpressure effect appears when the flow changes its speed; then the upper or lower fold has a different opening area. As the folds are under pressure the upper and lower folds are modeled by two strings with displacement x_1 and x_2. This is an enlargement of the Bernoulli model as it includes the vocal fold tension by adding string constants d_1 and d_2 and fold masses m_1 and m_2. When assuming both fold parts to be springs we can introduce a differential equation as follows

$$\frac{\partial^2 x_1}{\partial t^2} = \frac{1}{m_1}\left(P_1 L\, d_1 - k_1 x_1 - r_1 \frac{\partial x_1}{\partial t} - k_c(x_1 - x_2)\right), \tag{10.4}$$

$$\frac{\partial^2 x_2}{\partial t^2} = \frac{1}{m_2}\left(P_2 L\, d_2 - k_2 x_2 - r_2 \frac{\partial x_2}{\partial t} - k_c(x_2 - x_1)\right). \tag{10.5}$$

Here P is the subglottal pressure, L the folds length, and r_1 and r_2 are damping constants of the folds. Both folds are coupled with a coupling term k_c.

The model is able to produce a periodic oscillation of the folds as with normal singing. However, when varying the parameters bifurcations appear. Figure 10.2 shows bifurcations on changing the coupling constant k_c of the model, thus displaying a traditional bifurcation scenario as known from the logistic map [10.2]. The Lyapunov exponents shown in Fig. 10.2b are an indication of nonlinearities in the system and display the onset and amount of chaoticity in the regime.

Also falsetto voice transition can be explained using the two-mass model [10.15, 16]. Here the transition from chest to falsetto voice is a hysteresis loop as shown in Fig. 10.3. It takes more pressure to change from the chest to the falsetto voice then when coming back

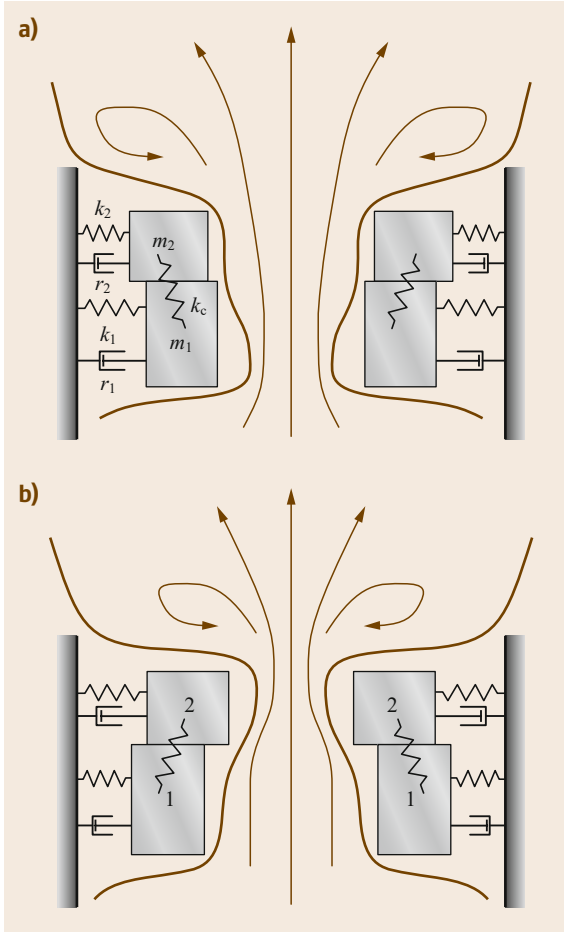


Fig. 10.1a,b Two-mass model of the vocal folds shown at two turning points of one oscillation cycle: **(a)** folds are more open at the supraglottal side, **(b)** folds are more open at the subglottal side (after [10.14])

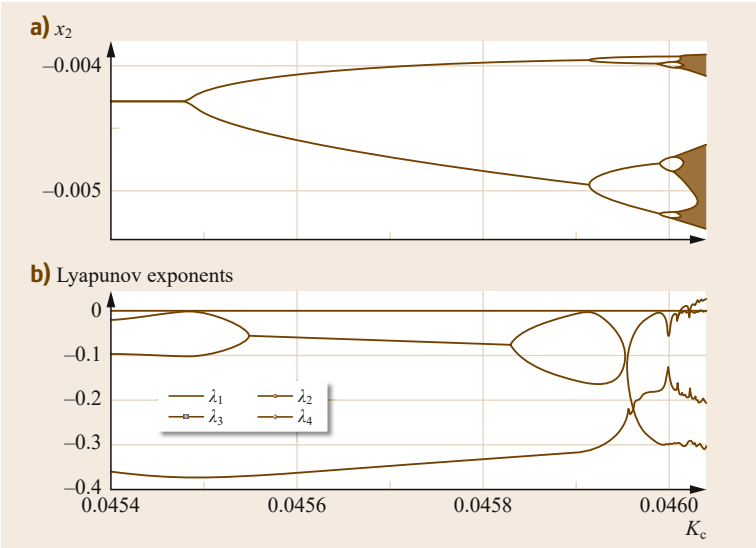


Fig. 10.2a,b Bifurcation scenario of a two-mass model of the vocal folds when varying the coupling constant between the vocal folds. From a certain region bifurcations appear and lead to an unstable rough voice. **(a)** Period-doubling at $K_c = 0.04548$ into two periodicities, followed by bifurcations with increasing K_c. At $K_c = 0.04611$ the system becomes chaotic. **(b)** Lyapunov exponents λ_1–λ_4 (*top* to *bottom* in **(b)**) calculated from **(a)**. $\lambda_1 = 0$ (bifurcated but non-chaotic oscillation) until $K_c = 0.04611$, where λ_1 becomes positive, indicating chaotic motion. $\lambda_2 = 0$ from $K_c > 0.04611$ (after [10.14])

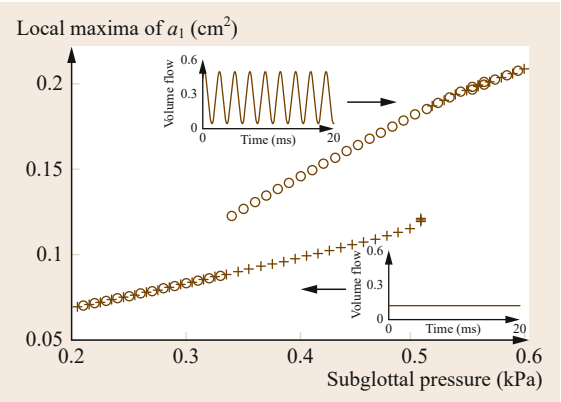


Fig. 10.3 Hysteresis loop between chest and falsetto voice. When changing from chest to falsetto a higher pressure is needed than when changing back from falsetto to chest voice

from falsetto to chest voice. This is a classical behavior of self-organizing systems which try to stay in an established regime as long as possible, stretching the boundaries of sudden phase changes to the other regime as much as possible.

So the two-mass model is able to determine sudden phase changes like tone onset and the change from chest to falsetto singing. It can also model bifurcations when the vocal folds have different tensions as discussed below.

10.2.3 Mucosal Wave Model

The two-mass model explains most of the vocal fold behavior. Still, it is a purely iterative model to which no analytical solution is known. To get an analytical understanding *Titze* proposed a mucosal wave model [10.5], which reasons that the vocal folds are basically a van der Pol oscillator like

$$m\frac{\partial^2 u}{\partial t^2} + (R - R_g)\frac{\partial u}{\partial t} + Ku = 0\,, \tag{10.6}$$

where u is the displacement of the folds perpendicular to the air stream P_s. Then m is the vocal fold mass, K its stiffness, and R its internal damping. Here P_s is acting on the folds not directly but through the relation of the open area a_1 and a_2

$$a_1 = 2L\left(u_0 + u_+\tau\frac{\partial u}{\partial t}\right) \tag{10.7}$$

$$a_2 = 2L\left(u_0 + u_-\tau\frac{\partial u}{\partial t}\right), \tag{10.8}$$

of the supraglottal and subglottal fold side, respectively. Then the pressure acting on the folds is built from the lung pressure P_s and the areas like

$$P_g = P_s\left(1 - \frac{a_2}{a_1}\right), \tag{10.9}$$

where then

$$R_g = \frac{2\tau P_s}{u_0}\,. \tag{10.10}$$

From the van der Pol oscillator an analytical solution is known with eigenvalues

$$\lambda_{1,2} = -\frac{R - R_g}{2m} \pm \sqrt{\left(\frac{R - R_g}{2m}\right)^2 - \frac{K}{m}}\,. \tag{10.11}$$

These eigenvalues have imaginary parts that are complex conjugates and therefore the attractor is a limit cycle if the real part is zero, which is the case for $R = R_g$. If $R \neq R_g$, the attractor is a focus which is either stable for $R > R_g$ or unstable for $R < R_g$. In the latter case, the driving of the folds by the air stream is greater than the internal damping of the folds and therefore the folds start moving as an unstable focus increasing the vibration amplitude. We can therefore calculate the onset pressure as

$$P_{\text{onset}} = \frac{Rx_0}{2\tau} \,. \tag{10.12}$$

10.2.4 Hopf Bifurcation

This mucosal model shows us analytically the reason for tone onset and gives estimates when this will appear. Still it arrives at an unstable limit cycle which theoretically could build up forever leading to unphysically large amplitudes. This is avoided by a Hopf bifurcation [10.17–19]. The basic idea of this formulation is to solve the equations analytically by linearizing the nonlinear equation system using a Taylor series in the neighborhood of the attractor point [10.2]. Mathematically, this is beyond the scope of this paper. Still the basic equation system is

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \alpha(P_s) & -\omega(P_s) \\ \omega(P_s) & \alpha(P_s) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} f(x,y) \\ g(x,y) \end{pmatrix} , \tag{10.13}$$

where P_s is the lung pressure again. On the rhs the first term is linear and the second term consists of all nonlinear terms which we neglect in the neighborhood of the attractor. The attractor has two coordinates x and y. A simpler way to understand the system analytically would be to use polar coordinates (s, θ) of radius and angle, respectively. Then a tone onset would be the transition of $s = 0$ to $s > 0$ and vice versa for a tone offset. After transformation of the equation system into polar coordinates and some math we arrive at

$$\dot{s} = dP_s s + as^3 = f(s) \tag{10.14}$$

$$\dot{\theta} = \omega + cP_s + bs^2 = g(s) \,, \tag{10.15}$$

with $d = \partial\alpha/\partial P_s(0)$ and $c = \partial\omega/\partial P_s(0)$. After examining the eigenvalues of the system four cases are possible:

- $d > 0, a < 0$: supercritical stable Hopf bifurcation
- $d < 0, a > 0$: supercritical unstable Hopf bifurcation
- $d < 0, a < 0$: subcritical stable Hopf bifurcation
- $d > 0, a > 0$: subcritical unstable Hopf bifurcation.

Bifurcation here means the onset of an oscillation so the transition from a point attractor to a limit cycle. In the terminology of nonlinear dynamics this is justified as an oscillation is expected with differential equations of second order. However, the Hopf bifurcation equation system is one of first order and therefore does not have an oscillation as a solution but an exponentiall decay. So an oscillation is not expected. Still it appears as a bifurcation of the system. As this bifurcation is appearing at a certain pressure level it indicates a sudden phase change from silence to a musical tone.

Lucero finds the vocal folds to be a subcritical unstable limit cycle. This explains tone onset and the hysteresis loop of on- and offset already discussed above. Figure 10.4 shows the results of these calculations where the pressure for tone onset needs to be higher than the one of tone offset as known experimentally.

The only problem here is that the bifurcation, the onset of the oscillation, in this system is unstable. This means that only a small change in singing would make it blow up to unreasonable large amplitudes. Therefore, one need to stabilize it by limiting the lung pressure.

10.2.5 Biphonation and Subharmonics

The singing voice has many more articulations than only a perfect harmonic pitch. Rough voice is well known to such musical styles as Blues, Rock, Metal, or Grunge. Such rough voice still has a periodicity and therefore a pitch, now with additional inharmonic components. To mention only a few, the throat

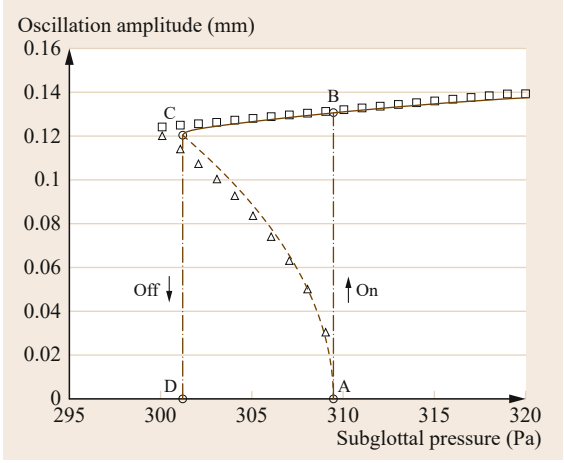


Fig. 10.4 Hysteresis loop of subglottal pressure and Hopf bifurcation limit cycle oscillation between tone onset at *A–B* needing a higher pressure than the tone breakdown at *C–D*, calculated from numerical simulation. (Reprinted with permission of [10.18]. Copyright 1999, Acoustical Society of America)

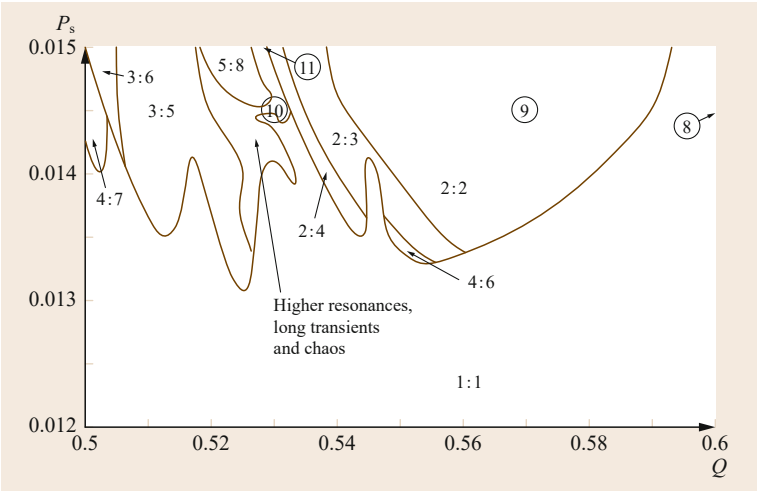


Fig. 10.5 Bifurcation diagram for vocal fold tension relation Q versus subglottal pressure P_s. The ratios are those of the maximum amplitudes for the left and right vocal fold. (Reprinted with permission of [10.21]. Copyright 1995, Acoustical Society of America)

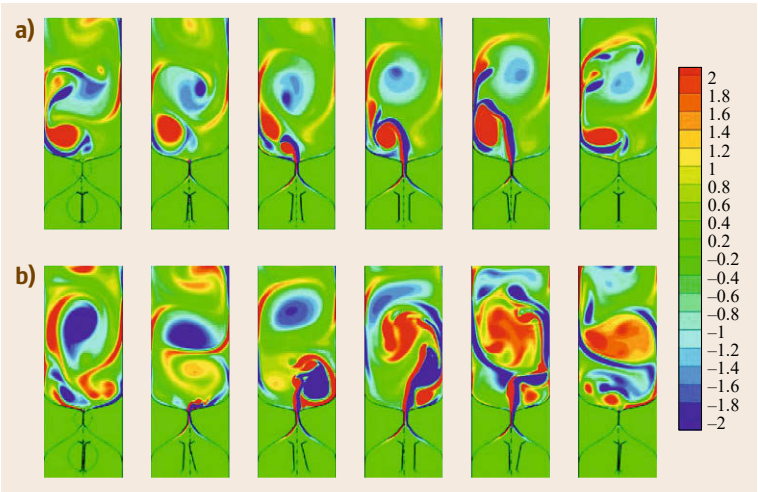


Fig. 10.6a,b Adjacent time points of a finite-element model of vocal fold flow dynamics for **(a)** a normal voice and **(b)** a voice with uneven fold tension displaying more turbulent motion with the altered folds

singing of Tuva or Tibet uses subharmonics where the singer normally performs an octave below his normal singing voice as discussed e.g., with *kagyraa* singing of Tuva [10.20]. Paul Pena, a Blues musician demonstrates the similarity of Tuvan subharmonic singing and blues howling style (CD: Paul Pena: Ghengis Blues. Six Degrees 2000). *Lee* compare Korean and Western singing in this way [10.22]. The role of vocal fold nonlinear expressiveness is discussed in [10.10] for contemporary vocal music.

These phenomena have been extensively studied using two-mass and related models [10.21, 23, 24] as well as experimentally [10.25].

The bifurcation diagram shown in Fig. 10.5 shows regions within which a stable oscillation of simple or complex integer oscillation ratios appear. Normal singing would be 1:1, throat singing would be 1:2 as an octave below. With 3:2 two tones would be present within one singer a musical fifth apart etc. The regions depend on the subglottal pressure and the relation between the vocal fold tension. These regions show sudden phase changes to other regions and are therefore stable over some parameter variations of pressure and tension relation.

A Finite-Element Model (FEM) of the air stream of vocal fold vibration [10.24] shows the turbulence appearing over the folds for two cases of symmetrical folds (both with the same tension) and asymmetrical folds. Clearly the turbulence above the folds is getting more complex with the asymmetrical case (Fig. 10.6).

In the model of *Xue* [10.24] it is interesting to see that the turbulent flow above the folds looks very much like the flow of an air stream entering a saxophone or clarinet mouthpiece. As discussed in the wind instrument part of this chapter, a laminar air stream entering a larger cavity becomes unstable and builds a first large

vortex which is the main sound source. So the next step in building a model for the singing voice could point in this direction of modeling the air flow in terms of the sound production and transition and possible reflections in the vocal tract.

10.2.6 Synchronization with Vocal Tract

Above we concentrated on the self-oscillating system of the vocal folds without taking the vocal tract or subglottal areas into account. We might simply consider the vocal tract as a resonator and not go into further detail here. Still strictly speaking the vocal tract is a tube which reflects the sound produced by the vocal folds and therefore is acting back on them. The mechanisms of a tube reacting to a fluid-driven self-sustained oscillation is discussed in more detail below with the organ. Also the organ labium is not a vocal fold. Comparing the singing voice with a saxophone we find a considerable difference. With the saxophone, the tube (the resonator) forces the reed (the generator) to go with its frequencies, as the tube is much less damped than is the reed. With the singing voice the situation is vice versa. The vocal tract (the resonator) is forced into the frequencies of the vocal folds (the generator), although again the resonator is much less damped. Still in a normal singing register the vocal tract resonance is normally above the periodicity of the vocal folds. This is reasonable as the vocal tract is shaping the voice in a formant way, enhancing higher spectral regions above the fundamental periodicity of the fold vibrations. So here we find that although the folds are more damped than the resonator, the generator takes over the vibration because its pitch is much lower than that of the vocal tract.

10.3 Harmonic Synchronization in Wind Instruments

When playing a saxophone, trumpet, or organ pipe under normal playing conditions, the played notes show a nearly perfect harmonic overtone spectrum. This is so familiar that we take this for granted. Still, when examining these instruments closer this harmonic series should not be expected at all as the instruments are highly nonlinear, show turbulence flow, uneven bores in flues and clarinets, shaped tubes like in trumpets or saxophones, a flared bell at the tube end, and other features that distort the simple behavior of a mathematical tube with a theoretically harmonic spectrum.

And indeed when overblowing a horn, trumpet, or saxophone the overtone series is never harmonic, but shows large deviations up to a third above or below the expected partial frequency. This is one of the major problems of instrument builders. Flute makers use a complex bore diameter shape over the flute length to account for the harmonic overtone series as well as for the amplitudes of the partials [10.26]. So when enlarging the bore at one point one overtone may fit better while at the same time another is deviating more. Additionally, the size and length of the finger holes change the length of the tube in a complex way and again lead to distortions of the harmonic series [10.27]. Here, over the centuries the builders established rules of thumb to arrive as closely as possible at a perfect series. The rest need to be adjusted by the player who can change playing pressure, lip tension, or the distance between the players lips and a labium to intonate the instrument correctly. With instruments of lesser quality it can be a pretty demanding task and in some cases a desired pitch cannot be reached at all.

With horns the flaring of the bell not only determines the timbre of the tone by enhancing higher partials with stronger flaring, larger bells lead to a frequency dependence of the tube end-correction. So for each frequency the tube has a considerably different length and therefore the harmonic series is no longer harmonic at all [10.28]. Although theoretically the Bessel horn, a horn with a bell flaring according to a Bessel function would again meet a perfect spectrum, when measuring the bells of existing instruments this Bessel shape is nearly never found.

Finally, the driving mechanism, the mouthpiece with reed or double reed or cavity as with brass instruments, or the labium shape and cavity with organ pipes or flues, are highly complex with a nonlinear driving mechanism. Here again we would expect a different theoretical tube end which is frequency dependent.

Still with all wind instruments within a normal playing regime the tones produced show a nearly perfect harmonic overtone series mostly far from the overtones which the instruments produce when they are overblown. This behavior is so robust that we would expect a simple mechanism to be behind it. Indeed, it is not possible to shape the overtone spectrum of one tone to a desired spectrum by changing the geometry of the instrument. At first sight this sounds very much like a synchronization or self-organization system where due to a nonlinearity in the driving mechanism as discussed below a very simple output is produced, the harmonic spectrum. So we need to discuss these instruments in detail to see if this mechanism exists and how it might work.

The fact that wind instruments leave this harmonic series when not played with enough blowing pressure, which produces noise, or with complex fingerings, which produce multiphonics, is again pointing in the same direction as a self-organized system would perform a simple output only for a certain parameter range. It then leaves the simple regime suddenly, which again is a property of wind instruments which suddenly change from noise to a tone at a certain lung pressure threshold. Also bifurcations would be expected from such a system which would be present with multiphonics played on nearly all wind instruments in contemporary, free improvised or modern classical music or free jazz. Here tutorial material teach hundreds of possible multiphonics with complex fingerings like with the clarinet [10.29].

This chapter reasons that this synchronization phenomenon arriving at a perfect harmonic series can be understood when examining the flow in the instruments, its change from laminar into turbulent flow, and its interactions with the tube and a possible reed.

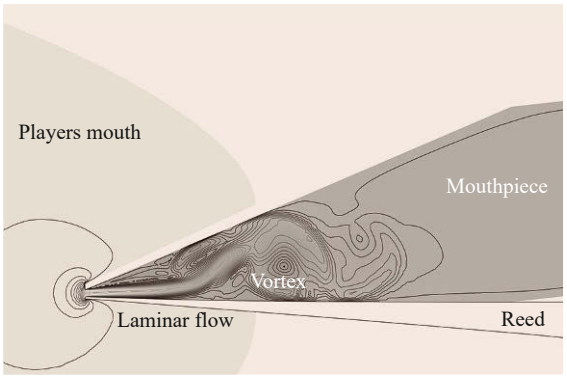
10.3.1 Navier–Stokes Flow Model

The Navier–Stokes flow equation can be used to understand the flow that appears when blowing into a wind instrument [10.30],

$$\partial_t u_i + u_j \partial_j u_i = -\frac{1}{\rho} \partial_i p + \nu \nabla^2 u_i$$
$$\text{with} \quad i = 1, 2, 3 \,. \tag{10.16}$$

It states that there is a balance of accelerations when comparing the flow velocity u and its interaction in all three directions $i = 1, 2, 3$ with the pressure p. So when the velocity changes, either in time as stated by the term $\partial_t u_i$ or in space as stated by the term $u_j \partial_j u_i$, it needs to be balanced either by a spatial change of pressure $-1/\rho \, \partial_i p$ at this point or by a viscous damping $\nu \nabla^2 u_i$ with viscosity constant ν.

Let us consider this in action in a saxophone mouthpiece as shown in Fig. 10.7. The pressure in a mouth causes a laminar flow into the mouthpiece which then becomes a large vortex. This vortex is caused by the term $u_j \partial_j u_i$, which describes a change of flow velocity between the directions. So as the laminar air stream is small it interacts with the air around it and as the air stream is freely moving in the cavity only slight differences between these interactions above and below the stream will drive it in one direction more than the other. This change of direction is exponentially increasing leading the stream to turn around itself by 360° building a vortex or eddy as seen in the figure.



10.3.2 First Vortex and Sound Production

This first vortex building is crucial for the synchronization in some ways. First, it is only caused by the nonlinear term in the Navier–Stokes equation $u_j \partial_j u_i$ and is therefore the nonlinearity in the system. If the equation was linear in this respect, so if the derivative of the flow velocity is not multiplied by the velocity of the other directions, the flow would be straight and not build a vortex.

Secondly, the vortex means that the flow does not move further in its original direction, it stops by building a vortex at a certain point. This means two things, first the flow is nearly completely dissipated within the mouthpiece and all energy traveling further into the tube is almost completely acoustic flow [10.31, 32]. Further, this also means that there is a strong spatial gradient at the point of the vortex. As discussed above, the Navier–Stokes equation is a balance of flow and pressure change. So when the flow changes strongly at a point in space the pressure changes strongly too at this point, a pressure gradient is built there. This pressure gradient is the sound source, so here the transfer of flow into acoustic energy is taking place.

In terms of the labium, the pressure gradient produced by the first vortex is experimentally shown in the case of a self-sustained oscillation at a labium in Fig. 10.8 [10.33, 34]. Here the three columns show the flow (column c)), the pressure at the upper labium side (column a)), and at the lower labium side (column b)) for adjacent time points from top to bottom over one cycle of oscillation. The sketch d) holds as a reference for all pressure plots displayed in columns a) and b). The pressure plots start at the labium tip on the left end of sketch d) and display the pressure at the different time

Fig. 10.7 Turbulent flow in a mouthpiece when driven by the pressure of the mouth. The reed is driven by the pressure difference between these two cavities. A large vortex is formed in the mouthpiece as the laminar flow from the mouth enters a large cavity (the mouthpiece) and therefore becomes unstable ◄

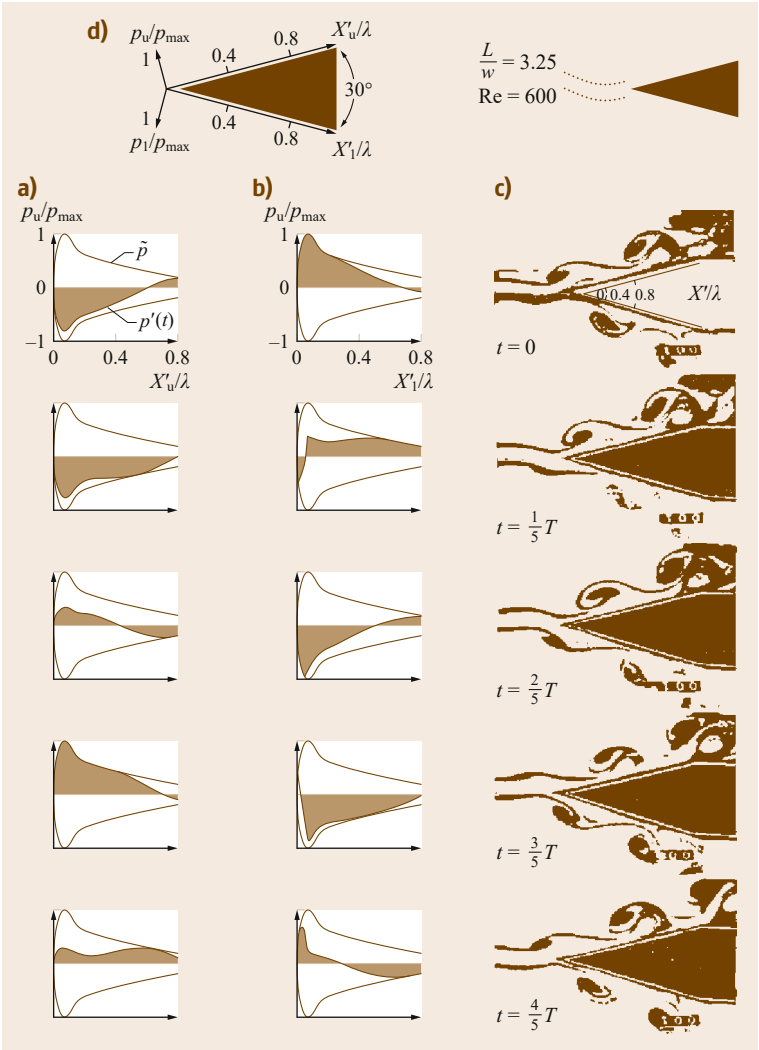


Fig. 10.8 Column **(a)** and **(b)**: pressure distribution $p'(t)$ (normalized by maximum pressure p_{max}) and column **(c)**: flow visulaization of the vortex street for pressure p_u. p_u is shown **(a)** above and **(b)** below the labium. The labium is sketched in **(d)** with labium tip at the left end and labium length as unit of oscillation wavelength λ like X'_u/λ for a flow of Reynolds number Re = 600. The rows in columns **(a–c)** are shown at five adjacent time points t over one cycle of a periodicity T (after [10.33])

points (*black area*) and the maximum pressure envelope over one period cycle (*black line*).

The point of interest for our reasoning is the strong pressure gradient at the tip of the labium shown in the pressure envelopes of the plots. This is the point where the large first vortex appears around the labium tip. So also for the labium instrument the first vortex is the point of a pressure gradient.

Now this gradient is changing fast in time according to the oscillation periodicity. So at this point at the labium, or at the vortex in the saxophone mouthpiece, a pressure impulse is produced which has a certain shape according to the geometry of the mouthpiece or labium cavity, the oscillation frequency, and the blowing pressure.

Still when using a Navier–Stokes model as discussed above we do not have any sound yet as this model is incompressible. Sound means a density change and therefore including a compressible Navier–Stokes model with a density ϱ varying over time and space like

$$\frac{\partial \varrho}{\partial t} = -\nabla(\varrho u_j) \,. \tag{10.17}$$

Using an incompressible Navier–Stokes model artificially skips the acoustics, still it is much easier to handle and therefore suitable as a first approximation. Nevertheless, the transfer from flow into acoustics is the core idea of wind instruments. It appears that the first pressure gradient at the vortex cannot move further as an acoustical pressure when using the incompressible Navier–Stokes equation as we artificially prohibit the acoustics. Still, when adding compressibility we allow the pressure gradient to continue as acoustical sound where the impulse is then traveling along the tube, is

reflected at the tube end, and then returns to the mouthpiece or labium again.

10.3.3 Phase Disturbance and Turbulent Damping

So now we have a pressure gradient or impulse traveling down the tube of the instrument. Indeed wind instruments are known to have a time series which consists of pressure impulses and are often modeled by synthesizers using rectangular-shaped oscillators. Still, as discussed above the pulse shape is very distorted when traveling down the tube. This is caused by a complex bore profile with flues, the tube shapes of saxophones or trumpets, the bell flaring at the tube end with frequency-dependent end-correction, and other aspects like finger hole size and tube thickness.

An interesting aspect is the shape of small areas of the instrument which may cause turbulent damping again because of the thin tube diameters. The acoustic flow can be fully described by the compressible Navier–Stokes equation, therefore it still remains a flow of pressure, velocity, and density fluctuations. Therefore, the acoustic pressure is interacting with the walls of the tube causing considerable damping within the tube which is affecting the playability of the instruments as well as their acoustic properties. This can only be explained by a tremendous damping caused already by only small turbulence production.

Such a case was found with bassoons in their s-shaped tube section [10.35]. Bassoon players need up to 15 kPa of lung pressure to drive their instruments, which physiologically is a challenge for the players' lungs. Therefore, reducing this pressure is highly welcome. When examining the acoustic flow through the bassoon tube, it appeared that for acoustic flow through s-shaped tubes where the pressure needs to change direction, turbulence appears in the tube which increases the onset threshold for playing on the instruments. By changing the geometry of the shape the onset playing pressure could be decreased.

A theoretical model for understanding turbulent damping is the k-ϵ Reynolds-Averaged Navier–Stokes model [10.30]. This was applied to a flute and compared to a normal incompressible Navier–Stokes model to account for the damping appearing through turbulence when blowing into the flute [10.36]. The efficiency of a flute is low where only about 3% of blow energy is entering the tube as acoustic energy. Then again only about 2% of these 3% is transmitted into the room as sound pressure making the flute a very low sounding instrument [10.37]. In a Finite-Element Reynolds-Averaged Navier–Stokes model it could be shown that when modeling the flute in-blow without taking turbulence into consideration the amount of energy entering the flute tube is about 50% while when modeling the instrument with turbulence it is only about 3% leading to a realistic model.

Therefore, most damping in the tube of wind instruments seems to be caused by turbulence, although there the acoustic energy traveling through the tube is transferred into flow again and then dissipated. So strictly speaking there is no fundamental difference between flow and acoustics and the compressible Navier–Stokes equation is able to account for both as well as for their coupling.

10.3.4 Triggering of New Impulse and Phase Alignment

So when the pressure impulse produced at the labium or in the mouthpiece has traveled through the tube and returns to the generator its phases are greatly disturbed. If this impulse would simply be reflected at the generator and travel back into the tube the resulting overtone spectrum of a tone would not be harmonic at all but would reflect the distorted series of overblown tones on the instrument. So a synchronization of the phases of the impulse needs to occur at the labium or in the mouthpiece to align these phases again leading to a harmonic series. The literature suggests two main approaches to this synchronization.

Models like van der Pol oscillators, as discussed in some detail in the section on the vocal tract (Sect. 10.2.3), are suitable for synchronizing organ pipes [10.38, 39]. These were applied to a slightly different problem of synchronizing two pipes. The synchronization occurs due to an energy transfer between the frequencies so that slower or decayed phases are pushed and faster phases are decayed in such a way that the phases are aligned again. This is reasonable in the case of two organ pipes coupling through space and is an effect that is known by organ builders and used to increase tuning stability of organs in regard to temperature fluctuations in churches due to weather conditions, which would otherwise detune the organ considerably.

The other model assumes that the returning impulse only triggers a new impulse, which is built in the generator [10.36]. So in saxophones the returning impulse moves the reed up or down, depending on whether the returning pressure is an under- or overpressure. Then the inflow of air into the mouthpiece changes and therefore the first vortex changes, too. This vortex then produces a new pressure impulse from scratch without taking the old one into account. As this new impulse is built in exactly the same way as the last one the synchronization is perfect and this new impulse then travels

down the tube again. So in this model the disturbance of the returning impulse can have any shape, as the returning impulse is either damped in the mouthpiece or radiated into the surrounding air via the labium. It need not be phase aligned again to arrive at a harmonic overtone series. It only triggers a new impulse and therefore can have any kind of distortion.

The trigger model is supported by the reed system to be a pure valve system. So although the impulse sent out from the mouthpiece has a rich harmonic overtone series, the reed moves almost completely in a sinusoidal manner without any harmonic overtones. The reed behaves like a simple valve opening or closing and is not contributing any harmonics. Therefore, the returning impulse can only trigger the fundamental periodicity as the reed is not reacting to the higher harmonics of the returning impulse.

Also the robustness and stability of the synchronization can be explained by the trigger model more easily. Aligning the phases of highly distorted returning impulses is complex and would be much more unstable and unreliable. Indeed, under normal playing conditions all wind instruments show this synchronization of partials almost completely, suggesting a very simple reason is behind this synchronization.

So when considering the reasons why, in a system of coupled oscillators, one oscillator takes over the other oscillator, in terms of its frequencies, with labium instruments it appears that the strong turbulent damping at the labium leads to the labium oscillator being heavily damped, while the tube is only weakly damped, and therefore the tube takes over the labium oscillation. With reed instruments the reed is much more damped than the tube and therefore here it is the reed taking over as already discussed by *Aschoff* [10.3].

10.3.5 Synchronization Condition

So we might generally say that the slaving of the generator by the resonator is due to the resonator being much less damped than the generator. Still, as we have found with the singing voice this might not always be the case and in this system, although the generator is damped more than the resonator, the generator prevails as its pitch is much lower than that of the resonator. A saxophone reed has a lowest eigenfrequency of about 1 kHz when attached to the mouthpiece. While playing this lowest eigenfrequency doubles to about 3 kHz [10.36] and therefore is greatly above the resonator frequency. Therefore, here both conditions, that of a heavier damping as well as a higher pitch, leads to a slaving of the reed by the tube.

With organ pipes the self-sustained oscillation of the labium is tuned to about the resonance frequency of the tube. Therefore, here the situation is not as simple and although we have perfect synchronization between the two it is not straightforward to say who synchronizes whom to what extent.

10.4 Violin Bow–String Interaction

Many musical instruments are bowed, like the violin, viola, cello, or double bass, but also the Indonesian *rebab*, the Chinese *erhu*, the Indian *esjay*, and hundreds of other instruments around the world. There are also some rare friction instruments which are played by rubbing over a surface, like the bowed glasses or the *lounuet*, a rubbed wooden block found in New Ireland, or bamboo tube zithers like the *kting ga-un* from Thailand [10.40]. The mechanism of bowing is a highly nonlinear one where the bow or a rubbing finger is alternatively sticking and slipping over the string or over a surface like glass or wood. The nonlinearity in this interaction is the strong and sudden change in the function of force of the bow acting on the string. A linear force function would only linearly change with bowing pressure where doubling the pressure would double the force of the bow acting upon the string. However, when bowing the force changes from being very strong during the sticking phase to a rather light force during the slipping phase, where the bow is only gliding over the string.

The system also shows all aspects of a self-organizing system [10.36, 41]. First, over a wide bowing pressure and velocity parameter range, it stays in the same regime, a periodic Helmholtz or sawtooth motion where the periodicity is determined by the string length. Adding to the string periodicity is a small effect from the bowing pressure, where higher pressures decrease the pitch slightly, which is known as the flattening effect. The system also shows sudden phase changes at certain pressure values either entering a subharmonic or a bifurcation regime with scratchy and noise sounds. In the subharmonic regime, the pitch is determined no longer by the string length but by the playing pressure, which in terms of synergetics can be described as a change of the ordering parameter from string length to bowing pressure. In the bifurcation regime, multiple periodicities appear which are again stable over a certain parameter range only to then suddenly shift into another regime with again inharmonic partials. Finally, the shifts between the regimes, like the Helmholtz and the subharmonic or the bifurcation regimes, show hystere-

sis loops where entering one regime needs a different playing pressure to coming back to it.

Historically, bowing was first prominently considered by *Raman* [10.42], who argued that when we hear a violin tone of constant amplitude there needs to be an energy balance between input energy supplied by the bow and sound energy radiated into space. When calculating the energies he used the velocities of bowing and string and therefore many later considerations took string and bow velocity as their starting point [10.43–46].

Another approach is to consider displacement of strings and bow first and discuss the system in a time-dependent manner, which was proposed by *Güth* [10.47]. He is especially concerned with initial transients and describes them as small initial impulses traveling along the string when the bow is not in a steady state but still struggling to achieve a regular periodicity. Models using finite-differences to model the system have the same argument because such models include displacement, velocity, and acceleration in a time-dependent manner. Theoretically there can be no fundamental difference between the models as displacement and velocity are directly connected via integration and differentiation. Still both viewpoints, that of velocity and that of displacement were leading to different models explaining the stick/slip interaction and the conditions leading to a stable sawtooth motion.

10.4.1 Bowing Force Model

Figure 10.9 shows the force function f over the string velocity v [10.48, 49]. The bow velocity in this plot is where the curve has its infinite slope and crosses the horizontal axis. If the bow is sticking to the string both velocities, that of the bow and that of the string, are the same and therefore the curve is in the region of the infinite slope. The slipping phase is displayed as the region left of the infinite slope. There the string velocity is less than the bow velocity. The three lines with constant slopes and intersections v_h, v_r, and v_c point to three different relative bow velocities. v_r and v_c are both extremes crossing $f(v)$ twice, v_r relates to the maximum force point. v_c is parallel to v_r, touching the rising curve of the force function only once. Conversely, v_h crosses $f(v)$ three times, indicated by the points. The region between v_r and v_c is a region of arbitrariness, where the system is still stable and will continue with its basic motion. If this region is left, either above v_r or below v_c, the stability is left and the system will no longer show sawtooth or Helmholtz motion. In other words, the normal Helmholtz regime of playing appears over a larger region of bowing pressure, not just with only one single bowing pressure value. This is crucial for playing violin, as only the Helmholtz regime produces a harmonic pitch depending upon the fingering enabling normal violin playing.

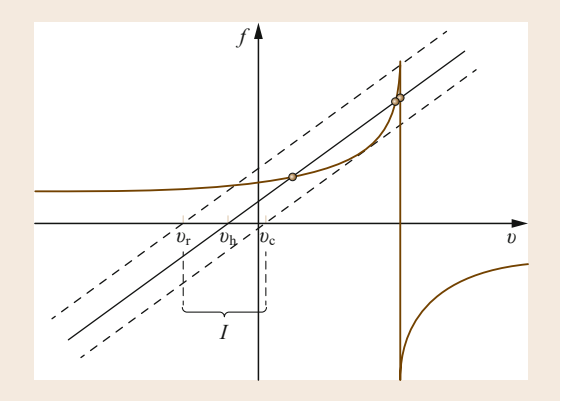


Fig. 10.9 Force f of the bow acting on the string depending upon the string velocity v. The positive force corresponds to the sticking phase, the negative force to the slipping phase. The point of sudden change is at the bow velocity v_h. The three slopes determine the region of ambiguous velocity-force dependency

10.4.2 Stick–Slip Condition Model

Another proposal is to model the string by introducing conditions for the bow to enter sticking or slipping in regard to the string [10.36, 50]. Such conditions can easily be introduced in a finite-difference time domain model (FDTD).

With displacement $f(x,t)$, damping D, the speed of sound c, and adding the bowing force $F(x,t)$, we have for the differential equation of the string

$$c^2 \frac{\partial^2 f(x,t)}{\partial x^2} - D\frac{\partial f(x,t)}{\partial t} - \frac{\partial^2 f(x,t)}{\partial t^2} = F(X,t)\,, \quad (10.18)$$

with the following stick-to-slip conditions:

- If the restoring force of the string at the bow point is higher than the bow force acting on the string or
- if the velocity of the string is higher than the bowing velocity

then the bow will tear-off the string.

If the second case is present we are in the stable regime of periodic sawtooth or Helmholtz motion. In the first case, a totally different string behavior appears, a subharmonic regime as discussed below.

10.4.3 Bifurcations and Subharmonics

Subharmonics and bifurcations have been widely studied with violin bowing [10.51, 52]. When leaving this

region of normal sawtooth motion by in- or decreasing the bowing pressure too much, two other regimes are reached:

- If the bowing pressure is larger than a threshold, a subharmonic regime is reached, where the played pitch depends on the bowing force.
- If the bowing pressure is smaller than a threshold, a bifurcation regime is reached, where inharmonic sounds are produced.

Subharmonic playing also produces normal pitches, still much lower than normal violin tones could be played. Bifurcation regimes produce noisy sounds and are more easily produced by reducing the amount of rosin on the bow leading to a low-volume, high-pitched noisy sound.

Figure 10.10 shows the time series for all three regimes. It was produced by a finite-difference solution of the string equation driven by a bow with constant velocity and linearly in- and decreasing bowing pressure. The pressure starts at $p = 0$ at the very left, increases up to the point of maximum amplitude in the middle of the plot, and linearly decreases again to become $p = 0$ again at the very right. Figure 10.11 shows a correlogram of this time series. A correlogram is like an autocorrelation of an autocorrelation of a time series, displaying only harmonic overtone series. Therefore, this plot displays the produced pitch of the time series.

The basic properties of bowing are shown in both plots:

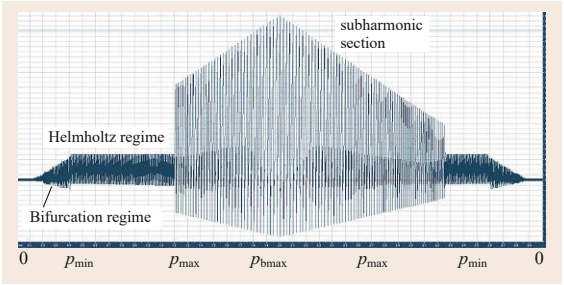


Fig. 10.10 Amplitude time series of the violin string/bow simulation, which was taken at the bridge point of the string, where the sound is transmitted to the radiating violin body; bowing pressure (*horizontal axes*) versus amplitude (*vertical axes*). The bowing pressure was increased from $p_b = 0$ to $p_b = p_{b\,max}$ linearly and then again linearly decreased to $p_b = 0$ again. The time series shows clear phase changes at certain pressure values. The Helmholtz regime is reached at p_{min} and becomes the subharmonic regime at p_{max}, which shows a more stable behavior with decreasing than with increasing pressure, a kind of hysteresis effect (Fig. 10.11). The bifurcation regime with lower pressure values than the sawtooth region also decreases amplitude and starts again at a certain threshold

- At the beginning with low pressures a complex bifurcation regime exists, as discussed below.
- Then at a sudden bowing pressure threshold a stable Helmholtz or sawtooth regime begins.
- Again at a certain threshold a subharmonic regime appears. As can be seen in the correlogram, the pitch of the played tone is strongly and linearly changing with linear increase of the bowing pressure. Also, the amplitudes in- and decrease linearly.
- The threshold at which the sawtooth regime is changing to a subharmonic regime is higher than the return threshold from subharmonic to sawtooth behavior. This is a typical hysteresis where an established regime tries to maintain itself as long as possible.
- During an increase of bowing pressure within the Helmholtz regime in the correlogram the pitch of the tone is slightly decreasing. This so-called 'flattening effect' of bowing is caused by the increased velocity difference between string and bow so as to tear-off the bow from the string with increased bowing pressure.

The bowing regime where the bowing pressure is too low to produce a sawtooth motion is very complex. Figure 10.12 shows the fundamental pitches during small subregimes within this bifurcation regime. These subregimes are again stable for a very small pressure region.

Overall, bowing is a very complex mechanism which, because of this complexity leads to a very simple output, a constant periodicity. This periodicity in the

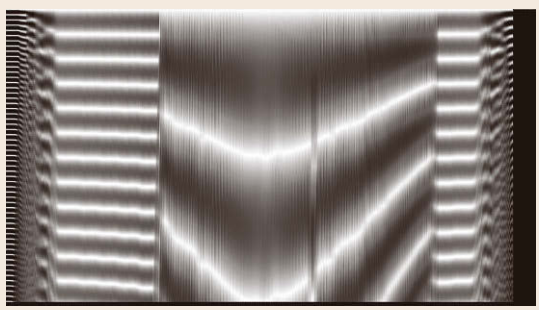


Fig. 10.11 Correlogram of the time series data of Fig. 10.4; bowing pressure (*horizontal axes*) versus fundamental frequency 100 Hz–20 kHz, linear scale (*vertical axes*). The different regimes can clearly be distinguished (see the caption for Fig. 10.10). The Helmholtz region shows a nearly constant fundamental frequency with a slight decrease of frequency known as the flattening effect. The subharmonic region clearly shows the dependence of the fundamental frequency on the bowing pressure. The bifurcation region shows small regions of constant fundamental frequencies between which phase changes occur

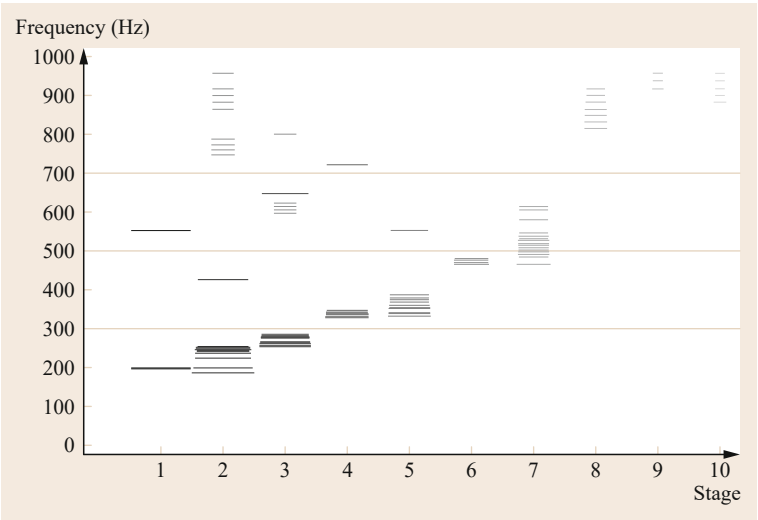


Fig. 10.12 Bifurcation scenario of the bowed string under the threshold pressure value needed to establish the Helmholtz regime (*Stage 1*); data from simulation

normal playing region has a fundamental pitch caused by the string length which only then makes normal violin playing by using fingerings possible at all. This mechanism also holds over a certain bowing pressure region making playing more convenient. Still, when leaving a certain pressure region the behavior of the system is very different, where the played pitch is determined by the bowing pressure itself. In contemporary forms of music these additional features of bowing are used to play low pitches on instruments which are normally not able to produce such low sounds.

10.4.4 Synchronization Condition

The reason for synchronization in the bow/string system and why sometimes the string and sometimes the bowing pressure determines the pitch becomes clear when using the stick–slip condition model. If the bow force is above a certain threshold, the damping of the string is too large to still be able to tear-off the bow from the string and therefore it no longer determines the pitch. Violin strings need to be damped much more than, e.g., guitar strings, for the system to work as the energy of the bow into the string must not survive more than one cycle around the string. Otherwise, the old traveling waves from previous tear-offs would disturb the present tear-off process and therefore a regular sawtooth motion would not be possible. Therefore, here the system which is damped less becomes more prominent and determines the pitch of the whole system.

10.4.5 Synchronization of Organ Pipes

Also synchronization may take place between musical instruments. Not many investigations have been carried out in this regard, those that have mainly deal with coupled organ pipes. This coupling between organ pipes standing next to each other or the coupling of organ pipes near a wall results in frequency synchronization between pipes [10.38, 39, 53]. Figure 10.13 shows the synchronization of a pipe and a loudspeaker standing next to it in relation to distance. *Abel* et al. use a van der Pol oscillator model to understand the synchronization assuming phase-locking when the pipes and frequencies are close enough. Varying these parameters leads to a d'Arnold tongue behavior, where with decreasing distance the range of frequency differences between the pipes where synchronization occurs is increasing in a nonlinear way.

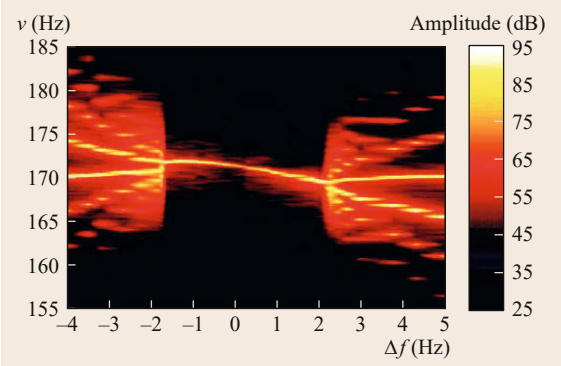
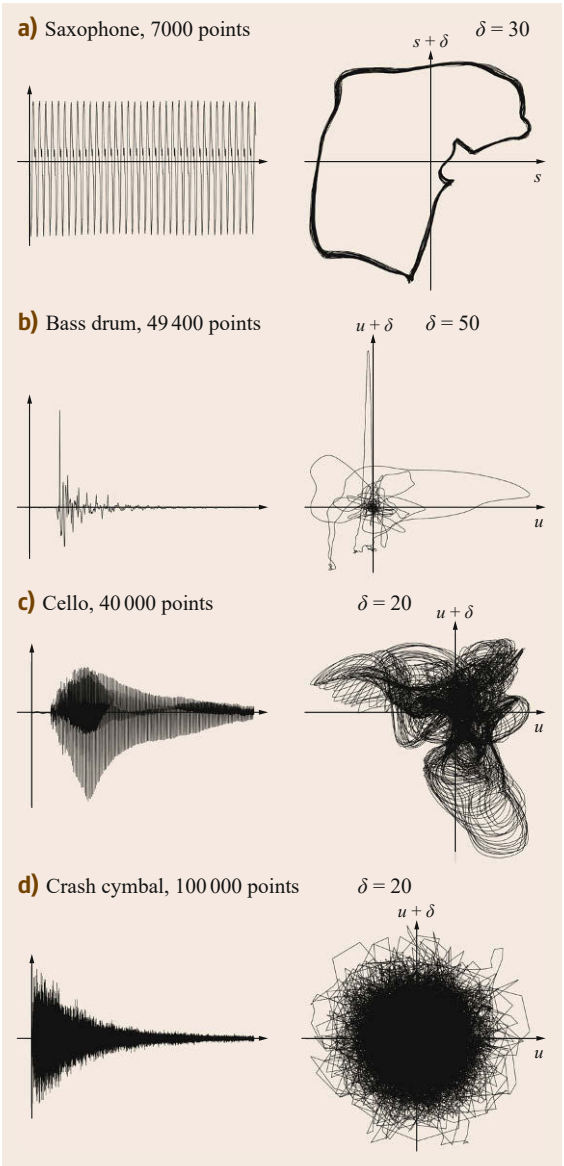


Fig. 10.13 Synchronization of an organ pipe and a loudspeaker standing next to it as the speakers frequency changes. Sharp onset and offset thresholds appear starting and ending the synchronization. Also the sidebands appearing in the nonsynchronized regions disappear during synchronization. (Reprinted with permission of [10.38]. Copyright 2006, Acoustical Society of America)

10.5 Fractal Dimensions of Musical Instrument Sounds

The term fractal is most often associated with self-similarity. The most common example is a snow flake which is complex in structure but is built following one simple rule of replacing lines in the geometry by stars, which then also have lines which are then once again replaced by stars ad infinitum. Then there is a self-similarity while operating on different levels of a scale. If there is one simple rule holding for all levels then the complex geometry can be explained by one simple construction rule. Also, as this rule holds for all levels of a scale from large to small, there is a scaling law characterizing the structure.



Still, the terms fractal and fractal dimension are derived from a mathematical standpoint. The relation between the length of a line r and the volume of a body C has the relation

$$C = r^D \,. \tag{10.19}$$

Here D is the dimension which for ordinary bodies has values of $D = 1, 2, 3$ and so is one-dimensional, two-dimensional, or three-dimensional. Now when writing this equation for the dimension we have

$$D = \log_r C = \frac{\log C}{\log r} \,. \tag{10.20}$$

So if there is a relation between the logarithms of a parameter C with changing r, the dimension is calculated by a fraction and may therefore be called fractal. Also as D may be different from an integer, a fractal dimension may have any value. Plotting this relation means using a log / log plot as both parameters C and r are present as logarithms. The fraction in such a plot means a slope and therefore we can say that the fractal dimension of a system described by $C(r)$ with changing r is the slope of a log / log plot.

When using measured data for such a structure, like when using a musical sound, of course this definition only makes sense when this slope is constant over a wide range of r. Then the time series is characterized by a fractal dimension. Such calculations have been performed for multiphonics of wind instruments [10.54], biphonations of the voice [10.9], initial transients for musical instrument sounds [10.36], or analyzing musical pieces of contemporary music [10.55].

There are many kinds of fractal dimensional calculations, a box-counting dimension, an information dimension, a correlation dimension, etc. which are shown to result in the same fractal number [10.2]. Still, the calculations are different where the fractal correlation dimension was built for time series and it is therefore often used with musical sounds. Still, other dimensions

Fig. 10.14a–d Four examples of musical sounds displaying their time series $s(t)$ on the left and their pseudo phase plots with different delays δ on the right in a two-dimensional embedding. **(a)** Limit cycle of a quasi steady state of a saxophone tone, **(b)** fixed point of a bass drum sound recorded with an accelerometer at the resonance membrane, **(c)** attack and decay of a cello tone with hard attack at the beginning and released bow, **(d)** a crash cymbal of a drum kit ◄

may be used for other purposes, like an information structure for musical onset detection [10.36].

10.5.1 Pseudo Phase-Space

All calculations of fractal dimensions start with the pseudo phase plot [10.56]. Here a one-dimensional time series $X(t)$ is embedded in an n-dimensional space by a delay variable δ

$$X(t) = \{x(t), x(t+\delta), x(t+2\delta), \ldots, x(t+\delta)\}\,, \tag{10.21}$$

with embedding dimension n.

Figure 10.14 shows four examples including a saxophone, a bass drum, a cello, and a crash cymbal in their time series and phase plots. The saxophone tone is shown in its quasi steady state only where the stable periodicity produces a limit cycle attractor. Still, the cycle shows deviations which is the quasi of the quasi steady state, where small amplitude and frequency deviations are displayed. The bass drum is a percussive sound showing strong amplitudes at the beginning of the sound and a fast decrease producing the knot at the axis origin. The cello tone is displayed with an initial transient with low amplitudes and additional inharmonic frequency components to the basically harmonic overtone structure. This plot shows a limit cycle but with very different cycle amplitudes and again the knot at the origin where the tone starts and ends. The crash cymbal has many inharmonic overtones, therefore coming close to a noise sound. Here no limit cycle is present and we have a chaotic attractor.

Phase plots can also be used to display the development of sound in detail. Figure 10.15 shows attractors at adjacent time points of a soft saxophone sound Fig. 10.15a, a violin sound with a hard attack Fig. 10.15b, and a Cambodian metalophone (*roneat deik*, Fig. 10.15c). All plots are normalized to the maximum amplitude of each plot. This ignores the amplitude decay of the sounds over time and therefore displays more details. The soft attack of the saxophone starts with noise at 0 ms, after which only slowly does a stable limit cycle with many deviations appear. The violin sound has a scratchy beginning with many inharmonic components shown at the beginning to then reach a very stable limit cycle. The last plot at 6000 ms shows the string still vibrating after bow release from the string. The *roneat deik* has an inharmonic structure and like the crash cymbal in Fig. 10.14 above begins with a noise-like attractor. After a while the higher harmonics have attenuated leaving basically the fundamental frequency vibrating where then a stable oscillation appears with only some inharmonic partials left.

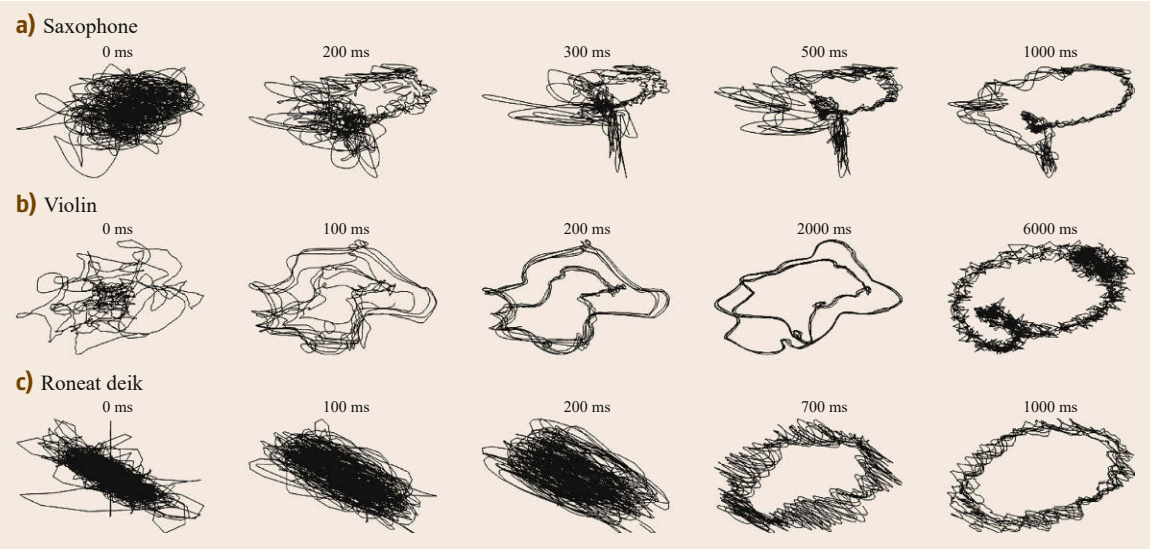


Fig. 10.15a–c Time development of phase plots of a saxophone tone with a soft attack (**a**), a violin tone with a hard attack (**b**), and a Cambodian metalophone (*Roneat deik*) (**c**) with an inharmonic overtone spectrum. The saxophone starts with noise and is only slowly reaching a quasi steady state. The violin sound has a scratchy beginning but is reaching a stable state, the 6000 ms plot displays the string after bow release. The Cambodian *Roneat deik* has an inharmonic overtone structure and therefore a chaotic plot which enters a cyclic motion when the higher harmonics have decayed and only the fundamental is still strong

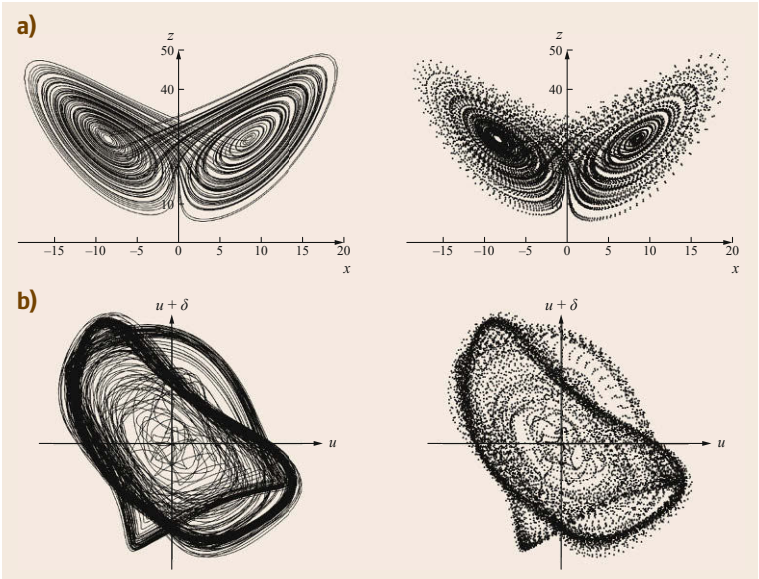


Fig. 10.16a,b Comparison of trajectories (*left plots*) and single point sets (*right plots*) of a quasi steady state and a transient attractor, **(a)** Lorenz attractor (x/z plane) as a quasi steady state attractor, **(b)** sinusoidal transient of first 100 Hz then 180 Hz with overall rising amplitudes

10.5.2 Fractal Correlation Dimension

So the pseudo phase plots display the fundamental behavior of musical tones very well and are therefore a reasonable starting point for further calculations. Here only the fractal correlation dimension is discussed as it is the one most widely used and numerically equivalent to all other dimensions.

Starting from the n-dimensional embedding X as discussed above to calculate a fractal dimension we need to have a parameter $C(r)$ changing with a scaling r. Here first all distances between two points X_i and X_j in the n-dimensional space are calculated. Then r is a maximum distance for which we count how many distances $C(r)$ are below r like

$$C(r) = \frac{1}{N^2} \sum_{i \neq j} H(r - |X_i - X_j|) \,. \tag{10.22}$$

Here, $H(x)$ is the Heaviside function with

$$H(x) = \begin{cases} 0, & \text{for } x \leq 0 \\ 1, & \text{for } x > 0 \end{cases} \,. \tag{10.23}$$

Then the correlation dimension is again the fraction of the logarithms of $C(r)$ and r like

$$D = \frac{\ln C(r)}{\ln r} \,. \tag{10.24}$$

If D is constant over a wide range of r then a general scaling law appears in the sound characterized by the fractal dimension value D.

Musically this makes great sense as the correlation was developed to count the number of dynamic subsystems of a system. It appears that a subsystem as taken by this algorithm is a harmonic overtone structure no matter how many overtones there are as long as they are all in harmonic relation 1 : 2 : 3 : . . . Another harmonic overtone series raises the dimension number by one as do all inharmonic components by themselves as they may be interpreted as a new subsystem which may consist of only one partial. Also large amplitude fluctuations raise the fractal correlation dimension by one. The correlation dimension starts with the normalized phase plot and therefore does not consider the overall amplitude of the sound at the time slice the phase plot is constructed of. This amplitude may be added later to account for it, still it is convenient to have both parameters separated at first, the chaoticity or complexity as displayed by the fractal dimension and the loudness of the sound.

So the algorithm behaves like this:

1. If only one harmonic overtone spectrum is in the sound, $D_C = 1$ no matter how many overtones are present.
2. Each additional harmonic overtone spectrum raises D_C to the next integer.
3. If only one inharmonic sinusoidal is added, D_C raises to the next integer making it suitable for detection of additional single inharmonic components.
4. Large amplitude fluctuations lead to a rise in D_C.
5. As the absolute amplitude is normalized, D_C does not depend upon amplitude.
6. If a component is below a certain amplitude threshold it is no longer considered by the algorithm.

Table 10.2 Fractal correlation values and length of the initial transients for a Tsuji classical guitar with a cedar top plate and rosewood body. The number in brackets with the d-string are the values for the pre-scratch if it was long enough to calculate the fractal dimension value. (*Ap* is Apoyando, *Ti* is Tirandu)

	Ap fortissimo	Mezzoforte	Piano	Ti fortissimo	Mezzoforte	Piano
e-String						
e^1	3.2	3.1	3.5	4.0	3.6	3.0
	30 ms	35 ms	35 ms	40 ms	30 ms	30 ms
g^1	4.4	3.9	3.7	2.8	3.5	2.7
	45 ms	40 ms	40 ms	30 ms	30 ms	25 ms
b^1	3.8	2.7	2.7	2.7	2.8	2.5
	40 ms	30 ms	35 ms	30 ms	30 ms	30 ms
e^2	3.8	2.4	1.6	3.0	1.5	2.0
	40 ms	30 ms	20 ms	30 ms	20 ms	25 ms
b-String						
b	3.5	2.5	2.5	3.5	1.6	2.1
	30 ms	25 ms	30 ms	40 ms	15 ms	20 ms
d^1	3.3	2.6	2.5	3.8	1.6	2.2
	30 ms	25 ms	30 ms	35 ms	15 ms	25 ms
$f\#^1$	3.2	3.8	3.4	3.5	1.6	3.1
	30 ms	35 ms	35 ms	35 ms	20 ms	35 ms
b^1	3.5	1.6	2.5	3.5	1.5	3.2
	30 ms	20 ms	30 ms	35 ms	15 ms	35 ms
g-String						
g	2.5	1.7	1.8	2.7	2.0	1.1
	25 ms	20 ms	15 ms	25 ms	20 ms	10 ms
b♭	3.1	2.5	1.6	2.2	1.5	1.3
	30 ms	25 ms	20 ms	20 ms	15 ms	15 ms
d^1	3.2	2.5	1.7	2.9	1.3	1.3
	35 ms	30 ms	20 ms	35 ms	15 ms	15 ms
g^1	2.5	3.5	3.5	3.2	2.5	2.8
	30 ms	40 ms	30 ms	35 ms	25 ms	30 ms
d-String						
d	(1.6)2.7	2.5	1.8	(1.1)2.7	(-)2.5	(-)3.2
f	(-)1.8	(1.7)2.5	1.6	(-)3.0	(-)2.5	(-)2.5
a	(-)2.4	(1.8)1.8	(-)1.2	(-)(-)	(1.4)(-)	(-)1.6
d^1	(-)1.8	(-) 2.8	(-)	(-)	(-)	

In the literature, it is sometimes stated that the fractal correlation dimension can only calculate steady states and not initial transients. This is not quite the case as can be seen when comparing a standard attractor, the Lorenz attractor of $D = 2.05$ with a transient sound of a clarinet changing a note as displayed in Fig. 10.16. Both phase plots show two attractors where we expect both to be $D = 2$. Indeed, the clarinet attractor is at $D = 2$. The difference between both attractors is that the trajectory of the Lorenz attractor travels a small amount of cycles on the left attractor then moves to the right and after some cycles there it moves back again. The clarinet attractor first stays at one attractor and then moves completely to the second one as this is the next tone played. However, the correlation dimension does not know about the temporal development of the trajectory as it is only taking all distances of the points into consideration. Therefore, to be precise when displaying the phase plot only the points should be displayed and not the temporal development between the points. Both cases are displayed in Fig. 10.16.

10.5.3 Initial Transients

So the fractal correlation dimension is very suitable for calculating the complexity or chaoticity of initial transients of musical instrument sounds. Table 10.2 shows the fractal dimensions of guitar tones of one guitar for the high e, h, g, and d-string for three volumes, fortissimo, mezzoforte, and piano, and two articulations *apoyando* (hard attack) and *tirandu* (soft attack). The dimensions are up to $D \approx 5$ meaning that four loud additional frequencies are added to the basically harmonic overtone series of the tone. The dimension does not

depend on volume or articulation very much which is expected when taking the physics of the guitar into consideration [10.36, 50].

Table 10.3 shows the mean values for three classical guitars for the two articulations. The *Lauenhardt & Kobs* has the smallest values and the *Tsuji* guitar has the highest value for *apoyando*. The subjective evaluation of guitar builders and experts in relation to these three guitars correspond to this finding concerning parameters like loudness or presence of the three instruments. Indeed, the inharmonic components of the guitars are perceived as a *knocking* at the tone onset which makes the sound more prominent or present. Therefore, higher values of fractal dimension of initial guitar transients point to a higher perceptual presence.

Further investigations with other instruments confirm this finding in general [10.36]. Violins and wind instruments have a wider range or articulations and therefore a wider distribution of fractal dimension values where hard violin attack sounds may have a dimension up to eight.

10.5.4 Mirliton

Also multiphonics [10.54] or the sound of mirliton, instruments where one finger hole is covered by a thin membrane [10.57] show a higher fractal dimensionality. Although not perfectly understood yet, mirliton instruments, as well as some multiphonic saxophone or clarinet sounds often show sidebands next to harmonic partials like

$$f_{n,m} = \frac{1}{(1+m)} n f_0 \pm m f_m \quad \text{with} \quad n = 1, 2, 3 \quad \text{and} \quad m = 0, 1, 2, 3 \ldots , \tag{10.25}$$

where f_0 is the fundamental frequency of the tone, nf_0 are its partials, f_m is a modulating frequency, and mf_m are the deviations around the harmonic partials with sideband frequencies decreasing around the harmonics.

Table 10.3 Mean and standard deviations for the two articulations for the three guitars

	Tsuji	Zander	Lauenhardt & Kobs
Apoyando			
Mean	2.9	2.5	2.0
Standard dev.	0.75	0.93	0.56
Tirandu			
Mean	2.5	2.6	2.3
Standard dev.	0.83	1.3	0.65

This leads to a rough and inharmonic sound with higher fractal dimensions. Figure 10.17 shows the time development of a Chinese *dizi* flute with the membrane unattached (first tone) and the membrane attached (second tone). Clearly the fractal dimension with the rougher sound of the attached membrane rises in general as well as with articulation.

10.5.5 Musical Density

The fractal dimension may also be used when investigating art music of the 20th century like free jazz, free improvised music, contemporary classical music, or sound-based music like Techno or Dance music. In many of these styles the temporal development of musical density is a prominent composition parameter as well as a prominent perception parameter with sound textures where no melodies, phrases, or cadenzas are present. As the perception of density is perceived instantaneously by listeners it needs to be considered by composers and musicians alike in these genres.

As an example, Even Parker's solo saxophone piece *Broken Wing* was calculated in terms of its fractal dimension [10.55] as shown in Fig. 10.18. Even Parker uses extended techniques of saxophone playing like complex tongue movement at the reed, multiphonics, and related techniques to arrive at an often nonharmonic tone production. The piece is perceived as starting with a low sound event density, then gradually increasing this and after reaching a maximum gradually becoming sparser again and finally ending in a low event density. The plot represents exactly this behavior for the per-

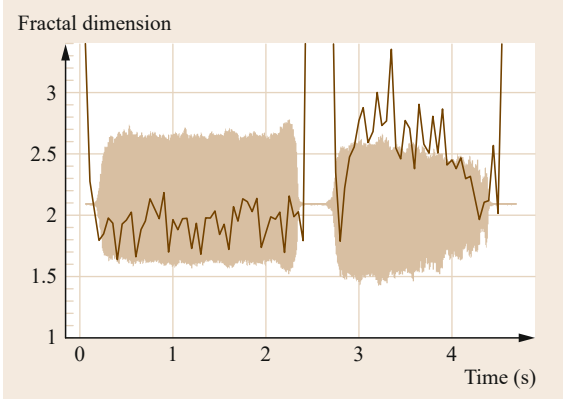


Fig. 10.17 Fractal dimension of two sounds of a Chinese *dizi* flute of same pitch, first played normally by closing the mirliton hole completely, second played as mirliton. The fractal dimension of the rough mirliton sound is considerably higher than with normal playing. Sections in between are noise with very high fractal dimensions

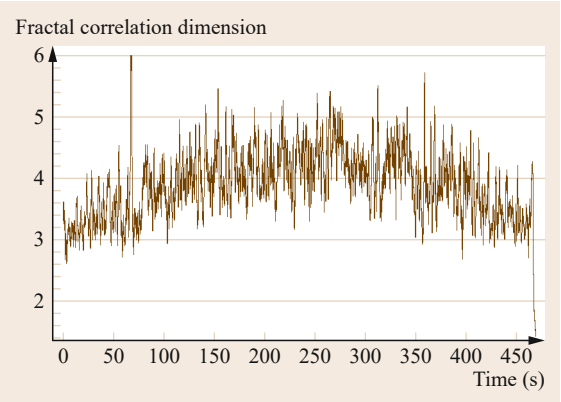


Fig. 10.18 Fractal dimensions of Even Parker's saxophone solo piece *Broken Wing* using extended techniques for saxophone tone production. One can clearly see the in- and decreasing behavior of the perceived density of the piece over 460 s ◄

formance and is therefore in very good correspondence with the event density musical feature for the whole musical piece.

So the fractal dimension displays many musical features:

- *Musical density* as often used in free improvised, electronic, contemporary classical, or experimental music where this parameter is crucial both for composition and perception of pieces.
- *Chaoticity of initial transients* as a measure of perceived presence and loudness of the musical tones.
- *Counting of the numbers of voices* in a musical texture as with homophonic or polyphonic music of all styles from Jazz, Rock, Metal, Techno, non-Western, or classical music.

10.6 General Models of Musical Instruments

Now, after our review of the main instrument families and their synchronization methods, we will sum the findings into general models to understand the phenomena. This general view has a top-down character and may not explain all details of the instruments at first. Still, the models can provide insight into basic aspects of synchronization.

10.6.1 Phase-Locking

As mentioned in some detail especially in the wind instrument section, the different frequencies or modes in an instrument have frequencies that do not match a harmonic overtone spectrum precisely and may deviate from it considerably. So one way of explaining synchronization is in terms of an energy transfer between the modes due to nonlinearities especially in the generator region. Then modes that are too slow can be pushed to catch up with faster modes and modes that are too fast can be drawn back to arrive at mode-locking and synchronize the phases.

One such general model proposed for wind and string instruments suggests a coupled equation system for modes [10.58]. Although a detailed description of the math is beyond the scope of this paper, the basic reasoning is that for displacement x_i and harmonic n_i of mode i driven by a force $F(\dot{x}_j)$ with eigenvalue λ_j like

$$\ddot{x}_i + k_i\dot{x}_i + n_i^2 x_i = \lambda_j F\left(\dot{x}_j\right) . \tag{10.26}$$

there is a solution that is a modification of the standard solution as we expect the amplitudes a_i and frequencies ω_i of the modes to change over time by phase Φ_i like

$$\dot{a}_i \sin\left(\omega_i t + \Phi_i\right) + a_i\dot{\Phi}_i \cos\omega_i t + \Phi_i = 0 . \tag{10.27}$$

Using only terms that slowly vary with ω, which are only considered here as only these are able to synchronize adjacent frequencies, we arrive at estimations of the amplitude and phase variations using

$$\langle\dot{a}_i\rangle = \frac{\lambda_i}{\omega_i}\left\langle F\left(\dot{x}_j\right)\cos\left(\omega_i t + \Phi_i\right)\right\rangle - 0.5k_i a_i , \tag{10.28}$$

$$\left\langle\dot{\Phi}_i\right\rangle = -\frac{\lambda_i}{a_i\omega_i}\left\langle F\left(\dot{x}_j\right)\sin\left(\omega_i t + \Phi_i\right)\right\rangle + \frac{\left(n_i^2 - \omega_i^2\right)}{2\omega_i} , \tag{10.29}$$

where the $\langle\rangle$ denotes that only slowly varying terms have been used. Now mode-locking is present when

$$qj = pi , \tag{10.30}$$

and

$$qn_j \approx pn_i , \tag{10.31}$$

with $p, q = 1, 2, 3, \ldots$ When assuming a nonlinear coupling function $F(\dot{x}_j)$, we arrive at the equation for mode-locking

$$p\left(n_i + \dot{\Phi}_i\right) = q\left(n_j + \dot{\Phi}_j\right) = pi\omega . \tag{10.32}$$

Here a common frequency ω holds for two modes n_i and n_j and its harmonics p and q when $\dot{\Phi}_i$ and $\dot{\Phi}_j$ are

appropriate. So as stated above, the phase changes slow down or speed up the modes and therefore lock them together to a common frequency.

Coupled systems of van der Pol oscillators also belong to this category as found with the singing voice. Similar models for musical instruments have been proposed in the literature, either a general view [10.59], or specifically for organ pipes [10.38], percussion instruments [10.60], and the singing voice [10.11].

10.6.2 Force Function Model

Another approach is a model that assumes a nonlinear generator and a linear resonator and has been proposed for violin–bow interaction, but was also proposed to hold for wind instruments [10.48, 49]. This model has already been discussed above in the bow–string section. The basic idea is that a force function of nonlinear nature is driving a system where the energy sent out by the generator (bow or mouthpiece) is traveling through a resonator (string or tube) and when returning is shaped by the force function again. Because of the nonlinear nature of this force function a regular motion appears which at first would also be present in a linear driver as no nonlinear distortion in the string or tube is assumed. Still, in addition to regular oscillation other regimes like subharmonics or double-slip or scratchy sounds can occur with violins when a parameter, the bowing or blowing pressure, reaches a certain threshold.

This model has analytical power because of the nonlinear driving mechanism, although in many cases the nonlinear function is not trivially found.

10.6.3 Impulse Pattern Formulation (IPF)

Another model sees musical instruments in terms of impulses starting from one system like a string, which travels through the instrument and interacts with the initial system again in a nonlinear way [10.36, 61]. This nonlinearity is found mainly in the damping taking into account the findings for musical instruments that in a coupled oscillator system the one with less damping forces the other into its oscillations. This model results in a logistic map with coupling constant α

$$g(t) = g(t_-) - \ln \frac{1}{\alpha} g(t_-) \,, \tag{10.33}$$

where a system g at different times $g(t)$ and a previous time $g(t_-)$ are interacting in an iterative way. The logarithm comes from assuming an exponential decay for the traveling wave due to damping. This damping may be the turbulent damping of a vortex as in wind instruments, the damping of a reed, or it might also be the damping of a wave traveling in wood.

This model can easily be enlarged to a multiple oscillator system where one system sends out an impulse which is then reflected at multiple points, like a string sending its energy to the top plate which is then transferred to the ribs, back plate, enclosed air, etc. Such a multidelayed version is

$$g(t_+) = g(t) - \ln \left[\frac{1}{\alpha} g(t) - \sum_{k=1}^{n} \frac{\beta_k}{\alpha} \mathrm{e}^{g(t) - g(t_{-k})} \right], \tag{10.34}$$

where β_k is the k-th impulse damping at the $t_{-k\text{-th}}$ step before the time step t with n+1 back-traveling impulses.

A stability analysis of the two systems shows that in a system of only one sending oscillator and one reflection point, as with wind instruments or bow-string interaction, a full bifurcation scenario appears as shown in Fig. 10.19.

Still, when using the multidelayed system and inserting the reflection strength found with a guitar these multiple reflection points stabilize the system and bifurcation and noise are no longer present as shown in Fig. 10.20.

This could explain why in stringed instruments where the string forces the body into its vibrations with all possible playing strengths a stable oscillation appears and no bifurcations, sudden onsets, or noise can be produced like with wind instruments or other systems with only two oscillators. The stability of a stringed instrument system is taken for granted and explained as a simple linear generator/resonator system. However, when investigating all instrument types such a simple relation is no longer present and therefore it is important to also find a reason for the stability in string–body interactions.

10.6.4 IPF and Initial Transients

An IPF system is also able to estimate the basic behavior of the initial transient of instruments as shown in Fig. 10.21. Taking different view points of a guitar, like the string, top plate, back plate etc. the impulse pattern as calculated by the IPF matches the basic behavior of the guitar parts as calculated by a finite-difference time domain (FDTD) method of the whole guitar body shown on the right of the figure. The FDTD displays the time series of the parts in every detail, while the IPF only shows the development of the impulses for the wave shapes to be added later if needed. It appears that the general length and complexity of the initial transients of both, FDTD and IPF, are very similar.

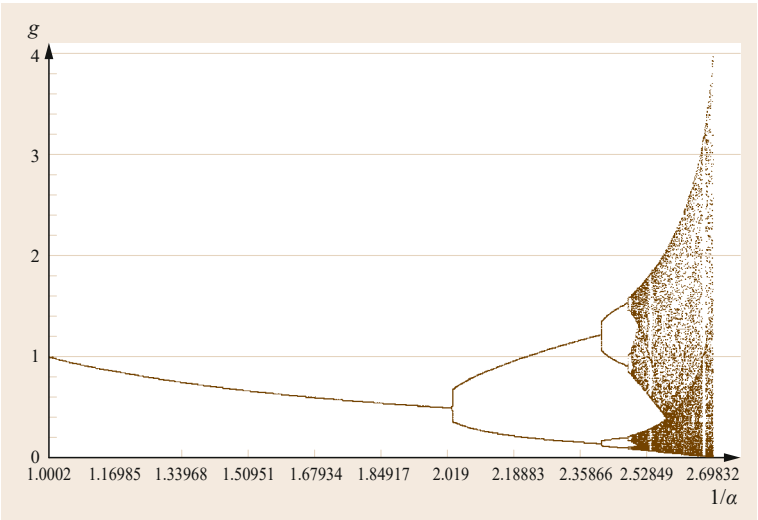


Fig. 10.19 Bifurcation scenario of the impulse time interval equation of one time point acting upon another time point for different values of $\frac{1}{\alpha}$, converged after the initial transient

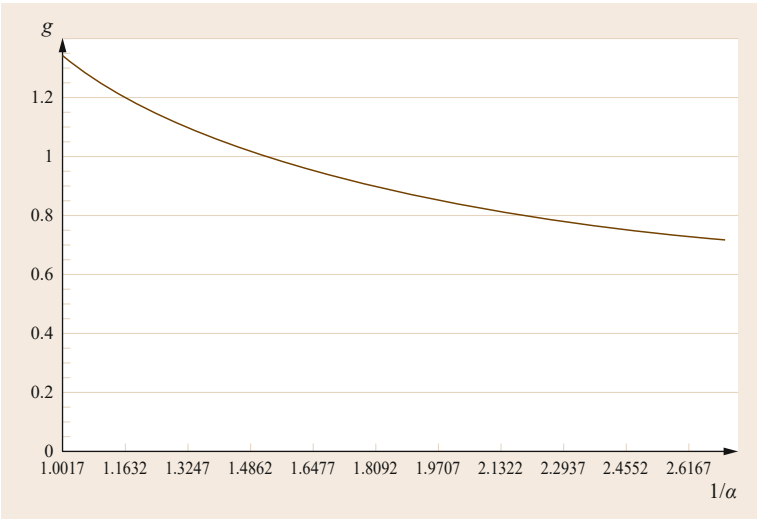


Fig. 10.20 Bifurcation scenario for the multilayered back-impulse equations with four back impulses ($n = 4$). Here $\beta_1 = 0.2$, $\beta_2 = 0.1$, and $\beta_3 = 0.05$. $1/\alpha$ is varied from $1 < 1/\alpha < 2.7$

Therefore, the IPF seems to be able also to explain the initial transient behavior of different guitar parts. As it does not display the whole time series but only the impulse development it is able to give more insight into the general behavior of these parts than the more complex time series.

10.6.5 Synchronization Conditions

We are now able to summarize the conditions for synchronization as found with musical instruments. Two of these conditions are known from synergetics and self-organization, the third one is found with the multidelayed system of stringed instruments of the IPF type.

The conditions for synchronization are:

- The less-damped system forces the more strongly damped system into its vibrations.
- The system with lower frequency forces the one with higher frequency into its oscillation.
- The lower dimensional system forces the higher dimensional system into its periodicity.

The three general types of synchronization summarized above of phase-locking, force function, and impulse pattern handle these conditions in different ways. Phase-locking formulations take damping and frequency relations into consideration. They can also include higher dimensional systems by coupling more equations, but this is often demanding in terms of finding a solution. The force function methods again consider damping and frequency but do not formulate higher di-

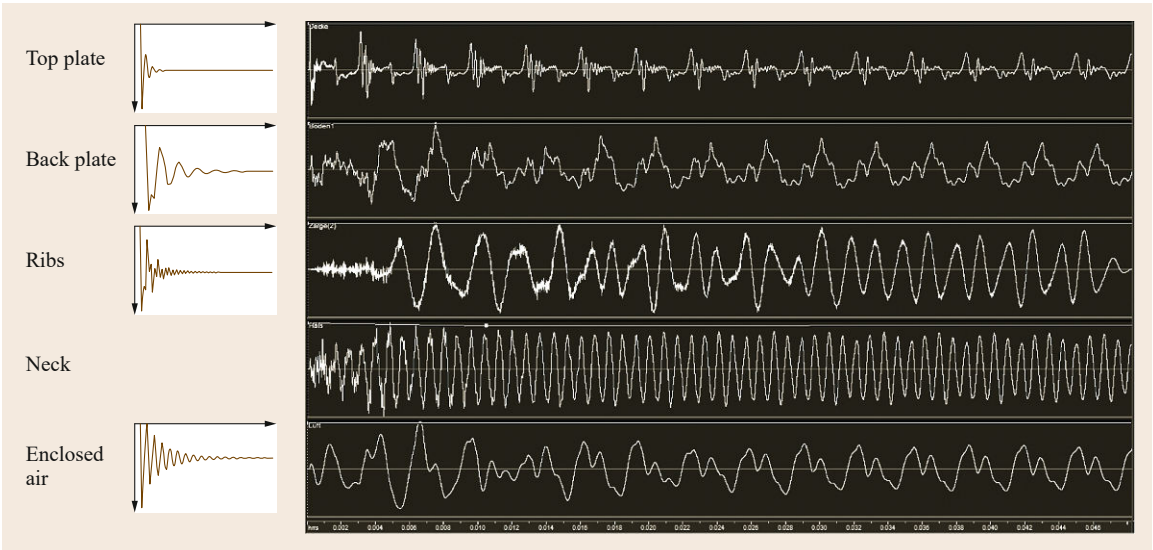


Fig. 10.21 Comparison between a finite-difference time domain (FDTD) model of a complete guitar geometry for the different guitar parts in its initial transient phase (*left*) and the impulse patterns calculated by a multiple-delayed system for the respective body parts (*right*). Note that the FDTD results are time series while the IPF series are the development of impulses without detail on the wave shapes. The FDTD and the IPF correspond very well in terms of length and complexity of the initial transient phase. (For details see [10.36])

mensional problems. The IPF takes all three into consideration. It also calculates the wave fronts or impulses and therefore is able to display the basic skeleton of initial transients. Still, it does not produce a time series at first which needs to be constructed later using an impulse shape, like a sawtooth, rectangle, sinusoidal etc.

10.7 Conclusions

Synchronization is perceptually important and hardwired in the human cochlea [10.62]. The reason for this might be evolutionary although this is just speculation. The self-sustained oscillation of the vocal folds is a highly efficient way of producing a tone as it synchronizes the otherwise broadband noise into single frequencies therefore reducing entropy considerably. These frequencies stand out above environmental noise and are therefore a very efficient way to communicate. So a harmonic overtone structure might be considered as a by-product of the attempt to efficiently produce loud sounds. Still then it is appropriate for the human ear to adapt to harmonic spectra as they are mainly produced by humans or animals which need to concern us more than inharmonic spectra of environmental ambient sounds like produced by wind or water. Indeed we find, e.g., in research of noise perception that noise of a tonal quality disturbs us much more than broadband noise. Also tonal music has been used to simulate intelligent beings in some ethnic groups such as spirits or animistic creatures [10.63]. From this viewpoint musical instruments imitate intelligent life because of their harmonic structure and might therefore be considered as a *universal language*. Also it is interesting to see that many musical instruments use a similar driving mechanism like that of the vocal folds which can be modeled in a similar way.

The converse may also be argued. Musical instruments have been used for the last 40 000 years [10.64] and one might expect even before that time. If the time span musical instruments exist would even be so large as to become salient in terms of evolutionaly changes, the precise fit of the human ear to musical sounds might also be an adaptation of the cochlea and auditory pathway to musical instruments. Then music would be part of our nature and therefore it would not be astonishing to find that it is such a prominent feature in practically all ethnic groups around the world.

All these assumptions are purely speculative in nature. Still, it is interesting to see how important it is for musical instruments to arrive at a perfect harmonic overtone spectrum, which is not at all self-explanatory, and where much effort is put into instrument building to achieve such a synchronization.

References

10.1 H. Haken: *Synergetics* (Springer, Berlin, Heidelberg 1990)

10.2 J. Argyris, G. Faust, M. Haase, R. Friedrich: *An Exploration of Dynamical Systems and Chaos* (Springer, Berlin, Heidelberg 2015)

10.3 V. Aschoff: Experimentelle Untersuchungen an einer Klarinette. [Experimental investigations of a clarinet], Akust. Z. **1**, 77–93 (1936)

10.4 J. Sundberg: *The Science of the Singing Voice* (Northern Illinois University Press, DeKalb 1988)

10.5 I.R. Titze: The physics of small-amplitude oscillation of the vocal folds, J. Acoust. Soc. Am. **83**, 1536–1552 (1988)

10.6 T. Fitch, J. Neubauer, H. Herzel: Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production, Animal Behav. **63**(3), 407–418 (2002)

10.7 P. Mergell, H. Herzel, T. Wittenberg, M. Tigges, U. Eysholdt: Phonation onset: Vocal fold modeling and high-speed glottography, J. Acoust. Soc. Am. **104**(1), 464–470 (1998)

10.8 P. Mergell, H. Herzel, I.R. Tietze: Irregular vocal-fold vibration – High-speed observation and modeling, J. Acoust. Soc. Am. **108**(6), 2996–2300 (2000)

10.9 A. Behrmann, R.J. Baken: Correlation dimension of electroglottographic data from healthy and pathologic subjects, J. Acoust. Soc. Am. **102**(4), 2371–2379 (1997)

10.10 J. Neubauer, M. Edgerton, H. Herzel: Nonlinear phenomena in contemporary vocal music, J. Voice **18**(1), 1–12 (2004)

10.11 K. Ishizaka: Equivalent lumped-mass models of vocal fold vibration. In: *Vocal Fold Physiology* (1981) pp. 231–244

10.12 I.R. Titze, Sh S. Schmidt, M.R. Titze: Phonation threshold pressure in a physical model of the vocal fold mucosa, J. Acoust. Soc. Am. **97**(5), 3080–3084 (1995)

10.13 J.J. Jiang, Y. Zhang: Modeling of chaotic vibrations in symmetric vocal folds, J. Acoust. Soc. Am. **110**(4), 2120–2128 (2001)

10.14 J.J. Jiang, Y. Zhang: Chaotic vibration induced by turbulent noise in a two-mass model of vocal folds, J. Acoust. Soc. Am. **112**(5), 2127–2133 (2002)

10.15 J.G. Švec, H.K. Schutte, D.G. Miller: On pitch jumps between chest and falsetto registers in voice: Data from living and excised human larynges, J. Acoust. Soc. Am. **106**(3), 1523–1531 (1999)

10.16 I.T. Tokuda, M. Zemke, M. Kob, H. Herzel: Biomechanical modeling of register transition and the role of vocal tract resonators, J. Acoust. Soc. Am. **127**(3), 1528–1536 (2010)

10.17 J.C. Lucero: Dynamics of the two-mass model of the vocal folds: Equilibria, bifurcations, and oscillation region, J. Acoust. Soc. Am. **94**(6), 3104–3111 (1993)

10.18 J.C. Lucero: A theoretical study of the hysteresis phenomenon at vocal fold oscillation onset-offset, J. Acoust. Soc. Am. **105**(1), 423–431 (1999)

10.19 J.C. Lucero, L.L. Koenig, K.G. Lourenço, N. Ruty, X. Pelorson: A lumped mucosal wave model of the vocal folds revisited: Recent extensions and oscillation hysteresis, J. Acoust. Soc. Am. **129**(3), 1568–1579 (2011)

10.20 P.Å. Lindestad, M. Södersten, B. Merker, S. Granqvist: Voice source characteristics in Mongolian 'throat singing' studied with high-speed imaging technique, acoustic spectra, and inverse filtering, J. Voice **15**(1), 78–85 (2001)

10.21 I. Steinecke, H. Herzel: Birfurcations in an asymmetric vocal fold model, J. Acoust. Soc. Am. **97**, 1874–1884 (1995)

10.22 M.-H. Lee, J.N. Lee, K.-S. Soh: Chaos in segments from Korean traditional singing and Western singing, J. Acoust. Soc. Am. **103**(2), 1175–1182 (1998)

10.23 D.A. Berry, H. Herzel, I.R. Titze, K. Krischer: Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions, J. Acoust. Soc. Am. **95**(6), 3595–3604 (1994)

10.24 Q. Xue, R. Mittal, X. Zhang: A computational study of the effect of vocal-fold asymmetry on phonation, J. Acoust. Soc. Am. **128**(2), 181–187 (2010)

10.25 F.A. Berry, H. Herzel, I.R. Tieze, B.H. Story: Bifurcations in excised larynx experiments, J. Voice **10**, 129–138 (1996)

10.26 T. Lerch: *Vergleichende Untersuchung von Bohrungsprofilen historischer Blockflöten des Barock (Comparative investigation of bore profiles of historical Barock recorder flutes)* (Staatliches Institut für Musikforschung. Preussischer Kulturbesitz Musikinstrumentenmuseum, Berlin 1996)

10.27 C.J. Nederveen: *Acoustical Aspects of Musical Instruments* (Northern Illinois University Press, DeKalb 1998)

10.28 A.H. Benade: *Fundamentals of Musical Acoustics* (Oxford Univ. Press, New York 1976)

10.29 G. Krassnitzer: *Multiphonics für Klarinette mit deutschem System und andere zeitgenössische Spielarten. (Multiphonics for clarinet with german system and other contemporary styles)* (edition ebenos, Aachen 2002)

10.30 P.A. Durbin, R. Pettersson: *Statistical Theory and Modeling for Turbulent Flows* (Wiley, Chichester 2001)

10.31 B. Fabre, A. Hirschberg, A.P.J. Wijnands: Vortex shedding in steady oscillation of a flue organ pipe, Acta Acust. United Acust. **82**, 863–877 (1996)

10.32 J.-P. Dalmont, J. Gilbert, J. Kergomard, S. Ollivier: An analytical prediction of the oscillation and extinction thresholds of a clarinet, J. Acoust. Soc. Am. **118**(5), 3294–3305 (2005)

10.33 R. Kaykayoglu, D. Rockwell: Unstable jet-edge interaction. Part 1. Instantaneous pressure fields at a single frequency, J. Fluid Mech. **169**, 125–149 (1986)

10.34 R. Kaykayoglu, D. Rockwell: Unstable jet-edge interaction. Part 2: Multiple frequency pressure fields, J. Fluid Mech. **169**, 151–172 (1986)

10.35 A. Richter, R. Grundmann: Numerical investigations of the bassoons aeroacoustic, J. Acoust. Soc. Am.

123, 3448 (2008)
10.36 R. Bader: *Nonlinearities and Synchronization in Musical Acoustics and Music Psychology*, Springer Series Current Research in Systematic Musicology, Vol. 2 (Springer, Heidelberg 2013)
10.37 J.W. Coltman: Sounding mechanism of the flute and organ pipe, J. Acoust. Soc. Am. **44**(4), 983–992 (1968)
10.38 M. Abel, S. Bergweiler, R. Gerhard-Multhaupt: Synchronization of organ pipes: Experimental observations and modeling, J. Acoust. Soc. Am. **119**, 2467 (2006)
10.39 W. Lottermoser: *Orgeln, Kirchen und Akustik (Organs, Churches, and Acoustics)* (Erwin Bochinsky, Frankfurt a.M. 1983)
10.40 C. Koehn: A bowed bamboo tube zither from Southeast Asia. In: *ISMA, Le Mans 2014* (2014) pp. 499–502
10.41 G. Müller, W. Lauterborn: The bowed string as a nonlinear dynamical system, Acustica **82**, 657–664 (1996)
10.42 C.V. Raman: On the mechanical theory of the vibrations of bowed strings and of musical instruments of the violin family, with experimental verification of the results, Bull. Indian Assoc. Cultivat. Sci. **15**, 1–158 (1918)
10.43 L. Cremer: *The Physics of the Violin* (MIT Press, Cambridge 1985)
10.44 A. Askenfeld: Measurements of bow motion and bow force in violin playing, J. Acoust. Soc. Am. **80**, 1007–1015 (1986)
10.45 P. Duffour, J. Woodhouse: Instability of systems with a frictional point contact: Part 1, Basic modelling, J. Sound Vib. **271**, 365–390 (2004)
10.46 P. Duffour, J. Woodhouse: Instability of systems with a frictional point contact: Part 2, Model extensions, J. Sound Vib. **271**, 391–410 (2004)
10.47 W. Güth: A comparison of the Raman and the oscillator models of string excitation by bowing, Acustica **82**, 169–174 (1996)
10.48 M.E. McIntyre, J. Woodhouse: Fundamentals of bowed-string dynamics, Acustica **43**, 93–108 (1979)
10.49 M.E. McIntyre, J. Woodhouse: On the oscillations of musical instruments, J. Acoust. Soc. Am. **74**(5), 1325–1345 (1983)
10.50 R. Bader: Whole geometry finite-difference modeling of the violin. In: *Proc. Forum Acusticum 2005* (2005) pp. 629–634
10.51 R.J. Hanson, A.J. Schneider, F.W. Halgedahl: Anomalous low-pitched tones from a bowed violin string, J. Catgut Acoust. Soc. **2**, 1–7 (1994)
10.52 M. Kimura: How to produce subharmonics on the violin, New Music Res. **28**, 178–184 (1999)
10.53 J. Angster, J. Angster, A. Miklós: Coupling between simultaneously sounded organ pipes, AES E-Library **94**, 1–8 (1993)
10.54 D.H. Keefe, B. Laden: Correlation dimension of woodwind multiphonic tones, J. Acoust. Soc. Am. **90**(4), 1754–1765 (1991)
10.55 D. Borgo: *Sync or Swarm. Improvising Music in a Complex Age* (Bloomsbury Academic, New York, London 2005)
10.56 V. Gibiat: Phase space representations of acoustical musical signals, J. Sound Vib. **123**(3), 529–536 (1988)
10.57 R.V. Velazques: Ancient aerophones with mirliton. In: *Proceedings ISGMA* (2004) pp. 363–373
10.58 N.H. Fletcher: Mode locking in nonlinearly excited inharmonic musical oscillators, J. Acoust. Soc. Am. **64**, 1566–1569 (1978)
10.59 S. Dubnov, X. Rodet: Investigation of phase coupling phenomena in sustained portion of musical instruments sound, J. Acoust. Soc. Am. **113**, 348–359 (2003)
10.60 K.A. Legge, N.H. Fletcher: Nonlinear generation of missing modes on a vibrating string, J. Acoust. Soc. Am. **76**(1), 5–12 (1984)
10.61 R. Bader: Theoretical framework for initial transient and steady-state frequency amplitudes of musical instruments as coupled subsystems. In: *Proc. 20th Int. Symp. Music Acoust. (ISMA)* (2010) pp. 1–8
10.62 P. Cariani: Temporal codes, timing nets, and music perception, J. New Music Res. **30**(2), 107135 (2001)
10.63 F. Messner: Friction blocks of New Ireland. In: *Australia and the Pacific Islands*, Garland Encyclopedia of World Music, Vol. 9, ed. by A.L. Kaeppler, J.W. Love (Routledge, London 1998) pp. 380–382
10.64 N.J. Conrad, M. Malina, S.C. Münzel: New flutes document the earliest musical tradition in southwestern Germany, Nature **460**, 737–740 (2009)

11. Room Acoustics – Fundamentals and Computer Simulation

Michael Vorländer

In room acoustics analytical formulas and computer simulations can be used to predict the acoustics of spaces, not only in terms of reverberation but other perceptual aspects, too, which are related to the perception of music or speech. In this context the room impulse response is the function of main interest. It can be measured by using sophisticated instrumentation and signal processing, or it can be simulated with computer models. In the process of auralization the data and signal processing enables one to listen into the simulated rooms in order to interpret the sound in the room aurally. In real-time implementation, this is a valuable extension of the technique of virtual reality.

Room acoustics is a well-established field with fascinating links to architecture and music. Knowledge concerning sound propagation in spaces, its physics and its effects on perception, is widely available. Efficient design tools are available, too, and experience with solutions for various applications is the basis for daily work in acoustic consulting. After all, room acoustic design and consulting is a quite straightforward process. This statement is true if elementary design rules and creation of acceptable room acoustic conditions are respected. Acceptable acoustic conditions, however, are not always given in rooms for speech and music, because standards and guidelines in room acoustics are not taken into account properly in the first phase of architectural design.

When it comes to requirements for excellent acoustic conditions, the picture is different. In prestigious projects the acoustic conditions need to fulfill high standards. What are these standards? In ISO 3382 [11.1] definitions and guidelines are given for the measurement and interpretation of room acoustic quantities. In famous textbooks, the most well known by *Beranek* [11.2] and *Barron* [11.3], we find a huge amount of data for concert halls and opera houses, and these data indicate target values and tolerance ranges for successful acoustic designs. Computer methods can help to optimize the design. Auralization, which is the most recently developed tool in room simulation, can provide audible results rather than numbers. Thus, an intuitive interface is provided, which allows the designer to experience the sound and its character.

In this chapter, the history and fundamentals of room acoustics are briefly summarized, the state of the art re-visited, and open questions discussed. These questions are related to crossdisciplinary research of classical room acoustics in combination with psychoacoustics and the technology of virtual reality.

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_11

11.1 Fundamentals of Sound Fields in Rooms

For a complete understanding of sound propagation physics we require a wave theoretical approach. Accordingly, a physically correct description and interpretation in enclosed spaces starts with the solution of the stationary wave equation with boundary conditions. For harmonic excitation

$$p = \hat{p}\mathrm{e}^{\mathrm{i}\omega t}\,, \tag{11.1}$$

the Helmholtz equation

$$\Delta p + k^2 p = 0 \tag{11.2}$$

is solved by using specific wave numbers, the eigenvalues k_n with the eigenfrequencies

$$f_n = \frac{c}{2\pi} k_n \tag{11.3}$$

for each eigenfrequency a characteristic distribution of the sound pressure in the room occurs. In a rectangular room, for example, with the dimensions $L_\mathrm{x}, L_\mathrm{y}, L_\mathrm{z} (V = L_\mathrm{x} L_\mathrm{y} L_\mathrm{z})$ and with hard boundaries ($R = 1$) these eigenfunctions (also called *modes*) are

$$p_{lmn}(x,y,z) = 8\hat{p}\cos\left(\frac{l\pi x}{L_\mathrm{x}}\right)\cos\left(\frac{m\pi y}{L_\mathrm{y}}\right) \times \cos\left(\frac{n\pi z}{L_\mathrm{z}}\right) \tag{11.4}$$

with l, m, n positive integers, and the eigenfrequencies

$$f_{lmn} = \frac{c}{2}\sqrt{\left(\frac{l}{L_\mathrm{x}}\right)^2 + \left(\frac{m}{L_\mathrm{y}}\right)^2 + \left(\frac{n}{L_\mathrm{z}}\right)^2}\,. \tag{11.5}$$

This principle holds for any room shape although the eigenfrequencies cannot be calculated in a comparably easy way. In general the number of eigenfrequencies in an interval between 0 and f drastically increases with f^3, while the density of eigenfrequencies within frequency bands increases with f^2.

$$\frac{dN_\mathrm{f}}{df} \approx 4\pi V \frac{f^2}{c^3} \tag{11.6}$$

Sound fields in rooms are usually excited with broadband and transient signals such as speech, music, or noise. Therefore, the total field is a rather complex superposition of eigenfunctions. With losses, damping constants, δ_ν, can be introduced, and the superposition reads

$$p(\boldsymbol{r}_0, \boldsymbol{r}) \propto i\omega \sum_\nu \frac{p_\nu(\boldsymbol{r}_0)\, p_\nu(\boldsymbol{r})}{K_\nu\left(\omega^2 - \omega_\nu^2 - 2\mathrm{i}\delta_\nu\omega_\nu\right)} \tag{11.7}$$

with ν, in short, denoting any combination of l, m, n, and the vectors $\boldsymbol{r}_0$ and $\boldsymbol{r}$ the positions of the source and the receiver, respectively. K_ν are constants depending on the relative amplitudes of the modes.

This function, the so-called *stationary room transfer function*, is a series of overlapping modes (Fig. 11.1). With increasing frequency the overlap of complex sound pressure becomes more and more dominant. The result at a specific frequency crucially depends on the relative phases of the modes involved. Furthermore, the relative phases depend on the complex reflection factors, on the speed of sound in the room (depending on temperature and humidity), and on the exact positions of source and receiver. This data is (a) not known precisely, and (b) it may slowly vary in time, for example the room itself may vary slightly simply in response to persons moving around. After all, according to the pioneering work by *Schröder* [11.4] and *Kuttruff* [11.5], the overlap range above the Schroeder frequency

$$f_\mathrm{Schroeder} = \sqrt{\frac{c^3}{4V\bar{\delta}_\nu}} \tag{11.8}$$

is quasistochastic. The room transfer function cannot be used as a basis for evaluation of the quality of the room response. It can be shown that the specific fine structure of the stationary room transfer function contains nothing but an estimate of the reverberation time (see below).

Typical Schroeder frequencies are given in Table 11.1.

Because of the stochastic nature of the transfer function it is useless to apply wave models for this frequency range. The apparently exact result is not relevant for actual transient excitation signals because the stationary case cannot be excited with these too short signals. For

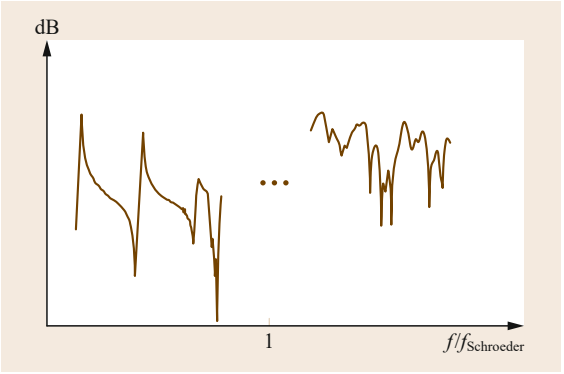


Fig. 11.1 Magnitude of a stationary room transfer function

Table 11.1 Typical Schroeder frequencies

Room	Volume (m^3)	$f_{Schroeder}$ (Hz)
Classroom	500	64
Auditorium	5000	32
Concert hall	50 000	16

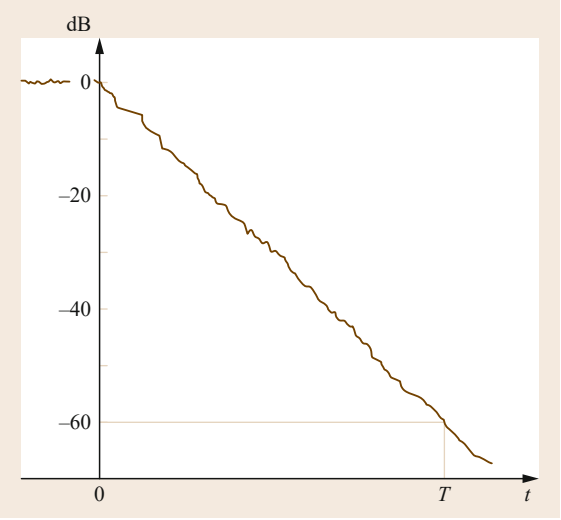


Fig. 11.2 Decay curve of a room

not too small rooms ($V > 500\,m^3$) these facts apply to basically the whole audio frequency range.

For interpretation of sound perception in rooms and for application in architectural design, another concept is required, which allows derivation of a set of quality metrics related to psychoacoustic effects in rooms such as reverberation, loudness, clarity, localization, spaciousness, etc. This is achieved by introducing a physical concept for the time domain and by interpreting the propagation of sound by geometrical propagation of rays rather than of waves.

Wallace Clement Sabine created the first scientific approach to understanding the acoustics of performance spaces. His famous *collected papers* [11.6] form the basis for room acoustics which we are still using today. The inter-relation between reverberation and volume and absorption was clarified and explained empirically. A little later, in the 1930s, *Norris* [11.7] and *Eyring* [11.8] presented a theory with more correct factors in the equations, which today, of course, are known as Sabine's and Eyring's reverberation time equations.

The reverberation time is defined as the time elapsed for a sound decaying by 60 dB after shut-off of a signal (Fig. 11.2).

With W denoting the acoustic power of a point source, the intensity I of a ray determined at a distance r is proportional to the squared sound pressure

$$I = \frac{W}{4\pi r^2} \propto p^2 \,. \tag{11.9}$$

With the assumption of incoherent rays, n of them add to

$$I = \sum I_n \,. \tag{11.10}$$

The assumption of incoherence would be violated in the case of stationary signals. Here, however, we assume broadband transient signals. Using this basic concept, statistical room acoustics can be introduced.

11.2 Statistical Room Acoustics

The basic assumption of statistical room acoustics is the existence of a *diffuse* field. In room sound fields this does not exist perfectly but the approximation is a very good guess. In a diffuse field, the angles of incidence at a receiver point are uniformly distributed in 3-D space. With this assumption it can be shown that the energy density at all positions in the room is constant, and all boundaries of the room (walls, floor, ceiling) are irradiated uniformly from all directions.

A ray with intensity I_0 is traced in a room with volume V and surface area S. It contributes to the energy density, w, with

$$e = e_0 = \frac{I_0}{c} \tag{11.11}$$

$$w = \frac{\sum e}{V} \,. \tag{11.12}$$

The expectation value of the time between two wall collisions is $1/\overline{n}$, with

$$\overline{n} = \frac{cS}{4V} \tag{11.13}$$

denoting the mean reflection rate.

11.3 Reverberation

For the derivation of the reverberation time formula we assume a room with uniformly distributed and directional-independent absorption. A diffuse field exists in the room. The sound field is modeled by rays filling the whole space, traveling at the speed of sound and bouncing between the boundaries (Fig. 11.3). Each ray initially has an energy e_0. At a boundary reflection it loses a portion of $(1-\alpha)$, so the remaining energy after the first reflection is $e_0(1-\alpha)$. Accordingly, after n reflections during time t the ray energy is

$$e(t) = e_0\,(1-\alpha)^{n(t)} = \frac{W}{4\pi c^3 t^2}\,(1-\alpha)^{n(t)}\,. \quad (11.14)$$

The expectation value of the number of reflections per time unit is (11.13)

$$n(t) = \overline{n}t\,. \quad (11.15)$$

The expectation value for the number of rays arriving at a receiver at time t during a time interval Δt is [11.9]

$$\Delta N = \frac{4\pi c^3 t^2}{V}\Delta t\,. \quad (11.16)$$

This leads to the energy density at a receiver point

$$w(t) = \frac{\sum e(t)}{V} = \frac{W\Delta t}{V}\,(1-\alpha)^{\overline{n}t}\,. \quad (11.17)$$

Re-arranging this equation for t and inserting, for the reverberation time T, a decay of 60 dB (i. e., a decay of $w(t)$ by a factor of 10^{-6}) leads to the famous Eyring reverberation formula

$$T = \frac{24\ln(10)}{c}\frac{V}{-S\ln(1-\alpha)} \approx 0.16\frac{s}{m}\frac{V}{-S\ln(1-\alpha)}\,. \quad (11.18)$$

For small absorption coefficients $\alpha \approx -\ln(1-\alpha)$ it can be simplified to

$$T \approx 0.16\frac{s}{m}\frac{V}{S\alpha}\,. \quad (11.19)$$

This equation is known as the Sabine equation. It holds for the large majority of ordinary rooms and performance spaces. Exceptions are spaces of complex geometry or any other special case in which no diffuse field can be expected. This may apply to coupled rooms, long or flat spaces, for example.

In the case of nonuniform absorption the reverberation formula must be modified by introducing the mean absorption coefficient or the equivalent absorption area, A, for N subsurfaces.

$$\overline{\alpha} = \frac{A}{S}, \quad A = \sum_{i=1}^{N} S_i\alpha_i \quad (11.20)$$

In the particle model, the sound can be attenuated by wall absorption but also by attenuation in the air. Sound traveling over a distance $x = ct$ is affected by an intensity loss of $\exp(-mct)$. Thus, the reverberation formula can be further extended to

$$T_{\text{Sabine}} \approx 0.16\frac{s}{m}\frac{V}{S\overline{\alpha}+4mV} = 0.16\frac{s}{m}\frac{V}{A_{\text{Sabine}}+4mV}\,, \quad (11.21)$$

$$T_{\text{Eyring}} \approx 0.16\frac{s}{m}\frac{V}{-S\ln(1-\overline{\alpha})+4mV} = 0.16\frac{s}{m}\frac{V}{A_{\text{Eyring}}+4mV}\,. \quad (11.22)$$

It should be noted, however, that the reverberation formula in Sabine's or Eyring's notation are approximations. The underlying assumption of a diffuse field is a theoretical model, which exists in rooms only in approximation. The interpretation of the equivalent absorption area A also leads to the possibility of considering it in relation to the mean logarithmic reflection loss, to the linear reflection loss, or just to the term in the denominator which expresses the losses associated with the boundaries. This factor A also appears in the solution of the sound level in stationary excitation.

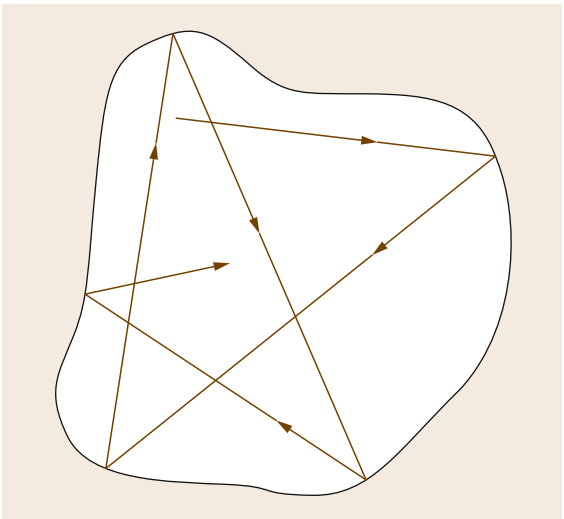


Fig. 11.3 Free paths between subsequent reflections

11.4 Stationary Excitation

In the case of a stationary sound power source the exponential energy propagation over time (11.14) must be integrated in order to calculate the total energy in the room. The energy balance is given by the steady state between the energy injected by the power source and the energy losses in the boundaries and the air. In this way the integration can be performed easily, and after normalization with the source sound power W it yields

$$w_{\text{diffuse}} = \frac{4W}{cA} e^{-\frac{A}{S}} . \tag{11.23}$$

This energy density can be compared with the energy density of the direct sound w_{direct}, observed at a distance r from the source

$$w_{\text{direct}} = \frac{W}{4\pi c r^2} . \tag{11.24}$$

The distance in which the direct and the diffuse field are equal is called the reverberation distance

$$d_{\text{rev}} = \sqrt{\frac{A}{16\pi}} . \tag{11.25}$$

The basic findings illustrate the influence of room volume and absorption on the perceived sound. This is actually the basic experience while listening in a room. Humans can estimate the room size and the state of the room interior (fully furnished, empty, etc.) very well. When it comes to more specific aspects of sound perception in rooms, however, the sound incident on the listener must be studied in more detail and in particular in dependence of the position of the source and the receiver.

11.5 Room Impulse Responses

In the 1950s, the *fine structure of reverberation* became the focus of interest. *Lothar Cremer*, in the first edition of his pioneering book [11.10], illustrated sound reflections using geometric constructions of rays and image source, a methodology which is still one of the standard methods in room acoustics today. He had identified the importance of reflections, their series of arrival, their density, and their global late decay. With these findings and with the availability of instrumentation for impulse response measurement, acoustic consulting was put on a scientific basis. From the 1950s, consulting firms could rely on a deeper understanding of sound fields and could advise architects accordingly. The basis is the room impulse response as illustrated in Fig. 11.4.

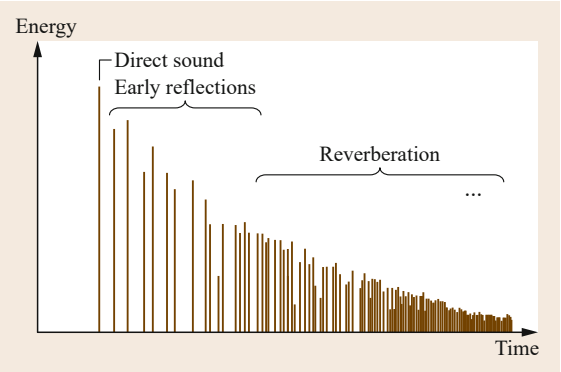


Fig. 11.4 Typical energy room impulse response

According to the concept of geometrical acoustics the reverberation is generated by subsequent sound reflections at the boundaries. A human listener in the room therefore perceives not just the direct sound but a series of delayed and attenuated reflections. They are delayed due to detours through the room and they lose energy due to the spread of the spherical wave behavior and to boundary absorption and air attenuation.

If the sound source emits a very short pulse, the room response contains a series of pulses, the density of which quadratically increases with time while their strength decreases quadratically in time and exponentially due to repeated absorption. Eventually, the energy decay integrated over short time intervals is an exponential function (11.17).

Concerning perception, the first component (the direct sound) determines the source localization. This effect is called the *law of the first wavefront* [11.10], or the *precedence effect* [11.11]. The reflections that arrive not later than 50–80 ms after the direct sound are perceptually very important. They support the loudness, a certain source widening by localization blur and they create the sensation of clarity in music perception. The reflections that arrive later than 80 ms are interpreted as reverberation. They create an impression of the room size and of the so-called listener envelopment (LEV; spatial impression). A strong reflection that significantly exceeds the exponential trend of the reverberation tail is considered an *echo*, which is usually a dramatic room defect.

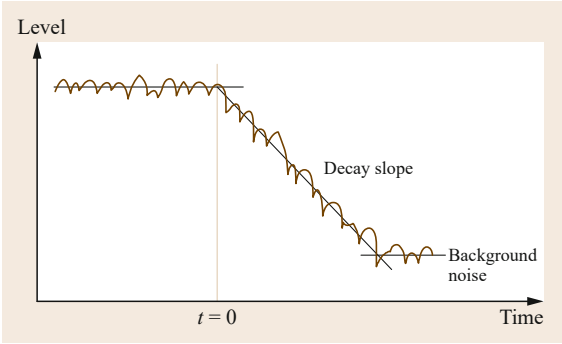


Fig. 11.5 Typical level versus time curve by using a switch-off of a random noise signal

The function of energy versus time depends on the positions of the source and the listener. The exponential envelope of the reverberation tail may be constant throughout the room but this does not at all hold for the early part.

11.5.1 Room Acoustic Measurements

In classical reverberation time measurement a noise source (loudspeaker excited with random noise) is switched on for a time sufficient to obtain an excitation in steady state. Then the source is switched off, and the decay of the sound level is observed.

As seen from Fig. 11.5, the stationary level (11.23) and the decay time (11.18) are obtained from the stationary excitation level and from the slope after switch-off. Statistical signal characteristics require averaging over several excitations until a stable and smooth mean decay curve appears.

In modern digital measurement methods, however, it is possible to obtain the room impulse response as a primary deterministic result. The reverberation time is calculated by postprocessing (Fig. 11.6). This approach requires more sophisticated equipment and signal processing but it is much more effective and theoretically more elegant.

According to *Schroeder* [11.12], the expected decay resulting from an infinite number of averages in one particular observation point may be obtained by processing the impulse response between the excitation signal (loudspeaker) and the observation point (microphone) directly and without tedious averaging.

This situation can be accounted for with a known impulse response $h(t)$, between source and receiver. When a room is excited by stationary white noise for a time sufficient to obtain stationary conditions and the noise is thereafter switched off at time $t = 0$, the ex-

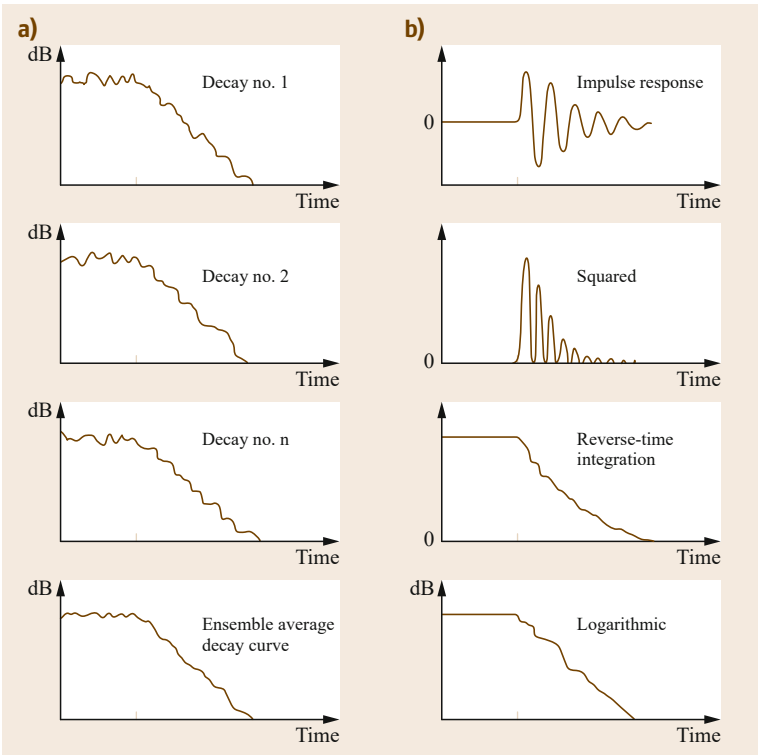


Fig. 11.6a,b Classical versus impulse response method. **(a)** In the classical method the expected decay is found by ensemble averaging of a number of individual decays based on noise excitation. **(b)** By application of the impulse response method the expected decay is found by postprocessing of the impulse response

pected level at any time $t > 0$ will be

$$L(t) = 10 \log \left[N_0 \int_t^\infty h^2(\tau) \mathrm{d}\tau \right], \qquad (11.26)$$

where N_0 is a constant specifying the signal power per unit bandwidth of the excitation signal.

Various methods may be applied to obtain the impulse response $h(t)$ or the transfer function $H(\omega)$ (which is linked to the impulse response by Fourier transformation). All such methods may be used if they are able to demonstrate reliable results within normal measurement conditions, i. e., conditions that are linear and time-invariant.

When an octave-band or fractional-octave-band filter is a part of the measured system, (11.26) will describe the expected decay according to the classical method for the applied filter band. Accordingly, the level plot can be evaluated for determination of the reverberation time.

The dynamic range necessary for reverberation time measurement is very high. As for level measurements, the upper limit for the time integral in (11.26) shall be limited in order to reduce the contribution from unwanted noise.

$$L(t) = 10 \log \left[N_0 \int_t^{t_2} h^2(\tau) \mathrm{d}\tau \right] \qquad (11.27)$$

Usually t_2 is set to the time where the level of the exponential decay in the impulse response $h^2(t)$ is equal to the extraneous background noise level. Different methods are described in the literature to compensate for the noise and the truncation in the integration interval [11.13]. If an evaluation range of 30 dB is used, and T30 is to be measured, the decay curve must be evaluated from linear regression between an upper limit of −5 dB and a lower limit of −35 dB. To add a safety margin and to avoid a truncation error, −40 or −45 dB dynamic range is required to ensure an undistorted decay curve down to −35 dB. Other definitions for reverberation time measurements are T20, T40, etc.. The early decay time (EDT) is evaluated from the initial part of the decay between 0 and −10 dB.

With strategies for subtracting the noise floor it is possible to extend the evaluation range. Estimation of the noise floor energy C can be performed by taking the first part of the impulse response, before the direct sound, or the late part which should show a constant short-time average energy. In this case, (11.27) changes to

$$L(t) = 10 \log \left[N_0 \int_t^{t_2} h^2(\tau) \mathrm{d}\tau + C \right], \qquad (11.28)$$

with C denoting the energy of the noise contribution. This procedure, however, requires manipulation of the impulse response with some arbitrarily chosen parameter. If the noise content remains in the measured response, backward integration will lead to a curved decay of two slopes. But even in this case the correct decay slope can be extracted, by using a nonlinear regression, as shown by *Xiang* [11.14].

All measurement methods can be based upon a common scheme of impulse response and transfer function determination. In the following part of this chapter, the fundamentals of modern impulse and correlation techniques are described.

The room is excited by a deterministic signal $s(t)$, and the impulse response $h(t)$ is calculated from the response to this excitation, $g(t)$. In modern digital methods the excitation signal is usually distributed over a longer period of time to facilitate high total energy. This procedure will enhance the dynamic range and reduce the influence of extraneous noise. Most of the methods may be described as FFT (Fast Fourier Transform), de-convolution, or correlation techniques applied to the measurement and processing of the impulse response. Usually results in building acoustics are expressed in one-third octave or octave bands. These filters can be included in the excitation signal, or in the signal processing of the received response.

The signal path expressed in the time domain reads (Fig. 11.7)

$$g(t) = s(t) * h(t) = \int_{-\infty}^{\infty} s(\tau) h(t-\tau) \mathrm{d}\tau \,. \qquad (11.29)$$

Exactly the same expressed in the frequency domain reads

$$G(\omega) = S(\omega) \cdot H(\omega) \,. \qquad (11.30)$$

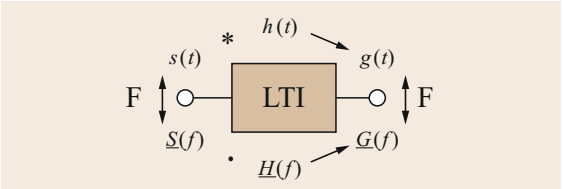


Fig. 11.7 Signal transmission through a linear time-invariant (LTI) system

While the key equation of a formulation in the frequency domain is

$$H(\omega) = \frac{G(\omega)}{S(\omega)} , \tag{11.31}$$

the same can be expressed in the time domain by so-called *de-convolution*

$$h(t) = g(t) * s^{-1}(t) \tag{11.32}$$

with $s^{-1}(t)$ being the signal with the inverse spectrum $1/S(\omega)$. $s^{-1}(t)$ is also called the *matched filter*. If $S(\omega)$ is a white spectrum, (11.32) can also be re-arranged into

$$h(t) = g(t) * s(-t) = \int_{-\infty}^{\infty} g(\tau)s(t+\tau)\mathrm{d}\tau , \tag{11.33}$$

which means that $h(t)$ is obtained from crosscorrelation of $s(t)$ and $g(t)$.

Obviously, the three equations above are absolutely equivalent for broadband white excitation signals. Differences, however, can be found in characteristics of the signal in relation to the capability of the components of the measurement chain and in relation to sensitivity against nonlinear distortions of time variances.

It is worth mentioning that the equations must be programmed in discrete form and be represented in the memory of a computer with a finite length. The integration will be read as summation, and the limits of the summation are finite. Accordingly, the excitation signal must have a certain length: a period. Now, if a repetitive excitation signal is used, the spectrum of the excitation will consist of narrow spectral lines where the distance between adjacent lines Δf will be given as the inverse of the time for one repetitive period T_{rep}

$$\Delta f = \frac{1}{T_{\mathrm{rep}}} . \tag{11.34}$$

In order to ensure that all modes of the room are excited the repetitive period must not be shorter than a certain fraction of the reverberation time for the room being measured. If, for example, the decay time of the impulse response is significantly larger than one period of the signal the tail of the impulse response appears as equivalent noise overlap in the subsequent periods. This effect is called *time aliasing*, and is similar to *frequency aliasing*, which happens if spectra are measured with too low a sampling rate. To be on the safe side, $T_{\mathrm{rep}} > T$ should be chosen.

The periodic nature of the source signals, however, offers a very big advantage. A repetition of the signal should give exactly reproducible results, measured and compared from one period to the next. This allows for the application of synchronous averaging, a procedure which reduces the effective background noise level by 3 dB per doubling of the number of averages. The reason is that the exactly repeated periods of the test signal add up *in phase* while the background noise is not correlated between the different periods and adds up energetically only. This averaging process is an advantage of deterministic signals in general. The gain in signal-to-noise ratio, with N denoting the number of averages, is $\Delta_{\mathrm{av}} = 10 \log N \mathrm{dB}$.

11.5.2 Digital Measurement Techniques

The three methods for determination of impulse responses expressed in (11.31)–(11.33) are described in more detail in this section.

The *FFT technique* is based on spectral division. There are two variants, the first of which uses any excitation signal, possibly from an arbitrary stationary noise source, and the other uses a deterministic source signal which is synchronized with the A/D sampling. Some traditional FFT analyzers operate with asynchronous noise sources.

With an asynchronous signal source the complex transfer function defined by (11.31) can be computed using the cross- and autocorrelation spectra between input and output channel, for instance from

$$H(\omega) = \frac{S^*(\omega)G(\omega)}{S^*(\omega)S(\omega)} = \frac{\Psi_{SG}(\omega)}{\Psi_{SS}(\omega)} \tag{11.35}$$

With deterministic and repeatable excitation, however, the measurement is straightforward and uses the formulation in the frequency domain. The input and output signal are measured simultaneously, transformed by FFT, and processed by complex division to obtain $H(\omega)$ (11.31). The most important requirement is that the source signal must have sufficient energy in the whole frequency range. Weak signal-to-noise ratios may cause errors in the spectrum division. Any broadband sweep, time-stretched pulses, or pseudorandom noise can be used.

After processing the spectrum division in the frequency domain, the impulse response is calculated finally by inverse Fourier transformation (Fig. 11.8).

The construction of a constant-envelope sweep with arbitrary spectral distribution, for instance, is possible when the group delay's inclination, or in other words, its derivative, is proportional to the expected energy $|H(\omega)|^2$ at each frequency [11.15]. For a discrete FFT spectrum, this means that the group delay has to increase according to

$$\tau_{\mathrm{G}}(\omega) = \tau_{\mathrm{G}}(\omega - \mathrm{d}\omega) + C \cdot |H(\omega)|^2 \tag{11.36}$$

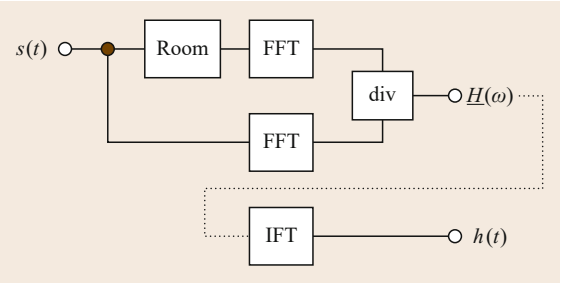


Fig. 11.8 Typical 2-channel FFT instrument

following a recursive structure of $\tau_G \to \tau_G(\omega - d\omega)$ in the FFT spectrum, and C denoting a normalization constant composed of the sweep's length divided by its total energy (ω_s is the sampling frequency)

$$C = \frac{\tau_G(\omega_{end}) - \tau_G(\omega_{start})}{\int_{\omega=0}^{\omega_s/2} |H(\omega)|^2} . \tag{11.37}$$

This type of signal is known from traditional sweep measurements and the tuned narrow-band filter. However, processing is much more sophisticated compared to the traditional narrow-band technique and it is possible to obtain the complex transfer function and accordingly the impulse response. The result is processed by a *matched filter* for de-convolution in the time domain (11.32). The matched filter is the excitation signal in reverse order

$$s^{-1}(t) = s(T_{rep} - t) . \tag{11.38}$$

An important point and a significant advantage of the sweep is that the problem of time aliasing can be significantly reduced by cyclic shift of the signal and extension by inserting *zeros*. The matched filter is then treated accordingly in the reversed way. Furthermore, *Müller* [11.15], *Farina* [11.16], and *Dietrich* [11.17] discussed the fast simultaneous measurement of the linear transfer function and nonlinear indicators like distortion products. Procedures for signal processing in the swept-sine technique can also be formulated in the frequency domain. If the matched filter convolution is performed by FFT, this technique is similar to the 2-channel FFT technique, but the differences (and disadvantages) can be found in the periodicity and the cyclic processing.

Another technique, the maximum length sequence (MLS) technique, became popular in the 1990s. The technique is a a special form of correlation measurement for obtaining impulse responses of LTI systems according to (11.32). The room is excited by a stationary MLS signal $s(t)$, and the output signal $g(t)$ is processed by crosscorrelation with the excitation signal.

Maximum length sequences are a special kind of correlation signal useable in the technique described above. The sequences are periodic binary pseudostochastic signals with an autocorrelation function almost identical to a periodic Dirac pulse. They are generated by use of a shift register with feedback loops. The exact feedback procedure is given by number-theoretical prescriptions [11.18].

The procedure of the MLS technique can be summarized as follows. The LTI system is excited with a stationary periodic sequence. In order to avoid time aliasing the period must be larger than the impulse response of the system as given, for instance, by the reverberation time. Crosscorrelation is used to extract the impulse response from the measured sequence.

$$\begin{aligned} &[s_{MLS}(t) * h(t)] * s_{MLS}(-t) \\ &\quad = s_{MLS}(t) * s_{MLS}(-t) * h(t) \\ &\quad = \Psi_{ss}(t) * h(t) . \end{aligned} \tag{11.39}$$

The important advantage of maximum length sequences is the availability of a fast correlation algorithm, the so-called *fast Hadamard transformation*, FHT. More elaborate investigations are found in *Rife* et al. [11.18], and practical aspects are discussed in [11.19].

11.5.3 Perception-Based Parameters Obtained from Impulse Responses

The subjective impressions were correlated with the physical properties of room impulse responses. One of the first observations concerning early reflections was made by *Thiele* [11.20] in 1952, who provided the basis for objective descriptors of the early-to-late energy integral ratio (*Deutlichkeit*). This was finally possible with the availability of multichannel equipment in anechoic rooms. Today, we consider the concept of early decay time as well known. This finding was given by *Vilhelm Jordan* [11.21] who discovered the relationship between early reverberation and subjective reverberance. Furthermore, during this period there were also studies on correlation of subjective impressions and objective parameters. Numerous publications appeared from research groups led by *Erwin Meyer* et al. [11.22–26].

In the following some of the quantities are introduced which were tested in laboratory environments, and applied in measurements in halls. The integrals are taken over the sound pressure impulse response $h(t)$, of the room with regard to particular source and receiver positions. It should be noted that the squared response $h^2(t)$, is proportional to the energy density.

Strength

$$G = 10\log\frac{\int_0^\infty h^2(t)\mathrm{d}t}{\int_0^\infty p_{10\,\mathrm{m}}^2(t)\mathrm{d}t} \tag{11.40}$$

with a reference to the sound pressure of the sound source used in a free field in 10 m distance, p_{10}^2.

Speech Intelligibility: Deutlichkeit

$$D = \frac{\int_0^{50\,\mathrm{ms}} h^2(t)\mathrm{d}t}{\int_0^\infty h^2(t)\mathrm{d}t}\,. \tag{11.41}$$

Transparency of Music, Clarity

$$C_{80} = 10\log\frac{\int_0^{80\,\mathrm{ms}} h^2(t)\mathrm{d}t}{\int_{80\,\mathrm{ms}}^\infty h^2(t)\mathrm{d}t}\,. \tag{11.42}$$

Apparent Source Width (ASW)

$$LF = \frac{\int_5^{80\,\mathrm{ms}} h_\infty^2(t)\mathrm{d}t}{\int_{0\,\mathrm{ms}}^{80\,\mathrm{ms}} h^2(t)\mathrm{d}t}\,, \tag{11.43}$$

with p_∞^2 denoting the squared sound pressure impulse response obtained with a gradient (dipole) microphone.

Listener Envelopment, Late Lateral Strength

$$LG = 10\log\frac{\int_{80}^\infty h_\infty^2(t)\mathrm{d}t}{\int_0^\infty p_{10\,\mathrm{m}}^2(t)\mathrm{d}t}\,. \tag{11.44}$$

11.5.4 Music Perception and Architectural Design

At the beginning of the 1970s many modern concepts for *good* room acoustics related to performance of classical music, opera, and speech in theaters were proven in theory and accepted as best practice. At that time, however, many halls were designed with modern architecture and construction methods. Usage of open concrete, mostly large flat and smooth surfaces, large and wide halls, particularly fan-shaped halls were preferred. And some of these halls did have their critics, although reverberation time, sound level, and clarity were found to be around the right order of magnitude.

James West [11.27] stated in 1966 that concert halls with good acoustics (Musikvereinssaal Vienna and Boston Symphony Hall) have a small width-to-height ratio. In 1968, *Harold Marshall* [11.28] confirmed that spatial impression is created by side wall reflections which are particularly strong in narrow halls. In the early 1970s, *Michael Barron* [11.29] in Southampton and *P. Damaske* and *Yoishi Ando* [11.30] in Göttingen identified the importance of lateral reflections as being underestimated. They pointed out the relevance of early lateral reflections for spatial impression. It affects the precision of source localization and gives an impression of diffuse sound incidence. Spatial impression still today is the most difficult component of multidimensional hearing in rooms (*Hörsamkeit*).

A fan-shaped hall shows indeed a lack of early lateral sound. A consequence was to implement side walls with segments of tilted angles, terraces, or vineyard areas, in order to split side wall reflections and to direct them to the audience at lateral angles.

After the 1970s, acousticians and architects could rely on quite stable and complete knowledge of the general principles of room shape and its effect on early and late reflections. The first issue of the book by *Heinrich Kuttruff*, published in 1972, already contained state-of-the-art knowledge of room acoustics [11.31]. Thereafter, although further detail could be added the general insight into room acoustics was complete.

It is interesting to note that the findings in room acoustic research were directly reflected in the hall shapes built. *Jürgen Meyer* [11.32] evaluated 157 halls with a seating capacity > 1000. He pointed out that the majority of room shapes were rectangular, fan-shaped, hexagonal, or arena style. Thus, hall shape has a large bearing on acoustic performance. Many details such as surface corrugations, stage, balcony and reflector design are important as well, as well as consideration of the directional radiation of musical instruments.

Finally, an extended issue of the international standard ISO 3382, *Measurement of the reverberation time of rooms with reference to other acoustical parameters* was developed and published in 1996 [11.1]. It not only contained guidelines for measuring reverberation time in auditoria but definitions of the other room acoustical criteria listed above.

11.6 Computers in Room Acoustics

The introduction of computers in acoustic measurements enabled the use of sophisticated impulse response measurement techniques for both real-room measurements and scale models. Impulse responses could be measured with high reproducibility and, thus, comparability. The first ISO 3382 already contained a hint that broadband excitation and postprocessing should be applied to obtain impulse responses [11.1].

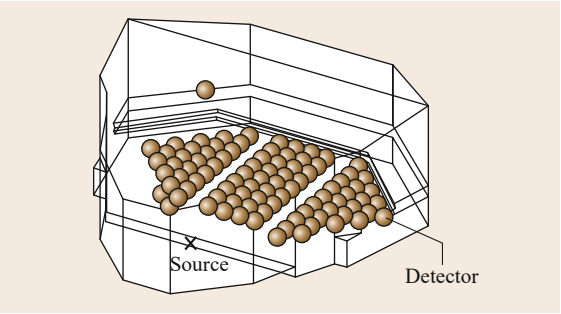


Fig. 11.9 Computer model of a room with a source (×) and a set of receivers (*brown spheres*)

Use of computers in room acoustics progressed greatly after 1985, during a period corresponding to the development of simulations. This development brought them to a level that allowed their application in consulting. Independently, *Dirk van Maercke* in Grenoble [11.33] (the first software listed was originally developed as EPIKUL at KU Leuven), *Michael Vorländer* in Aachen [11.34], *Bengt-Inge Dalenbäck* in Gothenburg [11.35], and *Graham Naylor* and *Jens Holger Rindel* [11.36] in Copenhagen worked on algorithms for simulating room acoustics on the basis of architectural CAD input data. These and other programs are still being used with success today. Round-robin tests [11.37–39] showed their efficiency and as well the limits of computer simulation of room acoustics. This led, among other findings, to the definition and implementation of scattering coefficients [11.40] in simulation methods. Similarly to observations in real sound fields with purely specular reflections (rooms with large flat concrete walls), it was noted that artificial decays based on purely specular models sound unnaturally cold and contrasted.

In room acoustic computer simulation a geometric model of the room is built using polygons (Fig. 11.9). The polygons are the portions of planes which form the room boundaries. Sound sources and receivers are considered to be small compared to wavelengths. The computer algorithm consists of equations of linear algebra for solving the problem of intersection of rays (vectors) with polygons (planes). The method is used for calculation of ISO 3382 criteria such as reverberation time (T), early decay time (EDT), definition (D), clarity (C80), center time (CT), lateral energy fraction (LF), or for auralization. The latter means that the computer results are reproduced binaurally through the D/A converter in the sound card.

Two principles are typically implemented, ray tracing and image sources. They are complementary in their pros and cons. Today, simulated room acoustics are applied in various fields with great success. Their well-developed algorithms help to create realistic acoustics during architectural planning. Acoustic simulation tools are also used for designing sound reinforcement systems in churches, stadiums, train stations, and airport terminals.

It is important to highlight the differences in the physical meaning of ray tracing and image sources. Ray tracing describes a stochastic process of particle radiation and detection. Image sources are geometrically constructed sources that correspond to specular paths of sound rays. Often, image sources are constructed by using rays, beams or cones, via a kind of ray tracing. Nevertheless, they are still *image source models*. The fundamental difference between image sources and ray tracing is the way contributions in impulse responses are calculated. Ray tracing only yields low-resolution impulse response data like envelopes in spectral and time domains. Image sources (classical or via tracing rays, beams, cones, etc.) may be used for exact construction of amplitude and delay of reflections with narrow-band resolution depending on the filter specifications for wall reflection factors, for instance.

11.6.1 Image Sources

We assume smooth walls which reflect the sound waves specularly with a reflection coefficient of one. In this case the sound paths between sources and receivers are unique solutions and use virtual (image) sources. This approach also solved the wave equation. In the case of reflection coefficients < 1 the approach is a quite good approximation. The main problem in this case is the assumption of plane-wave reflection coefficients in contrast to the spherical wave reflections in the precise wave approach. After all, the approximation is sufficiently correct if the source and the receiver are not closer to the boundaries than a few wavelengths.

Firstly, for a given room they must be constructed. The original source is mirrored at all wall polygons. In Fig. 11.10, the image source S_1 is created by constructing a perpendicular line between the source, S, and wall 1 and by continuing to the same distance behind the wall. The same is done for all walls. This yields the first-order image sources. With those the mirroring is continued to achieve the second-order image source, and so on. The order of an image source deflects the order of the reflection, i. e., the number of walls which were hit during the propagation between the sound source and the receiver.

Secondly, an audibility test must be performed. The result depends on the receiver position, and it is required for identification of the meaningful image sources. Each source is considered as the last element of a chain of image sources. At its indices we can identify

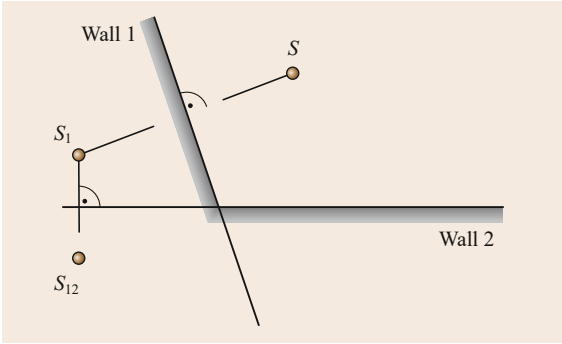


Fig. 11.10 Image source construction

the walls which were sequentially hit during the sound propagation in the room. This sequence must be physically consistent so that free lines of sight are present between the subsequent points on the polygons where the walls are hit.

A sequence of an i-th-order image source has i indices

$$S \to S_{n_1} \to S_{n_1 n_2} \to S_{n_1 n_2 n_3} \to \ldots \to S_{n_1 n_2 \ldots n_{i-1}} \to S_{n_1 n_2 \ldots n_i} \,. \tag{11.45}$$

The possible number of permutations and thus the image source number, N_{IS}, to be constructed for a room with n_{w} polygons is

$$N_{\mathrm{IS}} \approx (n_{\mathrm{w}} - 1)^{\overline{n}t} \tag{11.46}$$

with $\overline{n}t = i$ denoting the order of image source involved up to a time t in the impulse response (11.15). Of this extremely large number of image sources, however, just a very small portion is actually audible. To check this, the audibility test is introduced as follows (Fig. 11.11). We consider an example in a 2-D plane.

We start the process by checking S_{12} exemplarily. The receiver point, R, is connected to S_{12}. Its last index (no. 2) denotes the last wall (no. 2) involved in the sequence of reflections. If the intersection point, P_{12}, of the straight line connecting $R - S_{12}$ is located in the wall polygon no. 2, it's correct. Then we proceed by connecting the intersection point with the mother source of S_{12}, i. e., S_1 (the one with the same set of indices except the last). If the intersection point, P_1, of the straight line connecting $P_{12} - S_1$ is located in wall polygon no. 1, it's correct. As soon as one intersection point is located outside the polygon, the source is found to be inaudible. This applies in the example to S_{21}, because P_{21} is not inside polygon 1. This means that this reflection path is blocked by the room edges or by an obstacle in the room (in the case of concave rooms).

Accordingly, the core subroutine of the image source algorithm is the point-in-polygon test. It is crucial that the audibility is checked for the actual receiver

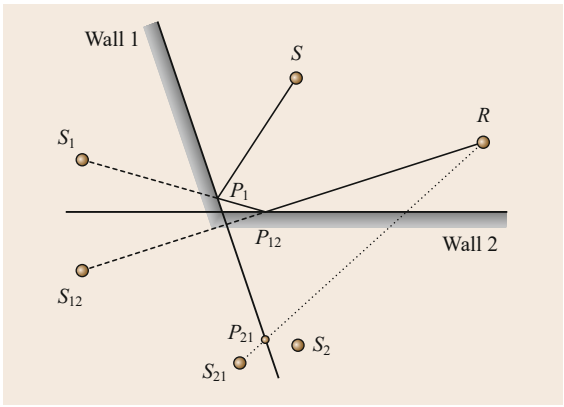


Fig. 11.11 Image source audibility test for the sources S_{12} and S_{21}. S_{21} is inaudible at R

position. Generally, image source cannot be excluded, except for constraints in the room geometry and imaging zones built by the polygon edge [11.41].

The audible image sources are postprocessed to account for their contributions to the impulse response in terms of amplitude and delay. The following factors influence the contribution of the image source number j (j is the sequence of wall reflections represented by $n_1 n_2 \ldots n_i$, see (11.45)).

$$\underline{H}_j = \frac{\mathrm{e}^{-jkr_j}}{r_j} \underline{H}_{\mathrm{source}}(\theta, \phi) \underline{H}_{\mathrm{air}}(r_j) \times \mathrm{HRTF}(\vartheta, \varphi) \prod_i^{n_j} \underline{R}_i \tag{11.47}$$

In this equation, $\underline{H}_j$ is the complex spectral component of the image source j. It contains the delay expressed in phase lag, the amplitude of spherical wave spread, the source spectrum expressed by the sound pressure radiation in source coordinates (θ, ϕ), the air attenuation over the traveled distance r_j, the spatial listener sensitivity expressed by the head-related transfer function in receiver coordinates (ϑ, φ), and the multiplied complex reflection coefficients $\underline{R}_i$ of the walls involved. Usually, however, the reflection coefficients are considered to be real-valued and taken from $R = \sqrt{1-\alpha}$.

The impulse response obtained by using image sources may look like the example shown in Fig. 11.4. The temporal resolution in the impulse is just limited by the precision of the distance between the image source and the receiver, which is a pure construction of straight lines. The accuracy in the impulse response in terms of the temporal series is very high. The accuracy of the amplitudes of the reflections depends on the input data used for the walls and the medium, the source, its directivity, and the listener head-related transfer function (HRTF).

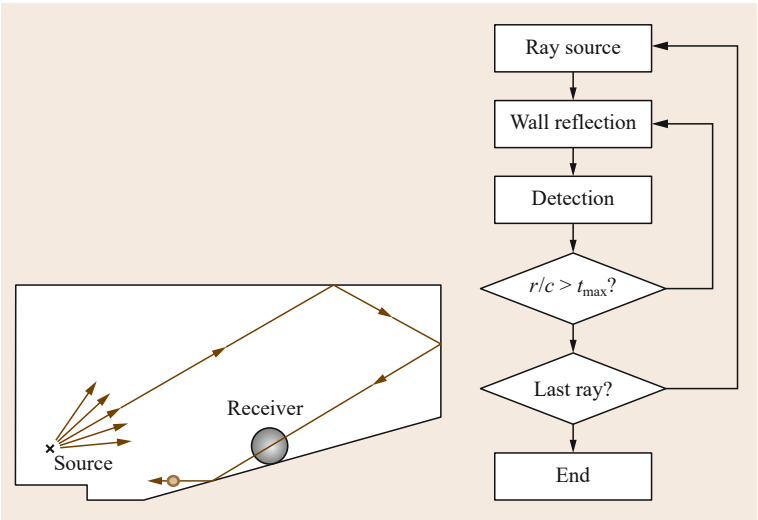


Fig. 11.12 Ray-tracing example and algorithm flow diagram

However, because of the problem of the rapidly growing computational effort, the image source algorithm is applicable to low reflection orders only. The long computation time means it is not useful for calculation of the reverberation tail. A very serious shortcoming is the prerequisite of pure specular reflections. So, the image source algorithm is a good choice for calculating the early specular part of the impulse response precisely.

11.6.2 Ray Tracing

Obviously there remains the problem of calculating the late part of the impulse response and the scattered and diffracted components. A well-established method in this respect is stochastic ray tracing (Fig. 11.12). The goal is to estimate the impulse response envelope over short time intervals without aiming at the exact fine structure formed by reflected, scattered, and diffracted sound. This approach is justified in the regime of reflection overlap within reasonable intervals of human temporal perception abilities. This statement simply means that humans cannot perceive the specific fine structure, so that an approximation by a statistical approach is sufficient. It is crucial, however, that the temporal, spectral, and spatial envelopes are correct. The corresponding averages are usually taken over 10 ms, critical frequency (or one-third octave) bands, and 5°, respectively.

In ray tracing, it is assumed that a sound source emits rays or particles at $t = 0$. In the following, the concept of a *particle* is used, which is in line with the wave-particle dualism of wave physics, too. It is crucial that numerous particles are emitted, and the results obtained are interpreted as a statistical mean of events. Events in this respect are reflections and detections are receivers, for example. Each particle is traced during its run through the room. The mathematics for this is quite similar to that described above for image source: vector algebra and intersections of vectors with planes are the basis, and the point-in-polygon test is the most relevant part.

Each time a particle hits a wall polygon, the collision point and the angles of the incident vector with the plane are calculated. Then the reflection is modeled by inverting the vector direction following a specular reflection, or by a random scattering process. In the latter case Lambert's law is often applied, which states that independent of the incidence angle the reflection angle is distributed according to a cosine function.

Absorption is modeled by reducing the energy by a factor of $(1-\alpha)$. The stop criterion of tracing one particle is defined by the maximum time in the impulse or by using a minimum energy. Absorption modeling can also be implemented by randomly deleting the particle if a random number $z \in [0, 1]$ is smaller than the absorption coefficient α. After the stop criterion is reached or the particle is deleted, the next particle is emitted from the source. This way the sound radiation into the full space, for example omnidirectionally, represented by some 10 000 particles is modeled. The particles are registered at receivers by counting the energy and arrival time.

The temporal resolution is limited by the fact that a finite number of particles are traced. Late reflections are thus represented just by a few particles, or just by one. The probability of later reflections represented by a particle is less than one. Therefore, the impulse response is not meaningful but randomly sparse. This is, by the way, a phenomenon similar to the effect of background noise limiting the meaningful temporal range of the impulse response.

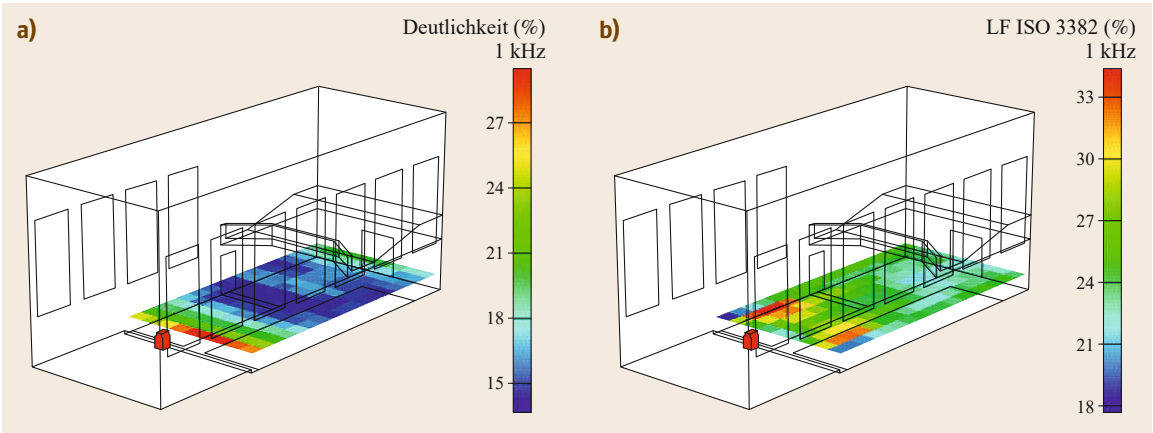


Fig. 11.13a,b Postprocessing of ray-tracing results into color maps of the ISO 3382 criteria D50 **(a)** and lateral energy fraction **(b)**

11.6.3 Hybrid Models

Because of the contradictory advantages and disadvantages of ray tracing and image sources an effort was made to combine the advantages in order to achieve high-precision results without too much complexity or computation time. Either ray tracing or similar algorithms such as radiosity were used to overcome the extremely high calculation time inherent in the image source model for simulation of the late part of the impulse response (adding a reverberation tail), or ray tracing was used to detect audible image sources in a kind of *forward audibility test*. The idea behind this is that a ray, beam, or cone detected by a receiver can be associated with an audible image source. The order, the indices, and the position of this image source can be reconstructed from the ray's history by storing the walls hit and the total free path. Hence the total travel time, the direction and the chain of image sources involved can be addressed to the image source. Almost all other algorithms used in commercial software are in a way dialects of the algorithms described above, and they differ in the way the mixing of the specular with the scattered component is implemented. The specific choice of dialect depends on the type of results wanted, particularly in terms of accuracy, spatial, and temporal resolution.

11.6.4 Wave Models

The boundary element method (BEM) and the finite element method (FEM) are also well-known algorithms in acoustics. In FEM, the field space for the acoustic problem must be discretized into suitable volume elements. Such mesh models are also available in the time domain, for example the finite-difference time-domain method (FDTD). Recent algorithmic implementations made use of the fact that FDTD requires updates of one mesh node at the actual time step just from the neighboring nodes at the last time step. Accordingly, the processing can be parallelized to a great extent, and the advantage of parallel computing leads to acceleration of the simulation by orders of magnitude.

The same speed increase was demonstrated for the acoustic radiance transfer method, which is based on discretization of the room boundaries and formulation of the integral radiation functions between the surface elements [11.42]. This method is very similar to the analytic approach of the energy radiosity method [11.31], but it can be extended to account for reflection phases and diffraction effects. Furthermore, significant parallel processing can be used due to the basically independent computation of algorithmic subroutines.

Actually high-speed wave model algorithms can exceed the Schroeder frequency. But the physical relevance of wave modeling above the Schroeder frequency is doubtful due to the stochastic nature of the pressure field. Therefore, all algorithms described above yield very similar impulse responses when it comes to perception.

The perception of room acoustics can best be assessed by listening to the results rather than by comparing color maps of ISO 3382 criteria (Fig. 11.13).

11.7 Auralization

The principle of auralization is illustrated in Fig. 11.14. It shows the basic elements of sound generation, transmission, radiation, and reproduction [11.43].

As soon as the interface between source signal and the propagation path is clearly defined, the acoustic situation can be projected into a model of signal processing. Provided the transfer functions of the elements are known by calculation or measurement, the signal transmitted in the room is processed by convolution (Fig. 11.15). The transfer function is accordingly the transfer function of a *filter*, and its representation in the time domain is nothing but the room impulse response $h(t)$.

This looks simple at first sight, but the task involves creation of a filter representing as many as possible features of the real signal path or multiple paths. To illustrate this task in more detail, some examples are given in the following sections. It is clear, however, that the requirements on the signal representation (bandwidth and sampling rate, for instance), on the accuracy of the data given, on the transmitting system, and on the quality of the audio reproduction system might be quite specific. Perception of sound signals has multiple dimensions, some of which are listed below:

- Kind of sound generation
- Direction of the event
- Movement of the source
- Own movement, environment (kind, shape, and size of the room).

The auralization filter must cover all relevant aspects of the specific case.

Impulse responses for auralization purposes must be generated with a temporal resolution adequate for the sampling rate of the test signals. A standard audio sampling rate of 44.1 kHz is usually used. In room acoustics, music or speech may be sampled at half that rate which still enables signals to be used with frequency components up to 10 kHz. This upper limit is also typically the limit of knowledge of absorption coefficients. Accordingly, the time resolution of the impulse response is less than 50 µs, which illustrates the demands on the simulation. Conventional ray tracing cannot be used since the statistical approach requires temporal histograms. The number of rays would be much too high if the (very small) time intervals must be filled with reflections. Conversely, image source algorithms are well qualified, at least up to a certain length of the impulse response.

It is clear that auralization involves binaural hearing. It's true that monoreproduction gives an impression of the simulated sound, but a very important feature of hearing in rooms is then omitted. In binaural technology, interaural delay and diffraction around the head are elementary parameters. Assuming linear sound propagation, both are included in the HRTFs or in corresponding HRIRs (head-related impulse responses). HRTF data are usually measured with probe microphones and a group of test subjects, or they can be measured with dummy heads.

Because of the fact that the later part of the impulse response doesn't need to be created with the same accuracy as the early part hybrid methods are often tried with a separation of the simulation into two or more steps. This is justified by the low energy contained in the late part and by masking. In auralization tests with running speech or music the later part of the reverberation tail is masked and noticeable only with impulsive or rhythmic signals. The early part is considered important up to a length of some hundred milliseconds at most. Methods for simulation of exact binaural impulse responses are described above. The fine structure in the late response is stochastic and can be taken, for instance, from a stochastic sequence of Dirac pulses. Octave band filters and summation enable transition into narrow-band or broadband sound pressure signals.

The appropriate playback arrangement is either headphone presentation or free-field reproduction with

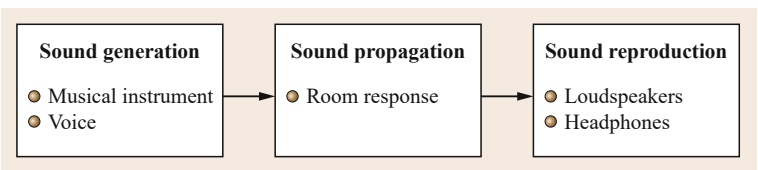


Fig. 11.14 Principle of auralization

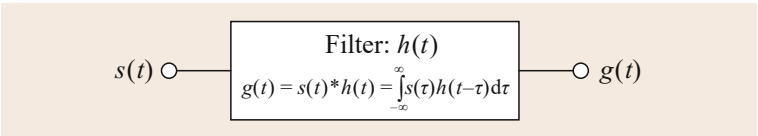


Fig. 11.15 Convolution of signal $s(t)$ with a room impulse response $h(t)$

loudspeakers and crosstalk cancelation. The equalization of the headphone or the loudspeakers is extremely important if the sound localization and spatial impression are to be created correctly. However, binaural technology has been well investigated and can be applied for these purposes.

11.8 Current Research Topics

Room acoustics is still not a solved problem. In spite of the standards and design tools there is still a need for research and transfer of knowledge to other engineering disciplines. Still too many meeting rooms, classrooms, auditoria, offices, train or metro stations, and airports have unacceptable acoustic conditions. Some recent research activities are described in the examples below.

11.8.1 Room Acoustics and Psychoacoustics

Densil Cabrera et al. asked why reverberance must be based on a level decay rather than on a loudness decay [11.44]. Time-dependent loudness also includes spectral masking as the temporal effects of cochlear sound processing. Models from psychoacoustics are available [11.45] and have already been adopted in a standard [11.46].

In fact, their loudness decay rates are well correlated with subjective reverberance. Furthermore, loudness decays can include signal properties if the decay curve is not derived from the impulse response but from signals convolved with impulse responses. Accordingly, the reverberance can be expressed in two aspects: (a) the sound reverb in signal transients and its effect on the signal modulation and (b) the decay after the final chord.

Concerning binaural aspects, hearing models of binaural processing are a powerful tool to evaluate localization, diffuseness, and spaciousness. Based on the Jeffress model of cross-correlation [11.47], diverse model approaches exist (*Lindemann* [11.48], *Dau* [11.49], *Breebaart* [11.50], *van Dorp Schuitman* [11.51], among others) and those will be used increasingly in room acoustic research.

11.8.2 Room Acoustic Measurements

A rather new field of research is the evaluation of measurement techniques (typically application of ISO 3382) concerning measurement uncertainty. The measurement uncertainty can then be compared with the just-noticeable differences (JND) in order to identify the significance of results. With the framework of the Guide to the expression of uncertainty in measurement (GUM) [11.52] it is possible to investigate the sources of errors influencing a measurement result and to specify the quantitative influence of each source of error on the overall uncertainty. In room acoustics, placement and directivity of dodecahedron loudspeakers were found to be crucial, as well as microphone and dummy head specifications [11.53, 54].

The relation between the subjective impression and in situ diffusivity at listener positions is an old problem and had already been described in the 1950s by Thiele. Directivity diagrams (*hedgehog plot*) illustrate the distribution of distinct specular and smooth diffuse sound incidence. But measurement of a detailed directional distribution is tedious and uncertain when it comes to specification of highly directional microphones. An innovative approach was introduced by Lokki and Pätynen at Aalto University in Helsinki, who use a loudspeaker orchestra which can provide reproducible situations on the stage as concerns the source positions and the directional radiation. Furthermore, they analyze the impulse responses with microphone arrays and spatial decomposition in order be able to reproduce virtual orchestra sounds for listening tests and sensory evaluations [11.55].

11.8.3 Virtual Room Acoustics

The basic algorithms of computer models were developed in the 1990s. Also at that time, the importance of scattering was recognized, thus leading to implementation of hybrid models. Their main feature is that deterministic specular components in the impulse responses are calculated separately from stochastic (diffusely scattered) parts. Both parts are combined in the end. It is important to mention that the physical sound field can be separated into those two parts as well. Auralizations using such kinds of hybrid model are in perceptive aspects close to recordings. To quantify the differences and to understand the reason for differences, though, is subject to research.

Besides the development of algorithms describing the physical sound propagation in rooms, many ideas from the field of computer graphics were also applied to room acoustic modeling, such as space partitioning or other methods to speed-up the algorithmic kernel for finding the ray intersection point with polygons. Because of severe differences of (narrow-band) light propagation in contrast to broadband sound propaga-

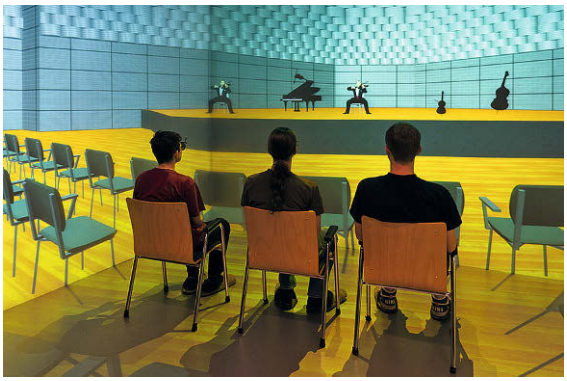


Fig. 11.16 Audiovisual real-time simulation of concert hall acoustics (photo by Peter Winandy)

tion and differences in the ratio of room and object sizes in comparison to wavelengths, the algorithms used for creation of computer images cannot be applied in room acoustics. Nevertheless, data management and dynamic object handling developed in computer graphics makes real-time auralization possible (acoustic virtual reality) (Fig. 11.16).

In a consortium of the Universities of Berlin, Aachen, Ilmenau, Oldenburg, and the Negev, Israel, several aspects of perception in room acoustics, and furthermore recording and reproduction of spaces is being investigated from various viewpoints [11.56].

The research group followed a coordinated effort to improve the complete signal chain from the numerical modelling, the data acquisition within numerical or real sound fields, the coding and transmission to the electro-acoustic reproduction by binaural technology or by sound field synthesis. A novel approach for the comparative evaluation of simulated environments did not only perceptively validate all improvements along the signal chain; it also allowed an evaluation of the plausibility and/or the authenticity of virtual room acoustic environments as a whole. Moreover, it brought forth better physical measures to predict the qualities of natural acoustic environments as well.

Another very interesting auralization project was presented by Woszczyk in 2009. *Virtual Haydn* [11.57] involved seven historic pianos, which were reconstructed by highly renowned artisans. They were played in nine virtual rooms created to represent rooms from Haydn's time, i. e., rooms in which Haydn could have played.

11.8.4 Array Technologies in Room Acoustics

The development of linear, circular, and spherical arrays for 3D sound field analysis inspired many investigations in room acoustics. Arrays are used for spatial decomposition and beamforming, thus for decomposition of complex sound fields into plane or spherical waves. Those waves of course stem from reflections. The advantage is that not only the temporal and spectral cues of room sound fields are captured, but their directional distribution as well. And apart from the possibility of multichannel sound recording, this technique is very interesting for analysis of room sound fields and reflection patterns.

One of the first findings with linear arrays was that clarity, C80, is very sensitive to small changes in the microphone position. If impulse responses are evaluated for adjacent array microphones just a few centimeters apart, clarity data differ more than the just noticeable difference [11.58]. This would indeed mean that the subjective impression of clarity differs significantly within an area of one concert hall seat. This does not match the listening experience at all, and accordingly the definition of the clarity index creates a too sensitive result.

The next example is application of the concept of wave field analysis (WFA), to room acoustics. In this way, rooms can be compared in regard to their wave propagation patterns [11.59], which gives a valuable insight into the spatial wave incidence and, thus, the amount of distinct specular wave fronts and diffuse components.

11.9 Final Remarks

In room acoustics, several dimensions of the overall listening impression were identified which show a good correlation with corresponding objective measurement data. It can be summarized that the three most important factors (loudness, reverberance, and spatial impression) explain most of the statistical variance when comparing the acoustic conditions in auditoria. Aspects of scattering and diffusivity and their relation to surface design are still under investigation, as well as robust descriptors of stage acoustics.

Apart from the definition of the most relevant perceptual factors, another crucial point is the listener's sensitivity to changes in a sound field in regard to those subjective aspects (just-noticeable differences) and to the robustness and significance of measurement results.

Cooperation of room acoustics with psychoacoustics and audio engineering will stimulate more ideas and innovation in research and concert hall design.

Room acoustics today is far from being black magic. Instead, there is a strict scientific interdisciplinary concept which will be further developed. Methods, knowledge, and guidelines from classical room acoustics will also be used for other purposes than acoustics in performance spaces, such as for train stations or airports, factory halls and offices.

References

11.1 ISO 3382: Measurement of the reverberation time of rooms with reference to other acoustical parameters (1995)

11.2 L. Beranek: *Concert Halls and Opera Houses – Music, Acoustics and Architecture*, 2nd edn. (Springer, New York 2004)

11.3 M. Barron: *Auditorium Acoustics and Architectural Design*, 2nd edn. (E&FN SPON, London 2010)

11.4 M.R. Schröder: Die statistischen Parameter der Frequenzkurven von großen Räumen, Acustica **4**, 595 (1954)

11.5 H. Kuttruff, M.R. Schroeder: On frequency response curves in rooms. Comparison of experimental, theoretical, and Monte Carlo results for the average frequency spacing between maxima, J. Acoust. Soc. Am. **34**, 76 (1962)

11.6 W.C. Sabine: *The American Architect 1900. Collected Papers Nr. 1* (Harvard Univ. Press, Cambridge 1922)

11.7 R.F. Norris, C.A. Andree: An instrumental method of reverberation measurement, J. Acoust. Soc. Am. **3**, 366 (1930)

11.8 C.F. Eyring: Methods of calculating the average coefficient of sound absorption, J. Acoust. Soc. Am. **4**, 178 (1933)

11.9 M. Vorländer: Revised relation between the sound power and the average sound pressure level in rooms, Acustica **81**, 332 (1995)

11.10 L. Cremer: *Die wissenschaftlichen Grundlagen der Raumakustik* (Hirzel, Stuttgart 1948)

11.11 J. Blauert: *Spatial Hearing – Revised Edition: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge 1996)

11.12 M.R. Schroeder: New method for measuring reverberation time, J. Acoust. Soc. Am. **37**, 409 (1965)

11.13 M. Guski, M. Vorländer: Comparison of noise compensation methods for room acoustic impulse response evaluations, Acust. Acta Acust. **100**, 320 (2014)

11.14 N. Xiang: Evaluation of reverberation times using a nonlinear regression approach, J. Acoust. Soc. Am. **98**, 2112 (1995)

11.15 S. Müller: Computer-generated pulse signal applied for sound measurements. In: *ASA Handbook of Signal Processing in Acoustics*, ed. by D. Havelock, S. Kuwano, M. Vorländer (Springer, New York 2005)

11.16 A. Farina: Simultaneous measurement of impulse response and distortion with a swept-sine technique. In: *AES 108th Convention, Paris* (2000)

11.17 P. Dietrich, B. Masiero, M. Vorländer: On the optimization of the multiple exponential sweep method, J. Audio Eng. Soc. **61**(3), 113 (2013)

11.18 D. Rife, J. Vanderkooy: Transfer-function measurement with maximum-length sequences, J. Audio Eng. Soc. **37**, 419 (1989)

11.19 M. Vorländer, M. Kob: Practical aspects of MLS measurements in building acoustics, Appl. Acoust. **52**, 239 (1997)

11.20 R. Thiele: Richtungsverteilung und Zeitfolge der Schallrückwürfe in Räumen, Acustica **3**, 291 (1953)

11.21 V.L. Jordan: Acoustical criteria for auditoriums and their relation to model techniques, J. Acoust. Soc. Am. **47**, 408 (1970)

11.22 W. Reichardt, O.A. Alim, W. Schmidt: Abhängigkeit der Grenzen zwischen brauchbarer und unbrauchbarer Durchsichtigkeit von der Art des Musikmotives, der Nachhallzeit und der Nachhalleinsatzzeit, Appl. Acoust. **7**, 243 (1974)

11.23 W. Wilkens: *Mehrdimensionale Beschreibung subjektiver Beurteilungen der Akustik von Konzertsälen*, Ph.D. Thesis (Technical University Berlin, Berlin 1975)

11.24 P. Lehmann: *Über die Ermittlung raumakustischer Kriterien und deren Zusammenhang mit subjektiven Beurteilungen der Hörsamkeit*, Ph.D. Thesis (Technical University Berlin, Berlin 1976)

11.25 D. Gottlob: *Vergleich objektiver Parameter mit Ergebnissen subjektiver Untersuchungen an Konzertsälen*, Ph.D. Thesis (Göttingen University, Göttingen 1973)

11.26 K.F. Siebrasse: *Vergleichende subjektive Untersuchungen zur Akustik von Konzertsälen*, Ph.D. Thesis (Göttingen University, Göttingen 1973)

11.27 J. West: Possible subjective significance of the ratio of height to width of concert halls, J. Acoust. Soc. Am. **40**, 1245 (1966)

11.28 H. Marshall: Levels of reflection masking in concert halls, J. Sound Vib. **5**, 116 (1968)

11.29 M. Barron: *The effects of early reflections on subjective acoustical quality of concert halls*, Ph.D. Thesis (Southampton University, Southampton 1976)

11.30 P. Damaske, Y. Ando: Interaural crosscorrelation for multichannel loudspeaker reproduction, Acustica **27**, 232 (1972)

11.31 H. Kuttruff: *Room Acoustics* (Elsevier, Amsterdam 1973)

11.32 J. Meyer: Trends in concert hall design – A review of the past 50 years. In: *26 Tonmeistertagung – VDT Int. Convention* (2010), in German

11.33 D. van Maerke: Simulation of sound fields in time and frequency domain using a geometrical model. In: *Proc. 12th ICA, Toronto*, Vol. 2 (1986), paper E11-7

11.34 M. Vorländer: Simulation of the transient and steady state sound propagation in rooms using a new combined sound particle – Image source algorithm, J. Acoust. Soc. Am. **86**, 172 (1989)

11.35 B.I. Dalenbäck: Room acoustic prediction based on a unified treatment of diffuse and specular reflection, J. Acoust. Soc. Am. **100**, 899 (1996)

11.36 G.M. Naylor: Odeon–another hybrid room acoustical model, Appl. Acoust. **38**, 131 (1993)

11.37 M. Vorländer: International round robin on room acoustical computer simulations. In: *Proc. 15th ICA, Trondheim* (1995) p. 689

11.38 I. Bork: A Comparison of room simulation software – The 2nd round robin on room acoustical computer simulation, Acust. Acta Acust. **84**, 943 (2000)

11.39 I. Bork: Report on the 3rd round robin on room acoustical computer simulation – Part II: Calculations, Acust. Acta Acust. **91**, 753 (2005)

11.40 ISO 17497-1: Acoustics – Measurement of the sound scattering properties of surfaces – Part 1: Measurement of the random-incidence scattering coefficient in a reverberation room (2004)

11.41 F.P. Mechel: Improved mirror source method in room acoustics, J. Sound Vib. **256**(5), 873 (2002)

11.42 S. Siltanen, T. Lokki, S. Kiminki, L. Savioja: The room acoustic rendering equation, J. Acoust. Soc. Am. **122**(3), 1624 (2007)

11.43 M. Vorländer: *Auralization – Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality* (Springer, Berlin 2008)

11.44 D. Lee, D. Cabrera, W.L. Martens: Equal reverberance contours for synthetic room impulse responses listened to directly: Evaluation of reverberance in terms of loudness decay parameters. In: *Int. Symp. Room Acoust., Melbourne* (2010)

11.45 J.H. Chalupper: Fastl: Dynamic loudness model (DLM) for normal and hearing-impaired listeners, Acust. Acta Acust. **88**, 379 (2002)

11.46 DIN 45631 A1: Calculation of loudness level and loudness from the sound spectrum – Zwicker method – Amendment 1: Calculation of the loudness of time-variant sound (2010)

11.47 L.A. Jeffress: A place theory of sound localization, J. Comp. Physiol. Psychol. **41**, 35 (1948)

11.48 W. Lindemann: Extension of a binaural cross-correlation model by means of contralateral inhibition. II. The law of the first wave front, J. Acoust. Soc. Am. **80**, 1623 (1986)

11.49 T. Dau, D. Püschel, A. Kohlrausch: A quantitative model of the effective signal processing in the auditory system. I. Model structure, J. Acoust. Soc. Am. **99**, 3615 (1996)

11.50 J. Breebaart, S. van de Par, A. Kohlrausch: Binaural processing model based on contralateral inhibition. I. Model setup, J. Acoust. Soc. Am. **110**, 1074 (2001)

11.51 J. van Dorp Schuitman, D. de Vries: *Deriving Room Acoustical Parameters Using Arrays and Hearing Models* (NAG/DAGA, Rotterdam 2009)

11.52 ISO/IEC Guide 98: Guide to the expression of uncertainty in measurement (GUM) (1993)

11.53 R. San Martín, I. Witew, M. Arana, M. Vorländer: Influence of the source orientation on the measurement of acoustic parameters, Acust. Acta Acust. **93**, 387 (2007)

11.54 I. Witew, A. Lindau, J. van Dorp Schuitman, M. Vorländer, S. Weinzierl, D. de Vries: *Application of GUM concepts on uncertainties of IACC caused by dummy head orientation* (DAGA, Berlin 2010)

11.55 T. Lokki, J. Pätynen, A. Kuusinen, S. Tervo: Concert hall acoustics: Repertoire, listening position and individual taste of the listeners influence the qualitative attributes and preferences, J. Acoust. Soc. Am. **140**(1), 551 (2016)

11.56 S. Weinzierl, M. Vorländer: Room acoustical parameters as predictors of room acoustical impression: What do we know and what would we like to know?, Acoust. Aust. **1**, 41 (2015)

11.57 T. Beghin: *The Virtual Haydn* (University of Chicago Press, Chicago 2015)

11.58 D. de Vries, E.M. Hulsebos, J. Baan: Spatial fluctuations in measures for spaciousness, J. Acoust. Soc. Am. **110**, 947 (2001)

11.59 A.J. Berkhout, D. de Vries, J.J. Sonke: Array technology for acoustic wave field analysis in enclosures, J. Acoust. Soc. Am. **102**, 2757 (1997)

Signal Pr Part B

Part B Signal Processing

Ed. by Jonas Braasch

This book section deals with signal processing techniques from the perspective of computational systematic musicology. Recent developments in the field have opened up major opportunities for researchers with technical backgrounds to pursue new ways of analyzing music using digital signal processing techniques. Traditionally, the primary work of systematic musicology focused on the manual analysis of music scores. In classical music, where the field of musicology originates from, the concept of work is based on music scores. While the score defines the work of a composer, the performance of these works is considered interpretation and thus of secondary interest. In contrast, the concept of work for many non-Western music genres as well as pop music and jazz is mainly defined through the actual sound of a performance or recording. Different rock bands, for example, use exactly the same chord progression, but the sound of these progressions is so unique that one often can tell the song before the singer starts. One cannot understand the fundamentals of this culture by just analyzing scores, and signal processing techniques are now used as an objective method to analyze the tempo, tonality, and timbre of this music, among many other features.

The underlying music analysis tools, which were introduced to better understand music, have meanwhile found their way into the music industry. Music streaming services use them to classify songs according to user preferences to maintain a competitive edge. The shift from storing music on individual media (e.g., vinyl, CDs) to large databases makes big data analyses possible, as media can now be accessed digitally. Instead of analyzing a few works by hand, general musical trends can be investigated by automatically analyzing a very large corpus of work using batch processes.

Meanwhile, signal processing tools are seeing an increased use in investigations of classical music as well, because there is an emerging focus on understanding the interpretation of works. This trend has to do with the fact that the historic body of classical music works no longer grows, and many major compositions have been analyzed multiple times. In contrast, the practice of interpreting these works is still a developing culture, leading to new opportunities to analyze works.

Moreover, signal processing methods have become important in organology, the science of musical instruments. It is common practice to develop a mathematical model of a musical instrument in order to understand how it operates and understand why certain trends occur. In this book section, we will cover one-dimensional approaches, where the acoustic propagation is simulated along one axis, for example along the resonator of a wind instrument, as well as three-dimensional methods, where the sound can travel across all three spatial dimensions.

Most chapters in this section deal with acoustical analysis methods, but techniques to process or synthesize sounds are also covered. Historically, systematic musicology has solely worked with passive methods, where the researcher merely observes his or her object of study. Meanwhile active research schemes have become more popular; for example, in Simha Arom's approach to experimental ethnomusicology he performs together with indigenous musicians to examine if his theories hold up to a practical test. Digital signal processing and synthesis methods can be instrumental in active research methods, for example in the virtual recreation of a historic music performance in a concert venue that no longer exists. This can be accomplished using wave field synthesis techniques in order to understand why the composer chose a given instrumentation in this venue.

The book section is organized as follows:

Chapter 12 covers the fundamentals of music studio technology. It deals with the whole studio production chain from capturing the signals using different microphone techniques, storing them on various analog and digital media, to different methods of playing back the signals using standard loudspeaker arrangements. In addition, the chapter outlines the complex signal processing techniques in studios that formed the artistic sound concepts of popular music. Practical introductions to digital signals, cables and connectors, and music synthesizers complete this reference on studio technology.

Chapter 13 deals with delay lines and digital waveguides. Both methods are used to understand sound propagation along a one-dimensional axis. This chapter covers the fundamentals including finite impulse response (FIR) filters, linear interpolation, sound reflection, lossy wave propagation, and comb filter effects. After explaining mathematical fundamentals, the chapter focuses on the acoustic simulation of a plucked string instrument by simulating the acoustic behavior of a damped string that is terminated at both ends after being plucked in its initial state.

Chapter 14 gives a mathematical introduction to the fundamental concepts of convolution, Fourier analysis, and cross-correlation. All three concepts are important signal processing methods used to analyze and synthesize sound. Convolution is used to simulate the transformation of a signal by a linear time-invariant acoustical system, Fourier analysis is used to understand how a signal behaves in terms of frequency, and

cross-correlation is used to compare the similarity of two systems.

Chapter 15 describes signal processing methods used to separate individual sound sources from a mixture. The techniques described in this chapter are fundamental to analysis of the complex sounds of ensemble music, where more than one musical instrument or voice is involved. Within the chapter, the repeating pattern extraction technique (REPET) is introduced, which detects periodic functions, e.g., in the form of beat patterns, establishes a model of the repeated patterns, and then segregates these pattern from the acoustic background using time/frequency masks. Another focus of this chapter is multipitch streaming to extract pitch contours of overlaid instruments. The contours are used to construct harmonic masks to segregate the individual instruments. The chapter concludes with the description of techniques to align the sound of a recorded performance to the underlying musical score.

Chapter 16 covers optical recognition methods to automatically extract music scores. Optical music recognition (OMR) is an instrumental process for archiving historic scores based on a symbolic notation. OMR provides a basis for automatic music analysis using the converted scores, where the extracted scores can be easily sonified using a music synthesizer. OMR has its roots in optical text recognition, but is affected by additional challenges that mainly result from the circumstance that music symbols, e.g., notes and staff lines, overlap, while text characters are generally isolated. The chapter provides an overview of different techniques to overcome the resulting technical challenges.

Chapter 17 deals with adaptive digital music systems. While it is often easier and therefore advantageous to describe a system mathematically using linear time-invariant systems, there are clear limitations to this approach. In general, musical instruments are nonlinear, time-variant systems, for example the bore of a wind instrument using keys to produces different pitches. In some cases, time-variant systems can be approximated using a sequence of time-invariant systems, but in many other cases only adaptive systems can simulate the behavior of time-variant systems with sufficient complexity. This chapter focuses on a number of techniques including Kalman filters and autoregressive moving average (ARMA) processes to provide expressive control of digital music systems in the joint time/frequency space.

Chapter 18 covers the fundamentals of wave field synthesis (WFS). WFS is an advanced method that simulates how sound propagates in three-dimensional space. It builds on the Huygens–Fresnel principle, which states that any given wave front can be described through the superposition of elementary spherical sound sources, and the Kirchhoff–Helmholtz integral. The latter demonstrates that by controlling the pressure and velocity along the boundaries of an enclosed space any three-dimensional sound field can be synthesized within it. This assumption works as long as the volume inside the enclosure is source-free. The chapter introduces the theoretical framework for WFS and discusses the practical approach and limitations of rendering sound fields using a two-dimensional array of loudspeakers as elementary sound-pressure sources.

Chapter 19 introduces the finite-difference method (FDM) in the context of musical acoustics. The FDM is a popular tool to describe a set of differential equations by discretizing a space and approximating these equations with difference functions. This method can be used for example by discretizing a geometrical model of a room using a constantly spaced grid (e.g., 10 cm along all three Cartesian coordinates). The difference equations are then solved numerically at the grid positions. The method is a useful tool where the environments are too complex to solve analytically and the dimensions for the desired frequencies are too small to use geometrical methods like ray tracing. This is the case for low frequencies in a small room or for the geometry of many musical instruments that are discussed in this chapter.

Chapter 20 discusses the fundamental problems, solutions, and applications of real-time sound processing. Classical applications where real-time processing is needed are digital music keyboards, live sound effect processors, and virtual reality systems. The chapter covers the main challenges for the design of real-time algorithms to minimize latency and cost of the computational operations. Common real-time algorithms will be explained based on software for field programmable gate array (FPGA) hardware of mobile devices. Mobile devices are becoming increasingly important for musicological research in the form of mobile field recorders, digital tuners and loudness meters, among other tools.

12. Music Studio Technology

Robert Mores

Music studio technology is reviewed with respect to the different tasks involved in recordings, broadcasts, and live concerts. The chapter covers microphones and microphone arrangements, signal preconditioning and sound effects, and matters of digitalization. It also covers equipment technology such as mixing consoles, synthesizers and sequencers. Historical and contemporary audio formats are reviewed including the issues of restoration. Practical matters such as signals, connectors, cables and grounding problems are addressed due to their significance to sound quality. The general trend towards audio networks is shown. Finally, speakers, reference listening and reinforcement systems are outlined, including some of the multidimensional formats.

This book chapter on music studio technology gives an overview of the state of the art in recording and reproduction. Music studios are usually specifically equipped to serve a dedicated task, like the recording of different genres, speech production, film sound production, synchronization, broadcast, sound reinforcement in live concerts, sound design, or art work. Therefore, Fig. 12.1 is only a rough overview of the most common elements of the workflow and equipment typically used for live concerts, broadcasts, or recordings.

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_12

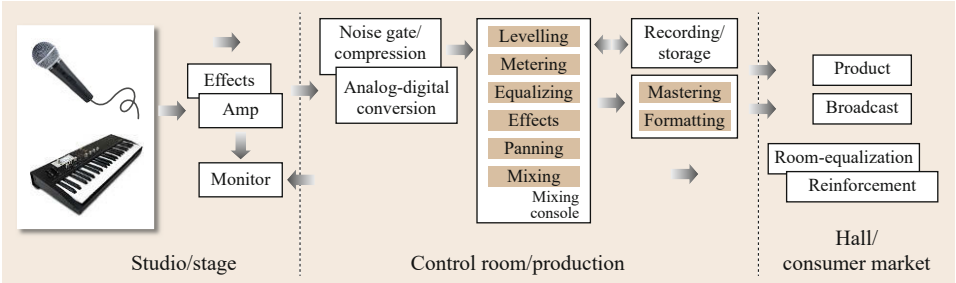


Fig. 12.1 Commonly used equipment and workflow in live concerts, broadcast and studio recording

In the studio or on the stage, there is usually only a limited demand for local amplification or individual effects where the audio signals are taken. Most of the signals are still analog and are converted to digital format in the control room after some noise-gating or dynamic compression. Analog mixing consoles are still popular in live performances and work without an analog–digital converter. First, the mixing console, whether analog or digital, preconditions any incoming signal, such as levelling, dynamic control, or equalization. Second, it facilitates enhancements, such as effects, or reverberation (Sect. 12.2). Third, it allows for the grouping of several sources and for panning to various channels for sound reinforcement or monitoring. For recording tasks, audio signals are stored individually or in groups before being postprocessed in terms of enhancement, mixing, or panning. Mastering is the final step in productions to customize formats or to compensate for limitations of preferred listening situations. Mastering might be done by means of the same mixing console, apart from extensive coding tasks. Today's desktop computing power allows for integration of all mixing console tasks together with the recording, the digital version of the noise gate, dynamic compression, and mastering and coding – all on one computer. This is common practice in small productions and a competitive challenge for professional producers.

This chapter describes the equipment and methods used in the presented workflow and is structured as follows:

1. Microphones and microphone arrangements
2. Signal preconditioning and effects
3. Digitalization
4. Mixing consoles
5. Synthesizer and sequencer
6. Historical and contemporary audio formats and restoration
7. Signals, connectors, cables and audio networks
8. Speakers, reference listening and reinforcement.

Some emphasis is given to microphones, signaling and formats because these issues are likely to be relevant in the projects and studies of musicology students, such as field work recording or retrieval of historical archives.

12.1 Microphones and Microphone Arrangements

There are two defining properties of microphones. Depending on its construction, a microphone can be sensitive to pressure or to pressure difference. Within pressure-sensitive microphones, the membrane, or diaphragm, faces the soundfield on one side and a reference pressure of an enclosed volume on the other side, effectively measuring absolute pressure independent of direction. In pressure-gradient microphones the diaphragm is exposed to the soundfield on both sides, which explains the sensitivity towards gradients normal to the plane of the diaphragm and the insensitivity towards in-plane gradients, see Fig. 12.2. The second property relates to the converter principle. Dynamic microphones follow the inverse principle of loudspeakers. The membrane drives a coil in a magnetic field and the induced signal is strong enough to be transmitted without in-microphone amplification. In condenser microphones the metalized diaphragm, together with a plate form the two sides of a capacitor and excitation of the diaphragm, will change the gap and therefore the capacitor's electrical parameters. The capacitor is usually DC-charged and the weak signal must be amplified within the microphone while the power for this task is facilitated by battery or a remote supply, see Sect. 12.7.2. In piezomicrophones the crystalline structure contains microscopic dipoles where pressure-induced deformation will bring about a weak signal that must be amplified. And, reaching even further in the past, carbon microphones employ the pressure-dependent resistance of anthracite to mod-

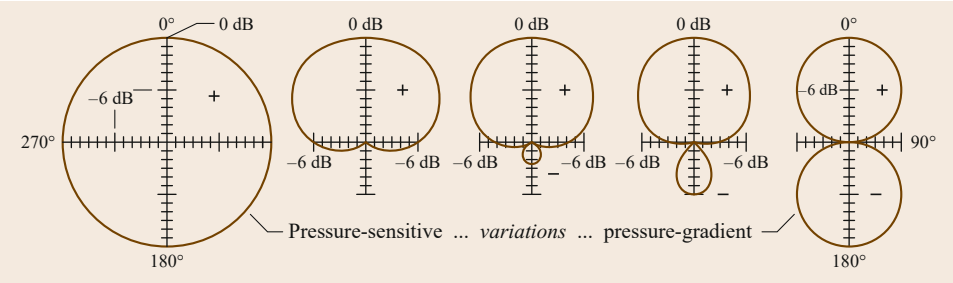
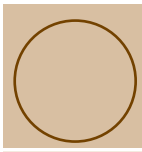


Fig. 12.2 Directivity polar diagrams for pressure-sensitive and pressure-gradient microphones and for constructive variations in between

Table 12.1 Main parameters of microphones of specific directivity

					
Directivity	Omnidirectional	Cardioid	Supercardioid	Hypercardioid	Figure-of-eight
Sensitivity $s(\theta)$	1.0	$0.5 + 0.5\cos(\theta)$	$0.366 + 0.634\cos(\theta)$	$0.25 + 0.75\cos(\theta)$	$\cos(\theta)$
Directivity γ	1.0	3.0	3.732	4.0	3.0
Beam angle −3 dB	–	65.5°	57.5°	52.4°	45°
Beam angle −6 dB	–	90°	77.8°	70.5°	60°
Sensitivity 90°	0 dB	−6 dB	−9 dB	−12 dB	$-\infty$ dB
Sensitivity 180°	0 dB	$-\infty$ dB	−12 dB	−6 dB	0 dB

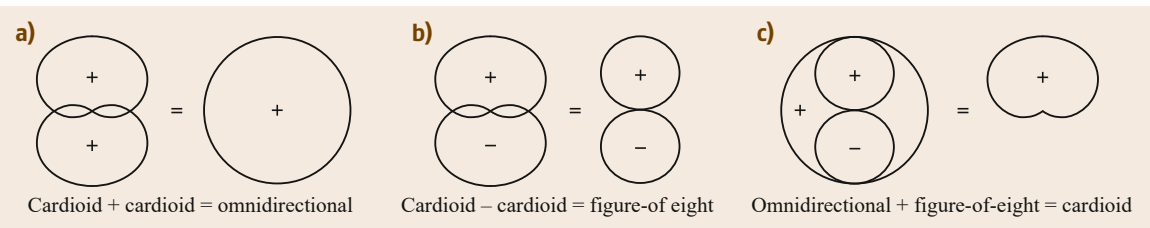


Fig. 12.3a–c Polar diagrams for microphones of varying directivity combined

ulate a DC source at low quality, e.g., high distortion ratio, low signal-to-noise ratio or a strongly limited frequency band. Rare examples of instrumentation use modulation of a laser beam as a principle. Some other instrumentation tasks feature a heated wire for measuring particle velocity at low frequencies. The physical principle behind this instrumentation is that the heated wire will be cooled down by particle movement.

Directivity factors can be further varied by construction (Fig. 12.2; Table 12.1) or by mutual addition or subtraction of signals from several microphones (Fig. 12.3).

Illustrations and directivity parameters presume perfect geometry. However, the outlined directivity patterns come close to reality only for low frequencies. Higher frequencies suffer from shadow, interference and bending effects (Fig. 12.4 for the MK 5 condenser microphone from Schoeps). In principle, the frequencies affected are inversely proportional to the size of the membrane and microphone. The smaller the microphone, the higher the fidelity at high frequencies.

The directivity type conversions illustrated in Fig. 12.3 presume perfect coincidence. However, two microphones cannot be in the same place at the same time. The distance between them will introduce comb filter artefacts. Figure 12.4 also indicates linearity over a wide range of frequencies. Manufacturers usually measure the frequency response on axis, $\theta = 0$. Other directions reveal frequency dependent attenuation. Consequently, the change of color/timbre is well perceivable for microphones when held off-axis. While striving for linearity, there are two major compensation issues. Pressure-gradient microphones respond with a leveraged bass range when held within the near-field of a source, the so-called proximity effect. In soloist microphones, for instance, this is usually com-

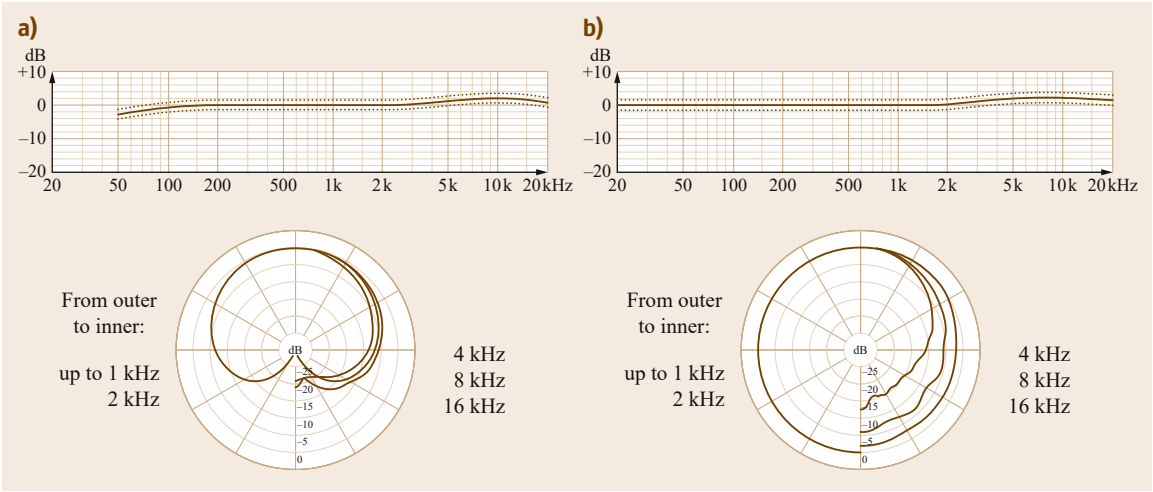


Fig. 12.4a,b Frequency-dependent directivity of a MK 5 microphone in its pressure-sensitive **(a)** and its pressure-gradient **(b)** mode

pensated for. The other major effect is that pressure-sensitive microphones respond with a leveraged treble range in the near-field. This is again compensated for in some microphones.

12.1.1 Coincident versus Spaced Microphone Arrangements

Multichannel recording facilitates spatial hearing while employing three basic principles of binaural listening:

(i) Direction-dependent intensity differences at the ears, or interaural intensity difference (IID)
(ii) Direction-dependent propagation delay, which is particularly dominant while localizing the first wavefront (precedence effect), or interaural time difference (ITD)
(iii) Directional color as determined by the shape of the head and the pinna.

A coincident setup follows principle (i) as microphones are arranged very close to each other to minimize propagation delay while the intensity difference will be achieved only by directionality of the microphones. Space between microphones introduces propagation delay to employ principle (ii) and it is the art of the recording engineer to translate source width and room dimensions into appropriate spacing. A dummy head employs all three principles.

12.1.2 Two-Dimensional Microphone Arrangements

The dimensions of recording and reproduction correspond to each other. Stereo or Dolby surround are two-dimensional and both microphone and loudspeaker arrangements are aligned with the surface plane. The left (L) and right (R) channels of a stereo mix are obtained from mid (M) and side (S) signals or from X and Y signals in coincident setups or from A and B signals in spaced setups (Fig. 12.5). The preferred angle α in XY arrangements is derived from the -3 dB crossover point of the cardioid directivity pattern, but other directivity patterns and other angles are used as well, e.g., $\alpha = 131°$ for cardioid, 115° for supercardioid, and 105° for hypercardioid. Spacing d in AB arrangements

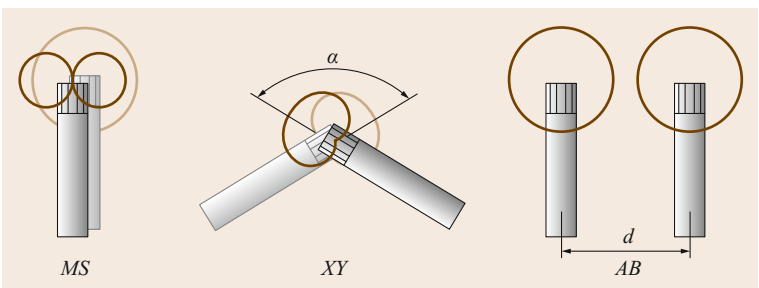


Fig. 12.5 Coincident and spaced microphone setup for stereo

varies from 17 cm to a few meters.

$$L = \frac{M+S}{\sqrt{2}} \quad R = \frac{M-S}{\sqrt{2}} \quad M = \frac{L+R}{\sqrt{2}} \qquad (12.1)$$

L and R result from the addition and subtraction of M and S (12.1). Reverse mixing re-establishes the omnidirectional M signal and is therefore a perfect mono downmix. XY signals directly correspond to L and R, and a mono downmix results in a wide subcardioid. A downmix of AB stereo results in comb filter artefacts and therefore in more or less annoying changes of timbre.

All three arrangements facilitate adaptation to the source width. If the contribution of S in (12.1) is chosen to be larger in relation to M, the captured sound source will appear wider. Likewise, such zooming effects will be obtained by larger α in XY arrangements and by a larger space d in AB arrangements.

A crossed figure-of-eight allows panning. Addition and subtraction will directly result in virtual figures-of-eight that are shifted by 45° (see the Blumlein arrangement in Fig. 12.6). Directions in between can be obtained by respective weighting factors. An additional coincident omnidirectional microphone, when combined with the Blumlein arrangement, will facilitate beam forming and panning. Both can be done even in the aftermath of recordings.

Arrangements for stereo recordings that combine spacing with the directivity of pressure-gradient microphones are shown in Fig. 12.7. The center C in the Decca tree is partially mixed to the left and right channels for stereo, and is routed directly to the center in surround recordings. Office de Radiodiffusion-Télévision France (ORTF) and Nederlandsche Omroep Stichting (NOS) arrangements are national standards but widely used in broadcast.

Arrangements for surround recordings are shown in Fig. 12.8. The most popular approaches are optimized cardioiid triangle (OCT), Ideale Nierenanordnung (INA, english: ideal cardioid arrangement), and the *Williams* [12.1] arrangement. Spacings a, b, c and d range from 8 to 116 cm and are specified for the individual arrangements and for individual angles α and β while following [12.2–4]. The three-microphone arrangement is supplemented by additional microphones for the surround channels. Established arrangements are the Hamasaki square, 4 cardioids spaced 1 to 3 m and far from the front microphones, or the IRT-cross, 4 cardioids spaced up to 1 m.

12.1.3 Three-Dimensional Microphone Arrangements

The principle of coincident arrangements can be extended to the third dimension. Using a double MS arrangement in the horizontal plane and a third figure-of-eight for the vertical direction establishes the three orthogonal directions. Virtually any direction can be achieved, again by multiplexing and by using weighting functions. Such arrangements can also be considered as sensing zeroth-order pressure with the omnidirectional

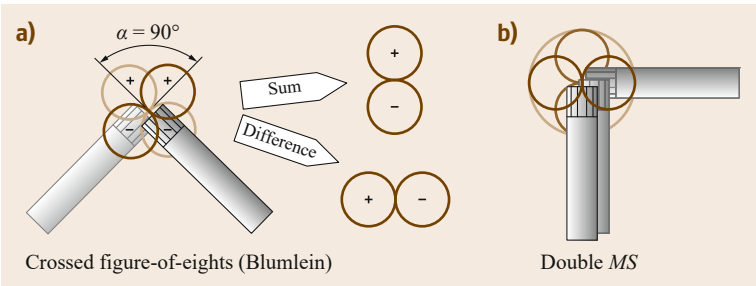


Fig. 12.6 (a) Blumlein and (b) double MS microphone setup

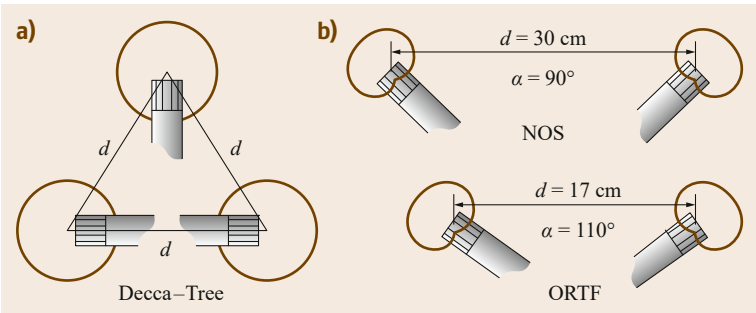


Fig. 12.7 (a) Decca tree, (b) NOS, and (c) ORTF microphone setup for stereo

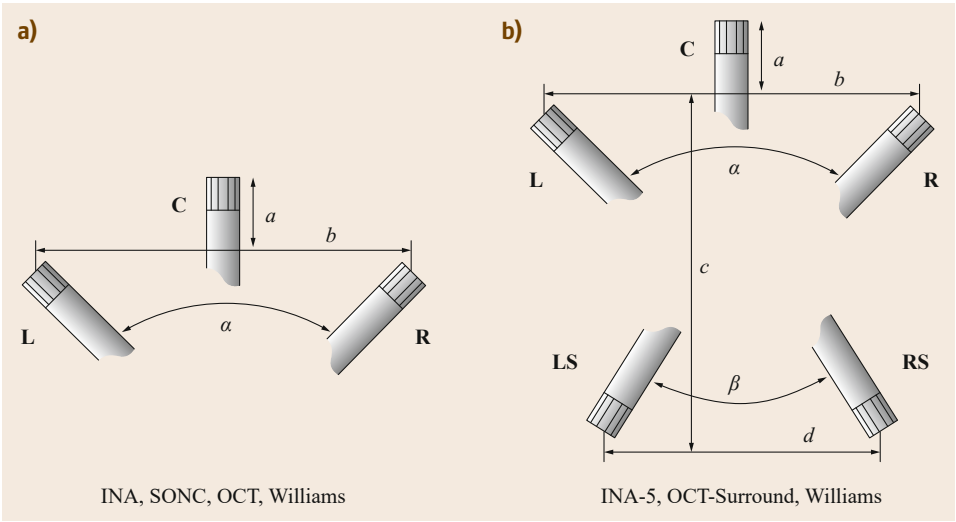


Fig. 12.8a,b INA, OCT and Williams setup for **(a)** stereo and **(b)** surround

microphone and first-order gradient pressure in three directions with the figure-of-eight microphones. Further orders can be developed and a higher spatial resolution obtained by using the Fourier–Bessel series as weighting factors in superimposed spherical harmonics, while also requiring more microphones. Three-dimensional systems of first order require four channels, and systems of second and third order require nine and 16 channels, respectively. Coincidence of the multiple microphones is a challenge. A smart solution was proposed by *Gerzon* in 1975 [12.5], that uses four pressure-gradient microphones arranged in a tetrahedron such that the membranes are in plane with the imaginary surfaces of the tetrahedron (Fig. 12.9). This soundfield microphone outputs the A-format set of signals *LF* (left front), *RF* (right front), *LB* (left back) and *RB* (right back). The B-format is obtained by the matrix (12.2) and can be rotated by angle φ and tilted by angle Φ, (12.3).

$$
\begin{aligned}
W &= \frac{LF + LB + RF + RB}{2} \\
X &= \frac{LF - LB + RF - RB}{2} \\
Y &= \frac{LF + LB - RF - RB}{2} \\
Z &= \frac{LF - LB - RF + RB}{2}
\end{aligned}
\qquad (12.2)
$$

$$
\begin{aligned}
W' &= W\,, \quad Y' = Y\cos\phi + X\sin\phi \\
X' &= X\cos\phi\cos\varphi + Y\sin\phi\cos\varphi + Z\sin\varphi \\
Z' &= Z\cos\varphi - X\cos\phi\sin\varphi - Y\sin\phi\sin\varphi
\end{aligned}
\qquad (12.3)
$$

The B-format can be projected to stereo by (12.4), where the imaginary part of the signal is obtained by the Hilbert transformation and a^* denotes the conjugate complex of a. Microphones with more channels support derivation of higher-order ambisonics (HOA). For instance, the 32-channel Eigenmike (Fig. 12.9d) supports third-order supercardioid.

$$
\begin{aligned}
L &= aX + b^*W + c^*Y \\
R &= a^*X + bW - cY \\
a &= 0.0928 + j0.255\,, \quad b = 0.4699 + j0.171 \\
c &= 0.3225 + j0.00855
\end{aligned}
\qquad (12.4)
$$

Spherical microphone arrays target sound sources and their polar diagrams (Fig. 12.10). Captured data facilitates modelling of the source and playback via a dodecahedron speaker, or other equally-spaced loudspeaker arrangements on spherical grids. Such playback in a real acoustical environment emulates the original sound source in that environment. Spatial resampling on the basis of Fourier transforms is facilitated by the equally spaced grid in the spherical arrangements so that the number of playback speakers may deviate from the number of microphones used in the recording. The captured source can also be listened to within virtually created acoustical rooms. For a tutorial on virtual acoustics, see [12.6].

Plane microphone arrays in rectangular, circular or helical arrangements are often used as acoustical cameras to identify sources of noise. Such cameras can also be used to identify the origin of sound radiation from a shell or a body or from an instrument or a singer. From the sound captured at the microphone array, the surface excitation can be traced back so that the radiation pattern can be fully described, at least for the spatial window covered by the array [12.7].

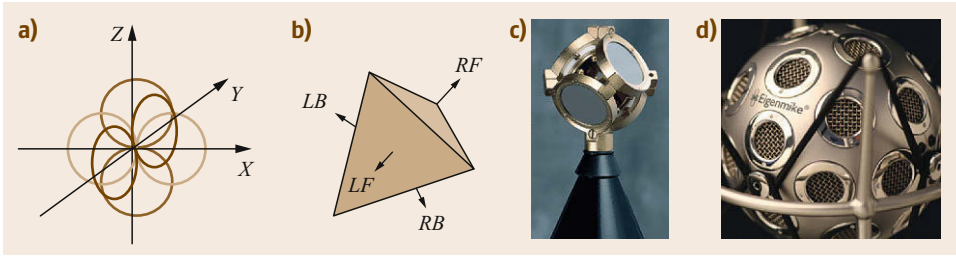


Fig. 12.9a–d Microphone arrangement **(a)** with three figures of eight, **(b)** with four gradient pressure microphones in tetra-hedron geometry, **(c)** the corresponding soundfield microphone manufactured by TSL Products and **(d)** the 32-channel Eigenmike (permission granted by mhacoustics)

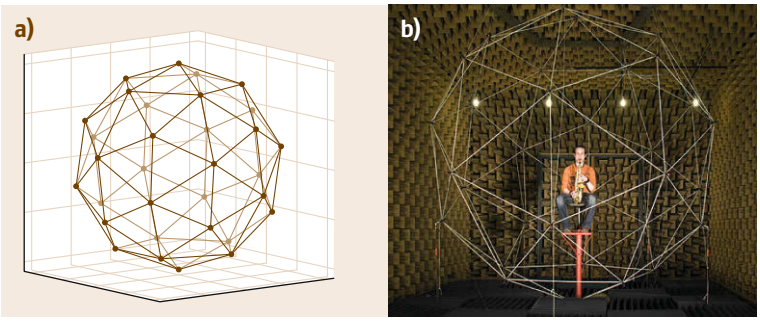


Fig. 12.10a,b Microphones on an equally-spaced spherical grid: **(a)** geometry and **(b)** application (© ITA RWTH Aachen University)

12.2 Signal Preconditioning and Effects

Signal preconditioning, such as noise gating, compression, or equalization, but also signal enhancement and effects have been implemented with the help of analog components and mostly in dedicated standalone equipment for a long time, while modern digital signal processing emulates such signal conditioning and modification.

12.2.1 Noise Gate, Compressor, and Expander

A noise gate mutes the signal in times of silence so that underlying noise cannot be heard anymore. Noise gating also reduces total noise, because individual contributions in mixes of several channels are likely to be noise-free. The controllable threshold level should be set just above the individual noise level so as to still allow pianissimo passages to be gated. Compression is necessary where the signal's dynamic range is larger than what can be captured or can be adequately processed, or larger than can be appropriately reproduced. For instance, brass instruments or percussions reveal a large dynamic range of 40 dB and up in practice. Microphones are capable of processing such a dynamic range, too. Compression effectively applies a high gain factor for low input levels, and – beginning from an adjustable limiter level – a lower gain factor for high input levels. There is usually a soft transitional crossover between the two linear ranges. Controllable attack and release times define the temporal dynamics and require masterful adjustment, because unfavorable settings cause audible distortion, artefacts, or roughness. Expansion is the supplementary, or reverse, processing step to increase the dynamic range. Here, the gain increases with the input level, and thresholds and temporal dynamics can likewise be controlled.

12.2.2 Levelling

Input levelling is an important preconditioning measure. Source levels and impedances depend on the source type (microphone converter principle, see above, or the type of pickup), and the range of levels additionally depends on the musical performance and the setup. Input levels that are too high will cause clipping and therefore distortion, while input levels that are too low are at risk of suffering from a signal-to-noise ratio that is too low during pianissimo passages. Input levelling therefore adjusts the anticipated input range to the range of the subsequent signal processing. In practice, the setting will be done after observing musicians during their

Table 12.2 Digital word formats and their dynamic range

Format	Minimal amplitude	Maximal amplitude	Max. dynamic range (dB)
16-bit fixed point	$-1 \ldots +1$	$-32\,768 \ldots +32\,767$	90.3
20-bit fixed point	$-1 \ldots +1$	$-524\,288 \ldots +524\,287$	114.4
24-bit fixed point	$-1 \ldots +1$	$-8\,388\,608 \ldots +8\,388\,607$	138.5
32/40-bit floating point	$\pm 1.175 \times 10^{-38}$	$\pm 3.403 \times 10^{38}$	1529
64-bit floating point	$\pm 2.225 \times 10^{-308}$	$\pm 1.798 \times 10^{-308}$	12 318

Table 12.3 Equalizer bands according to DIN 45651 and DIN 45652

Octave band center frequency f_O (Hz)	Third band center frequency f_O (Hz)			Bandwidths B (Hz)
31.5	25	31.5	40	22.4 – 28 – 35.5 – 45
63	50	63	80	45 – 56 – 71 – 90
125	100	125	160	90 – 112 – 140 – 180
250	200	250	315	180 – 224 – 280 – 355
500	400	500	630	355 – 450 – 560 – 710
1000	800	1000	1250	710 – 900 – 1120 – 1400
2000	1600	2000	2500	1400 – 1800 – 2240 – 2800
4000	3150	4000	5000	2800 – 3350 – 4500 – 5600
8000	6300	8000	10 000	5600 – 7100 – 9000 – 11 200
16 000	12 500	16 000	20 000	11 200 – 14 000 – 18 000 – 22 400

fortissimo performance. The observed upper limit will be kept clear by some headroom and will be set comfortably below the technical upper limit of subsequent circuits. The peak program meters (PPM) used for observation usually follow a logarithmic scale and provide a visual overload warning. All of this is true for analog mixing consoles and analog recordings. It is good practice to condition input signals before analog to digital conversion as well. Even though the dynamic range within digital systems might be quite large – depending on the word length and on the arithmetic format (Table 12.2) – the analog-to-digital converter has its own dynamic range due to the inherent noise and quantization error.

12.2.3 Equalization

Equalization is another important preconditioning measure. Individual instruments or voices might suffer from unbalanced timbre. Sometimes microphone positions suffer from local excess due to standing waves or there is some impact or other noise from the surroundings to be diminished. Equalization is also relevant for output signaling, for instance, to ultimately balance mastered sound or to compensate for room acoustical deficiencies when working with public address (PA) systems.

Equalization has its origin in the analog domain where usually first- to fourth-order filters have been used to achieve 6–24 dB attenuation per octave, respectively. Active filters also allow amplification, and the so-called peak filters are defined by their center frequency f_0 and gain. Full-parametric filters also facilitate the control of the Q-factor, $Q = f_0/B$ with bandwidth B. Shelving filters control the low- and high-frequency end of the audio band, rather than specific spots within the audio band. Digital implementations usually follow the analog model. Graphical equalizers are banks of filters of fixed frequency and bandwidth to control gain and attenuation for specific frequency bands in the range of ± 12 dB to ± 15 dB. Filters are arranged in such a way that neutral settings result in a unity frequency response. Typical implementations of graphical equalizers follow the definition of frequency octave bands or of musical thirds, according to DIN 45651 and DIN 45652 (Table 12.3). In practical applications therefore, the 10-band equalizers cover the range 22.4–22 400 Hz, in octaves, and 27-band equalizers cover the range to 35.5–18 000 Hz, in thirds.

12.2.4 Metering and Instrumentation

While the PPM is obligatory in mixing consoles, there are further instruments to assess signal quality. The vector scope superimposes the left and right channel in a vector space, i. e., the amplitude of one channel is plotted versus the amplitude of the other channel. This indicates the correlation or decorrelation of the two channels. For the sake of spaciousness, individual channels in stereo or Dolby surround formats should be strongly decorrelated. Metering allows the control of microphone setup and cabling, since reverse polarity and other failures in a switch matrix are spontaneously

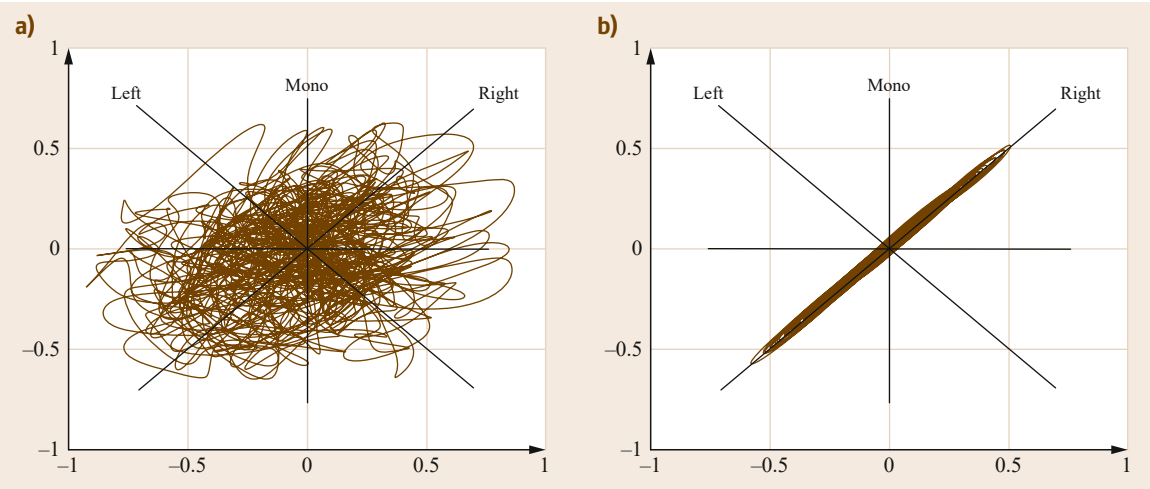


Fig. 12.11a,b Vectorscope plot of a decorrelated stereo signal **(a)** and a signal that is fully panned to the right channel **(b)**

identified. Mixing results can be visually inspected for a desired degree of decorrelation.

In Fig. 12.11, a spacious stereo signal is represented on the left side. Signals panned to either side are represented on either diagonal axis (Fig. 12.11b), mono signals are represented on the vertical axis, and mono signals with a reversed polarity on either channel are represented on the horizontal axis. In a corresponding manner, surround meters prompt the correlation in combination with the signal level for all the five channels in relation to each other.

12.2.5 Distortion, Harmonizer, and Enhancer

Distortion is occasionally a desired enhancement of sound. Strictly speaking, distortion is a widely used term and deserves a structured approach. Linear distortion is achieved by linear operations or processing and is reversible, e.g., by equalization. Nonlinear distortion is the result of nonlinear operations or processing and is irreversible. Nonlinear distortion can be further differentiated into harmonic and nonharmonic distortion. Harmonic distortion may result from operation along a nonlinear amplifier characteristic or from clipping (Fig. 12.12). Nonharmonic distortion adds spectral components that do not follow the series of overtones that are usually found in natural sounds.

The effect of clipping can be calculated analytically using Fourier series. Note that multitudes of the fundamental frequency are developed, which are perceived as harmonics belonging to the fundamental tone. Here, the three-fold, five-fold and further odd-numbered multitudes of the fundamental at 500 Hz are created. The saturation of tubes has a similar effect. However, there is a transitional crossover that effectively weakens the development of higher orders. Saturation at the preamp tube causes what musicians call distortion, while saturation at the output stage is called overdrive. Tubes are likely to create asymmetric clipping, which promotes even-order harmonics. These even-order harmonics match the first, second and third octave of the fundamental (2-fold, 4-fold, and 8-fold), and the fifth in the next but one octave above the fundamental (6-fold). Therefore, the added components fit the musical scale well, which is perceived as a warm and rich enhancement of tone. Harmonizers develop even-order harmonics rather than odd-order harmonics. So-called enhancers and harmonizers use high-pass or band-pass filters to selectively generate higher-order harmonics. The tune parameter facilitates selection while the drive controls the power dosage.

Harmonic distortion can be measured by the distortion factor k,

$$k = 100\% \sqrt{\frac{U_2^2 + U_3^2 + \ldots U_n^2}{U_1^2 + U_2^2 + U_3^2 + \ldots U_n^2}}$$
$$= 100\% \sqrt{\frac{U^2 - U_1^2}{U^2}}\,, \qquad (12.5)$$

where U is the root mean square (RMS), of the signal voltage and U_i is the RMS voltage of the i-th harmonic. The equation suggests that there is nothing else than harmonics in the sound. However, the general form $U^2 - U_1^2$ in (12.5) implies nonharmonic noise, too, in compliance with DIN IEC 60268-2 and AES17. The total harmonic distortion (THD) is the same measure and often denoted in dB. An alternative approach relates the

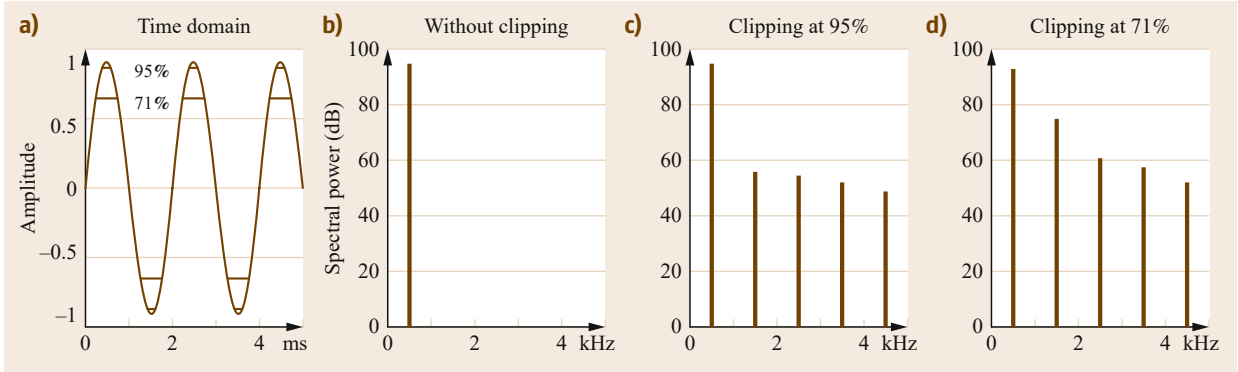


Fig. 12.12a–d 500 Hz sinus **(a)**, and its spectrum without clipping **(b)** and with clipping **(c,d)**

harmonics to the fundamental rather than to the total signal, according to IEEE 1241.

$$\mathrm{THD} = 20\log_{10}\sqrt{\frac{U_2^2+U_3^2+\ldots U_n^2}{U_1^2}}$$
$$= 20\log_{10}\sqrt{\frac{U^2-U_1^2}{U_1^2}} \qquad (12.6)$$

These measures usually qualify the linearity of amplifiers or transformers, and k, or THD, should then be as low as possible. While measuring harmonic distortion, however, k would represent the wanted distortion effect. Finally, nonharmonic distortion can be achieved, for instance, by operating circuits outside their specification, or by adding a chaotic rule set in digital feedback loops, for instance, by operation at supply voltages that are below specification. Nonharmonic components do not comply with the definition of k or THD, but should in general be quantifiable by respective measurement equipment.

12.2.6 Delay Effects, Flanger, and Phaser

Delay and superposition of signals cause manifold perceptual changes of sound in terms of both timbre and localization. The addition of a signal and its delayed representation cause a comb filter effect, or amplification and damping with spectral regularity. Delay by Δt will cause elevation of frequencies $f_{\mathrm{peak}} = n/\Delta t$ and depression at $f_{\mathrm{dip}} = (2n-1)/2\Delta t$ (Fig. 12.13). Such a delay is naturally existent with the well-known echo in any acoustical room and human listening is conditioned to this.

Comb-filter effects are easily audible. Experts can identify the existence of an echo even at levels as low as −18 dB. Figure 12.13b shows a delayed signal at −6 dB.

The parameters to delay effects are: signal delay, level differences, number of delayed signals, time-variance of the delay (modulation), spatial directions of a signal and its *echo*, and the use of feedback. Feedback returns part of the outcome of the delayed signal back to

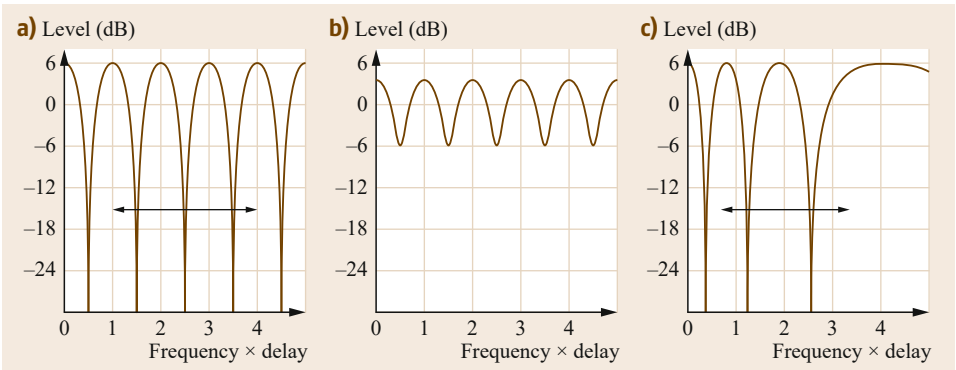


Fig. 12.13a–c Spectral presentation of superimposed signals, **(a)** superposition of a broadband source signal and its delayed version having the same level, the two-sided arrow indicates the variation resulting from a modified delay, **(b)** superposition when the delayed signal is 6dB lower than the source signal, **(c)** for the phaser effect a low delay is periodically varied over time

Table 12.4 Sound effects based on delay and their parameters

Effect	Delay	Modulation	Feedback
Echo	> 100 ms	No	Yes
Double tracking	20 . . . 40 ms	No	No
Multitrap delay	Several individually adjustable delays	Yes/no	Yes/no
Chorus	15 . . . 30 ms	Yes	No
Flanger	1 . . . 10 ms	Yes	Yes

the input. These parameters rule the perception of timbre and space. Echoes can even cause the perception of pitch, especially if signals are broadband signals with little or no tonality. The variation of these parameters results in widely used effects. Echoes use fixed delays above 100 ms with feedback (Table 12.4). Double tracking is used on voice signals and works with delays at the threshold to echo. Multitrap delay uses feedback and modulation to generate nonperiodic patterns and rhythms. For the chorus and the flanger effect, a low delay is periodically varied over time (modulation), which effectively expands and compresses the grid of multiple zeros in the spectral domain. The flanger effect typically works with feedback and with a lower delay than the chorus effect. The flanger effect creates a complex timbre with some tonal character and often sounds like a passing jet, while the chorus effect creates a darker timbre, spaciousness, and fullness of sound.

The phaser effect also has a grid of zeros that is expanded and compressed across the frequency axis, however, the zeros are not equally spaced (Fig. 12.13). The delayed signal passes few or multiple stages of allpass filters, each representing an individually controllable shift of phase.

12.2.7 Reverberation

In natural rooms, early reflections are usually distinct and often separately distinguishable. Reflected sound is fed back into the room and repeated reflections will finally dissolve into a diffuse sound, or reverberation. Reverberation is a desired sound enhancement and human listening is used to it. Recordings might already contain a balanced portion of reverberation or reverberation might be added during sound postproduction after recording in a less reverberant – or dry – studio. In the predigital era, reverberant rooms or tunnels, together with a speaker and several microphones, were used to create reverberation. In the 1950s, a 2 m^2 reverb plate made of steel was used in many studios together with input and output transducers to create mono or stereo reverberation. Today, reverb springs are still commonly used in vintage guitar amps.

An example of an early digital reverb is Schroeder's circuit with four parallel feedback loops with reciprocally allocated peaks and notches, followed by two allpass filters [12.8]. This prototype and its descendent variations, as well as the widely used Lexicon models 224L, 480L, and 960L all have in common that individual parameters of the reverberation are controllable. Examples are room type (church, hall, chamber), reverberation time, structure of early reflections, low frequency reverb (bass multiply), critical frequency for high-frequency attenuation (roll-off), spectral decline of reverberation time (treble decay), and delay between direct signal and reverberation signal (predelay). Time-variant random modification of delay or timbre emulates movement and the liveliness of reverberation.

Another digital technology for reverberation is sampling. The impulse response of a room can be instrumentally sampled and then convoluted with a dry source signal. Such samples contain the true response of a natural room like a fingerprint. Convolution requires substantial calculation power, for instance, $\approx 3 \times 10^{10}$ operations per second for a 5 s impulse response sampled and convoluted at 48 kHz. Fast convolution algorithms [12.9] help to reduce the number of operations but even then convolution-based reverberation only worked on dedicated hardware in the 1990s (Sony DER-777, Yamaha SREV1). Today, sample-based reverberation also works on desktop computers and audio work benches (AWB). The sampling approach, in its simplicity, has only a few parameters to control, e.g., predelay and reverberation time.

In the production workflow, reverberation serves different objectives. Dry sources – sound sources with little or no reverberation – such as a synthesizer sound or a spot microphone need additional reverb. Reverberation technology is also used to enhance or correct the reverberation that comes with the recorded natural reverberation. Sound engineers are cautious not to take too much natural reverberation during the recording, because there is no convincing way of eliminating surplus reverb. Adding reverb, however, is convenient. Reverb is eventually used to emphasize individual instruments, such as a snare drum or guitar. Finally, reverberation is used to fuse the final mix by a common virtual room. In film sound productions, reverberation is used to match voice and sound effects with the ambient noise or with the visual impression of a film scene. For dubbing tasks, reverberation is the key to retrospectively place the dubbing voice into the given scene of the actor in an authentic way.

Reverberation equipment most commonly processes mono, stereo, or surround signals. For stereo signals, a mono signal is derived from the input signals and the reverb is added to both channels.

12.2.8 Vocoder

Vocoders were initially engineered in the telecommunications industry to encode speech signals for efficient transmission. But early in the 1940s, vocoders were already being used to enhance sound. The main idea is to decompose the source and filter parameters of the glottis and the vocal tract to finally transmit parameters. These are used to recompose the voice signal based on a common model. The parameters are: voiced/unvoiced sound, pitch frequency and formant frequencies. On the parametric level, individual components can be substituted. For instance, a guitar can be played to drive a human voice.

12.3 Digitalization

Digitizing signals is a two-step procedure: first, taking samples from a signal at constant time intervals Δt, or at constant rate $f_s = 1/\Delta t$, and second, quantizing the sample amplitude in the value domain with a specific resolution, the quantum q. Therefore, the digital signal is discrete in two domains (Fig. 12.14).

For every sample taken, the digital representation of the signal contains a superposition of the original signal $x(n\Delta t)$ and the quantization error. The digital sound vector $x[n] = x(n\Delta t) + q[n]$ will eventually be converted back to the analog domain for playback. The quantization noise, which is part of the digital signal, will then be converted as well and will be part of the played back audio signal. This quantization error is usually statistically independent from the input signal and therefore the sign of the maximum $q/2$ noise contribution in individual samples is random. As shown in Fig. 12.14 the sign might change rapidly, effectively causing noise of high frequency, or less rapidly, causing noise of relatively low frequency. The spectral density is therefore assumed to be uniform from DC to $f_s/2$. And the energy contained in the noise signal is equivalent to $q^2/12$, while the energy of the signal $x[n]$ is equivalent to the square of the efficient level expressed in multiples of q. Since signals peak at maximum $\pm q^{k-1}$ for bipolar representations of amplitude and for word size k, the ratio of signal to quantization noise is limited by the signal character and the word size. Roughly, each of the k bits contributes 6 dB to the signal-to-noise ratio (SNR). A 16-bit audio format has a potential of 96 dB SNR. However, this maximum value is not achieved since:

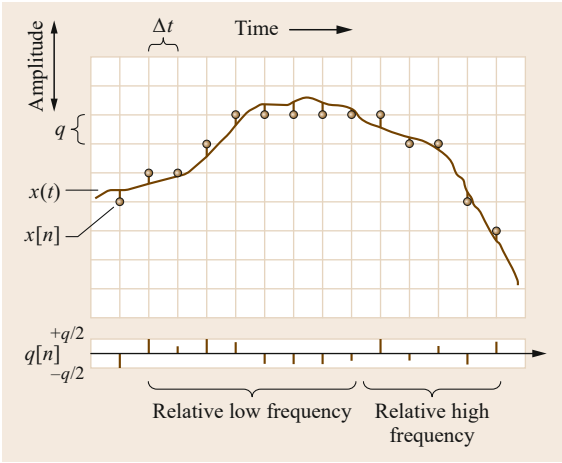


Fig. 12.14 Analog signal $x(t)$, its digital representation $x[n]$ and the quantization noise $q[n]$

1. The careful sound engineer will allow some headroom during recording
2. Signals are transient or they strongly decline and therefore contain little average energy
3. The dynamics of a musical piece already demands a good portion of the total dynamics and silent passages suffer from a lower SNR
4. Each digital processing step and its inherent rounding implies supplemental quantization noise.

In the time-domain, the sampling of a signal can be considered as some kind of gating. An ideal sampling model begins with the multiplication of the input signal with a periodical Dirac impulse. System theory suggests that such multiplication in the time domain is equivalent to a convolution in the frequency domain. The Fourier transform of a periodic Dirac impulse is a spectrally periodic Dirac impulse and, if convolved with a band-limited input signal, forces periodicity of the spectrum of the sampled sound.

This principle is visualized in Fig. 12.15 and formally expressed by

$$\begin{array}{ccccc} x_s(t) & = & x(t) & \cdot & \delta_{\mathrm{PER},\Delta t}\,, \\ \updownarrow \mathrm{FT} & & \updownarrow \mathrm{FT} & & \updownarrow \mathrm{FT} \\ X_s(f) & = & X(f) & \otimes & \Delta_{\mathrm{PER},f_s}\,, \end{array} \qquad (12.7)$$

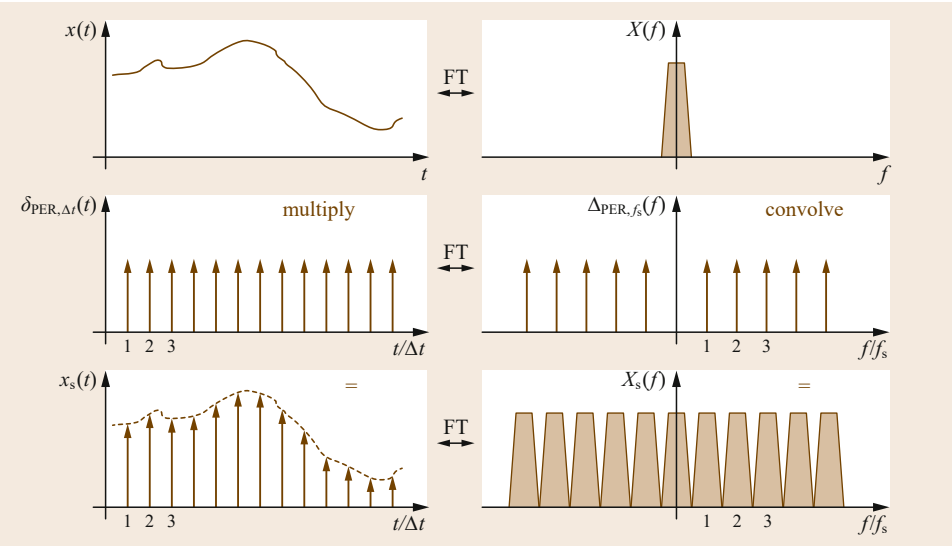


Fig. 12.15 Analog signal $x(t)$, the sampled signal $x_s(t)$ in the time domain and the respective signals $X(f)$ and $X_s(f)$ in the frequency domain

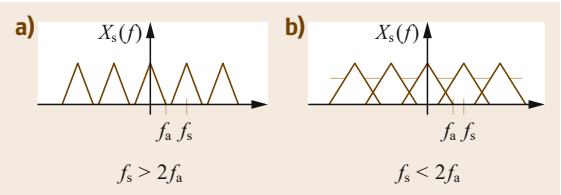


Fig. 12.16a,b Spectrally perpetuated signal without **(a)** and with **(b)** unwanted overlap

with the Fourier transform (FT) mutually relating the time domain and the frequency domain with each other.

The nature of sampling not only implies periodicity of spectral components, but also their superposition. Figure 12.16 shows an example of a signal with properly limited bandwidth that results in no overlap (Fig. 12.16a) and an example of an inadequately limited bandwidth resulting in overlap (Fig. 12.16b). Such overlap is easily audible. Therefore, the basic sampling theorem of Shannon asks that the maximum frequency of the analog input signal, f_a, is limited to half of the sampling frequency f_s.

$$f_s \geq 2f_a \tag{12.8}$$

A priori band limitation of signals is usually achieved with antialiasing filters, with the cut-off frequency set at $f_s/2$. However, the desired attenuation can only be achieved to a certain degree. For instance, the typical fourth-order low pass filter has an attenuation of only 27 dB at f_s, i. e., 3 dB at the cutoff-frequency ($f_s/2$) plus 6 dB per order one octave above f_s. High-end studio equipment not only uses well-defined antialiasing filters but also the shape of the samples to control out-of-band noise. The sampling model above can be extended by convolving each Dirac sample with a certain impulse shape, $i(t)$, which in return implies a multiplication of the periodical spectrum with a weighting function $I(f)$, which is, again, the Fourier transform of the impulse shape $i(t)$.

$$\begin{array}{ccccccc} x[n] & = & x(t) & \cdot & \delta_{\mathrm{PER},\Delta t} & \otimes & i(t)\,, \\ \updownarrow \mathrm{FT} & & \updownarrow \mathrm{FT} & & \updownarrow \mathrm{FT} & & \updownarrow \mathrm{FT} \\ X_A(f) & = & X(f) & \otimes & \Delta_{\mathrm{PER},f_A} & \cdot & I(f)\,. \end{array} \tag{12.9}$$

Figure 12.17 illustrates such an extension using the abstract rectangular shape for $i(t)$, effectively shaping the spectrum with the corresponding Fourier transform, a $\sin(f)/f$ function, while the zero crossings can be defined by the impulse width.

12.3.1 DM and SDM

Besides these so-called Nyquist samplers, oversampling technologies are widely used today, for instance delta modulation (DM) or sigma-delta modulation (SDM). Instead of the two-fold sampling frequency (12.8), DM and SDM use oversampling, i. e., a 32-, 64-, 128-, or 256-fold sampling frequency. The principle of periodicity remains. However, the size of the periodically multiplied spectral slices – as defined by the sampling frequency – is much wider, and therefore, the problems around the antialiasing filter are much more relaxed and noise is effectively reduced.

The principle behind DM is that the absolute value of an input is not quantized but the difference between successive samples. Therefore, the word size does not

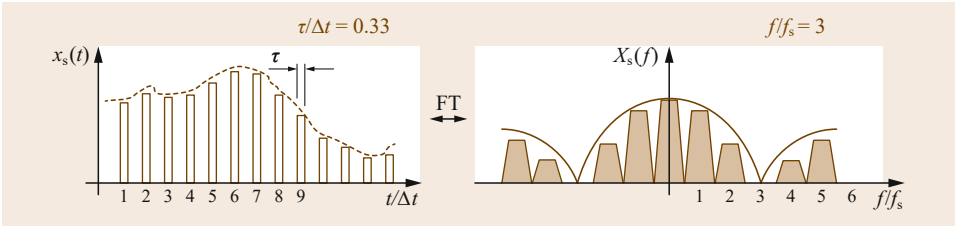


Fig. 12.17 Sampling with impulse shaping and the respective spectral weighting function

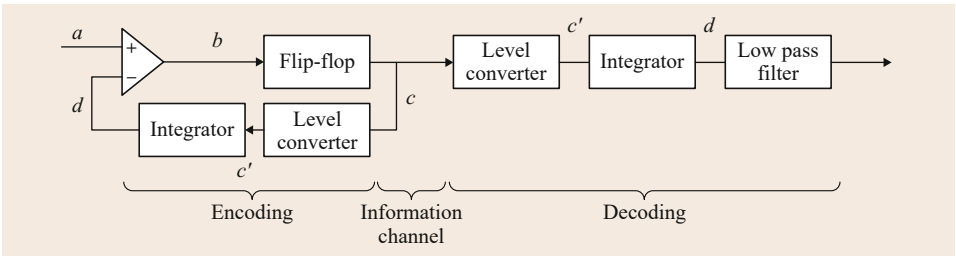


Fig. 12.18 Schematic of a delta modulator

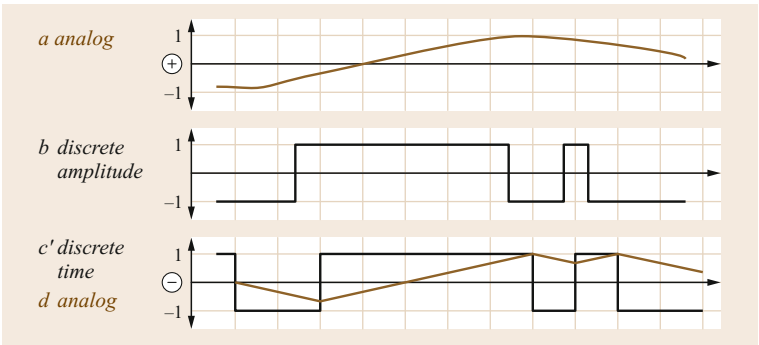


Fig. 12.19 Signal d follows input signal a by up/down integration, as identified by signal b after comparison

need to cover the full input range, but only the expected maximum difference between samples. This is also the basic principle of data compression by prediction. Here, the word size is even reduced to only one bit through oversampling, or in other words, the sampling rate is so high that the difference between successive samples is less than a bit. The analog input signal a in Fig. 12.18 is compared with the signal d, which is constructed stepwise to follow signal a. Signal b directly outputs the result of the comparison between a and d (Fig. 12.19). Signal b is latched by a flip-flop to finally constitute the digital signal c. The level-converted digital signal c' controls the integrator to further follow signal upwards or downwards, as a result of comparison. In the receiver, the corresponding level conversion and integration steps will constitute the same signal d, which will represent signal a after some additional low-pass filtering.

DM has been extensively used in speech transmission. Its signal fidelity is limited by the slope, which is in return predefined by the converter and integrator's time constant. Slope overload is avoided by an extension of the sampling principle, where several predefined slopes can be iteratively adapted to the input. Adaptive delta modulation (ADM) is widely used today.

Another way to prevent slope overload is to use an additional integrator – or sigma – before signal comparison (Fig. 12.20 for SDM). The additional integrator in the signal chain asks for an additional complementary differentiation step, which is obtained by omitting the integration step on the decoder side. While the density of ones and zeros in a DM bit stream represents the positive and negative slopes of the analog signal, in a SDM bit stream it represents the absolute signal level of the analog signal.

In terms of noise, SDM has an additional advantage over DM. While the transfer function of the input signal is unity, the transfer function of the quantization noise is $H(z) = 1 - z^{-1}$, due to the feedback loop. This implies high-pass filtering of the quantization noise. For instance, in systems with a sampling of $128 \cdot 20\,\text{kHz}$, the noise at 20 kHz is additionally reduced by 26 dB and the noise at 5 kHz is reduced by 40 dB. SDM systems are

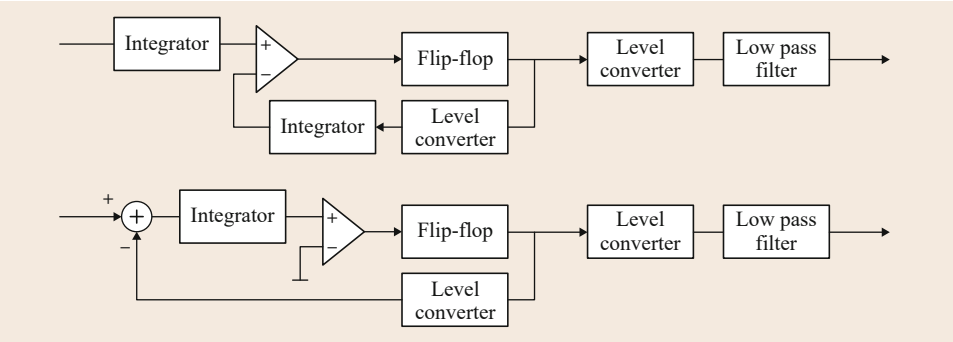


Fig. 12.20 Schematic of the sigma-delta modulator

usually implemented in three cascades and the resulting low-pass filter of the order of three attenuates noise by 78 dB at 20 kHz and by 120 dB at 5 kHz.

The SDM bit stream can be stored directly, e.g., in the superaudio CD format (SACD). However, it must usually be converted to linear formats of, for instance, 16, 20, or 24 bits by decimation in order to allow for standard signal processing and mixing technologies to be applicable. While the aliasing problem is more relaxed with SDM, the art of achieving fidelity lies in the decimation filter, which can vary from simple averaging to sophisticated higher-order approximation of the underlying input signal. Decimation reduces the sampling rate while increasing the word size.

12.4 Mixing Consoles

Mixing consoles are used to mix a larger number of input signals to fewer output signals, such as stereo channels or PA channels. They are typically arranged in a group of input signal channels, one for each input, a set of signal channels to arrange subgroups, and signal channels for mastering the outputs. An *8-4-2* console has 8 input channels, 4 subgroups and 2 master channels. In split consoles, all channels are dedicated to one task and are separately organized, while inline consoles combine input and subgroup tasks within one signal channel and are able to process two signals simultaneously.

An input signal channel typically comprises an input gain control, equalization, a control for effect send, monitor send, and auxiliary send, panning and a fader for the mix. The input gain adjusts the input level. For instance, a microphone signal at −50 to −30 dBu will be amplified to studio reference level at +4 dBu (1.23 V) or +6 dBu (1.55 V). For dBu, see Sect. 12.7.2. The reference is 0 dBu = 0.775 V. Equalization has been adressed in Sect. 12.2.3. In mixing consoles, typically only two or three frequency bands will be controlled. The signal is then amplified and spectrally shaped, if necessary, and is ready for the mix and for further processing. The signal will be used for musicians' monitoring, for effects, for recording or will be assigned to subgroups. Some signals require delay rather than additional effects, for instance spot microphones, or delay lines in a PA setup where PA speakers far from the stage must be delayed according to the Haze effect to avoid irritating the source location. For all of these purposes, auxiliary outputs, but also specific effect-send and monitor-send outputs are provided and controlled by dedicated faders. Signals returned from effect channels can often be inserted inline within the signal channel. So-called inserts deliberately detour the signal over an external device. A panning fader directly maps the signal to the stereo master channels. The sliding fader finally controls the level and therefore the contribution to the mix within a group or the master signal. Consoles differ in the way the auxiliary outputs are arranged in relation to the fader. There are prefader and postfader outputs, sometimes with options. The signal flow within the consoles is arranged by matrices and busses.

Subgroup signal channels are similarly arranged. However, they usually do not have equalization, but rather additional inputs and inserts. Subgroups are directly panned to the stereo master channels and controlled by sliding faders. Stereo panning principally employs both binaural factors for localization, interaural time differences (ITD) and interaural level differences (ILD). However, in mixing consoles, the level difference is commonly the only employed principle due to the desired monocompatibility because stereo signals with time differences between the two channels would cause undesirable comb-filter effects in mono downmixes. There is no standard on how to dose the level for

Table 12.5 Standards for metering signals

	DIN scale	NORDIC scale	British scale	EBU scale	Digital peak meter
Standard	DIN IEC 60268-10 type I		DIN IEC 60268-10 type IIa	DIN IEC 60268-10 type IIb	IEC 60268-18
Scale range	−50 . . . +5 dB	−42 . . . +12 dB	0 . . . 7	−60 . . . +9 dB	
Reference level	*0 dB*	*0 dB*	*6*	*+9 dB*	*0 dB*
Reference voltage	1.55 V	0.775 V	1.94 V	2.18 V	
Integration time	5 ms	5 ms	10 ms	10 ms	1 sample/10 ms
Hold time	1.7 s	1.7 s	2.8 s	2.8 s	1.7 s

the left and right channel in relation to a *linear* panning scale. In the middle position, where −3 dB would seem appropriate, some consoles set the level at −4.5 dB for both channels.

The master channels are organized as stereo channels in most cases. They are equipped with a sliding fader, additional inserts, and level metering. The most common forms of level metering are the volume unit meter (VU) and PPM. A VU meter is usually implemented in the form of a coil instrument and roughly images subjective loudness. It has a range of −20 VU to +3 VU, with the reference of 0 VU = +4 dBu (1.23 V). The VU meter is calibrated to sinusoids and is slowly acting. The integration time is between 35 ms and 300 ms and is therefore not a good choice to control digital clipping. PPM is a better choice for controlling transient overload, since the integration time is between 5 and 10 ms. Some PPMs are equipped with a fast option of only 0.1 ms. There are several different European traditions and standards for metering, which are summarized in the DIN IEC 68268-10 (Table 12.5).

A talkback channel from the control room to the studio or to the stage is a common feature in the master signal channel as well as headset monitoring.

Digital mixing consoles in professional environments are standalone devices, like the analog mixing consoles, and often supplemented by modern hard-disc recording systems. In semiprofessional environments they might be represented by as little as specific software that runs on a standard desktop computer with external microphone preamps and external ADC and DAC devices. The digital console offers flexibility in terms of configuration, complexity and signal processing steps. However, the shared processor and memory resources must be carefully assigned and monitored.

12.5 Synthesizer and Sequencer

A synthesizer is known to be an electronic system to create sound. The history of synthesizers began with hardware-only systems, such as the Theremin, the Trautonium or the Hammond organ in the 1930s and evolved to computer-based software systems, such as Reactor or Max/MSP. Synthesizers are used for both emulation of existing instruments and creation of fancy new sounds or timbres. Often enough, the creativity of technicians and musicians brought along instruments with original sound characteristics, such as the Moog and its little brother, the Minimoog. Early synthesizers were monophone and had to be played manually for every single performance, in the 1970s polyphone systems and systems with program and storage capacities facilitated complexity and reuse.

In correspondence with musical instruments and natural sound creation, the electronic sound creation often follows the source-filter model: a generator process followed by filtering and eventually followed by a resonator or a model of radiation.

For pitched musical sounds, the source is generally an oscillator that generates an impulse train, for instance the jet stream cycle at the reed, the glottis or the labium, or the Helmholtz motion of a bowed string. Even the plucked string generates an impulse train at the bridge. Sharp impulses inherently host a wide spectrum and the repetition of pulses causes the repetition of impulses in the spectrum or the harmonic structure of pitched sound (12.7). These impulses are not necessarily of the same level. The form of the individual impulses rules the general spectral decline (12.8). The plugging or bowing position on a string already introduces formants, irrespective of additional subsequent spectral shaping by resonances or damping. The resulting envelope represents the outcome of the filter part. For voice, the related formants host the vowel quality, for musical instruments the envelope hosts the timbre.

Synthesizers use similar construction principles; the basic elements are oscillation, filters, and modulation.

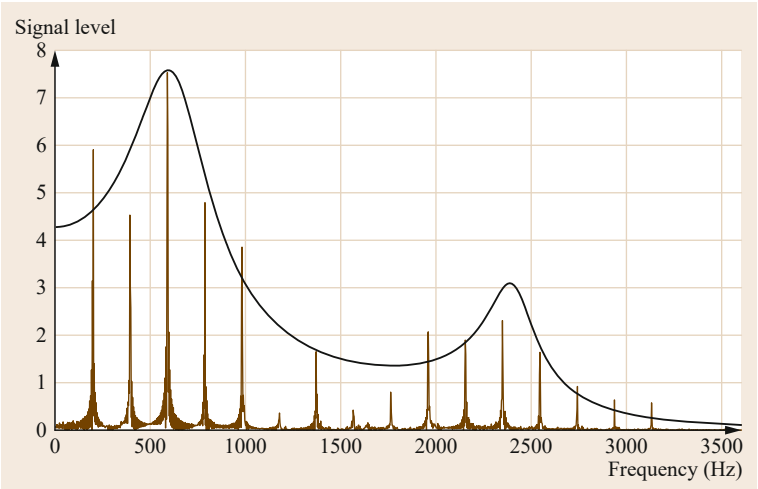


Fig. 12.21 Harmonic spectrum of a 200 ms long section of sound for the note g played on a cello, radiated sound recorded by microphone and the approximation of the spectral envelope by linear prediction (LP) of order 8

Oscillators might generate harmonic wave shapes or impulse trains. As already outlined, impulse trains inherently host a harmonic structure. Periodical functions, such as a saw tooth or a rectangular function host a harmonic structure, too, as the commonly known Fourier correspondences suggest. A voltage-controlled oscillator (VCO) usually generates a harmonic wave shape while the frequency is controlled by an input voltage. Therefore the pitch can be controlled or the sound can be modulated, e.g., by vibrato. Oscillators can be built by analog circuits or by digital loops.

Filters allow equalization, spectral shaping, phase modulation, or the construction of specific formants. For instance, the spectrum of the cello in Fig. 12.21 can be interpreted as having two formants. Time-invariant filters can be used to create the flanger or the phaser effects, see Sect. 12.2.6. Again, filters can be implemented in analog and digital fashion.

In a modulator, the multiplication of signals results in spectrally wide or less distinct signals, or in a frequency shift. The effect depends on the frequency of the modulating signal $s_m(t)$ and of the modulated audio signal $s_a(t)$.

$$s(t) = s_a(t)\, s_m(t) = s_a(t)\, \hat{s}_m cos(\omega t + \phi) \quad (12.10)$$

FM synthesis has its origin in broadcasting and has been used for sound synthesis since the 1980s. With FM synthesis the argument of the trigonometric function is modulated rather than the amplitude.

$$s(t) = \hat{s} \sin(\omega_c t + I \sin(\omega_a t)) \quad (12.11)$$

This standard notation is quoted to understand the nature of frequency modulation. In practical application, the trigonometric function in the argument of the other trigonometric function is directly replaced by an audio signal, which in return might represent superimposed harmonic functions. The nature of the nested function is such that modulation creates a series of sidebands at multiples of ω_a around the carrier with frequency ω_c. The amplitudes of the sidebands are determined by $J_k(I)$, the k-th Bessel function of first kind, while I is the modulation index and a parameter to the synthesis. A main sound effect of FM synthesis is vibrato. Another effect is the creation of a harmonic structure to sounds that only come along with a fundamental frequency.

Additive synthesis is the principle of piecewise construction from elementary periodical functions and allows for the creation of a harmonic structure. For example the Hammond organ uses tone wheels for the fundamental tone and each of its harmonic overtones. The player controls the timbre by activating respective registers. The Hammond organ by construction hosts a summation point for many signals. Today's digital versions of a Hammond organ reuse the additive method.

In subtractive synthesis, a spectrally wide signal is disposed to achieve the desired sound. Referring to the cello example, repetitive sharp impulses initially host all harmonic overtones without spectral shaping and the following filter, the one that approximates the envelope in the above example, introduces the formants and effectively subtracts all harmonic overtones above 3500 Hz, which finally brings the synthesized sound close to its natural paragon. The Moog synthesizer uses the subtraction method.

In granular synthesis, sound elements of smallest possible size in the time and the frequency domain are used to construct sound. The grains consist of a harmonic signal windowed by a Gaussian envelope, for instance. The widely-used short-time Fourier transform (STFT) also uses windowing for the sliced input vector,

and the output coefficients reveal how well the windowed input relates to a harmonic signal. In contrast, the inverse transform or synthesis rebuilds the original signal by adding all tiny elements together. In (12.12) the window can be thought of as being applied to the sliced input familiar in STFT. It can also be thought of as being applied to the harmonic function, representing a grain. The same is true for synthesis. The windowed complex harmonic function serves as a grain for synthesis, while the coefficients $F(\omega)$ select the type of grain and control the weight.

$$\text{FT} \quad F(\omega) = \int \overbrace{f(t)\, w(t-\tau)}^{\text{windowed input}} \mathrm{e}^{\mathrm{j}\omega t}\mathrm{d}t \quad \text{analysis}$$

$$\text{IFT} \quad f(t) = \int F(\omega) \underbrace{w(t-\tau)\, \mathrm{e}^{\mathrm{j}\omega t}}_{\text{windowed weighting function}} \mathrm{d}\omega \quad \text{synthesis} \tag{12.12}$$

Such composition allows for the synthesis of tones with harmonic overtones, since grains can be concatenated in an overlapping fashion to establish continuous pure tones. But transients can also be created in the same manner since transients can be transformed into the frequency domain as well. Strictly speaking, the granular synthesis is a special form of the additive method. The analysis/resynthesis method can also be applied with other transforms, such as the wavelet transform, or with analysis methods that decompose into a representation of noise plus transients plus sinusoids.

For the sampling method, samples of desirable sounds are taken by recording and are used as elementary functions in an additive and/or parameterized approach. Resampling is one of the methods used to adjust the pitch for a given sample. Pitch and envelope are typically controlled separately.

In physical modelling synthesis, physical representations of a musical instrument are used to create sound numerically. This might be a discrete geometrical model with physical properties using numerical methods such as the finite element method (FEM), finite difference method (FDM) or boundary element method (BEM). Alternatives are the above-mentioned source-filter method and the wave-propagation model.

The reader is referred to *Poli* [12.10] for an overview of musical signal representations.

The objective of synthesizing sounds with fidelity or creating sounds of sensation deserves a brief discussion. First, Fig. 12.21 shows – on purpose – the envelope of the nonregular harmonic spectrum of a real instrument. The approximated envelope does not represent the irregularities in the structure. Fluctuations in the tone cause an ever-changing sound texture and the analyzed spectrum changes for every section of sound. Irregularities and fluctuations are well perceivable to listeners and significantly add to the desired senstion of tone.

These, however, ask for additional rule sets or modelling under the synthesis paradigm. Second, the noise floor, or in general the nontonal components in a sound are not yet covered by the sinusoids. Such components must again be modelled separately.

Third, the harmonic spectrum is not as regular as it first appears. The stiffness of strings, for example, causes a frequency shift upwards for higher overtones. This is significant. For example, in the middle register of a piano, the frequency of the 17th harmonic is already a factor of 18 above the fundamental rather than a factor of 17. Such effects again ask for extended modeling.

Fourth, the above-mentioned components and methods are likely to produce sounds where the partials are in a coherent relation to each other. Coherence implies stationary phase conditions between signals. For instance $\cos(\omega t)$ and $\cos(2\omega t)$ are coherent, as are $\cos(\omega t)$ and $\sin(3\omega t)$. Loops or mutually synchronized loops typically produce statically arranged sounds, which are likely to cause only little sensation. Breaking up the coherence again requires additional rule sets.

Finally, digital modelling ignores the impact of impedance. Digital signals can be copied without losing power. This is not possible for analog signals. The source impedance is a major factor for signals under load. For instance, the tone wheel of the Hammond organ carries a sinusoid wave shape and the digital model of it therefore only loops a digitized sinusoid. On the real instrument, however, pressing a key implies that a short-circuit coil will be opened and the induced voltage faces the high impedance of the open-circuit coil at a particular instance, which causes the well-known *dirty* sound attack of the Hammond organ. Modelling a signal by levels, i. e., voltages or currents, is not enough to emulate such effects.

12.5.1 MIDI

In studios, a variety of control signals are used to remotely control devices or to switch signals. Among these control-signal-only formats is the Musical Instrument Digital Interface (MIDI). It does not carry audio signals, but key parameters of sound, such as pitch, onset and dynamics. It was designed to interface musical instruments, in particular synthesizers, but mi-

Table 12.6 MIDI control signals for $n = 0$ to 15 individual channels. Data bytes cover the range 0 to 127 (always headed by 0)

Status + MIDI channel	Data byte 1	Data byte 2	Action	Description
8*n*	*kk*	*vv*	Key off	Stop playing note *kk* apply release velocity *vv*
9*n*	*kk*	*vv*	Key on	Start playing note *kk* apply dynamic level *vv*
A*n*	*kk*	*vv*	Polyphonic aftertouch	Apply dynamic level *vv* to the already played note *kk*
B*n*	*cc*	*vv*	Control change	Change operational Status *cc* of a controller to the value *vv*
C*n*	*pp*		Program change	Assign musical instrument *pp* to channel *n*
D*n*	*vv*		Monophonic aftertouch	Apply dynamic level *vv* to all notes in channel *n*
E*n*	*vv*	*ww*	Pitch bending	Pitchwheel Status set to the value *wwvv* (14 bit)
F0...7	*xx*		System message	Specific control data *xx* for up to eight devices

grated into other applications, such as effects, sampling, fading, panning and even the control of lightshows. Composing, arranging and music writing have taken advantage of the direct back-path from keyboard to music.

MIDI is a serial interface operating at 31 250 bps and using DIN5 connectors. A control command consists of three bytes that need a total of 960 µs for transmission. For example, the three bytes are (i) on- or off-keying, (ii) key/note and (iii) dynamic level (Table 12.6).

Apart from the essential control data for the music to be played, there is a wide range of sound controls specific to the employed devices (B*n* in Table 12.6). Among the 127 mostly predefined commands are the control of the modulation wheel, pedal, portamento, sostenuto, legato, harmonic content, attack time, release time, brightness, stereo balance and effect control, i. e., depth of tremolo, chorus or phaser. MIDI also facilitates synchronization of devices so that several playback devices, synthesizers, keyboards and recording devices can act together. The MIDI clock subdivides each beat of a predefined beats per minute (BPM) into 96 ticks. The limits of MIDI are the control of scales when not tempered and the speed of control. Standard MIDI files (SMF) are indicated by the *.mid or *.kar extension. MIDI may also be captured in RIFF-RMD files with the extension *.rfi or within the Extensible Music Format (XMF). The sample dump standard (SDS) combines MIDI together with linear PCM audio for coordinated playback and synthesis.

Sequencers are, in principle, recording and playback devices for sequences of MIDI control signals, e.g., the complete arrangement of a song. Hardware sequencers are standalone devices with control through front-panel functions and a keyboard. Sometimes a sequencer is already hosted within a keyboard. Software sequencers work on the basis of computers and go beyond programming, recording and playback, since they also employ graphical interfaces for editing and audio sample processing. Such a combination of editing audio and MIDI together is referred to as a digital audio workstation (DAW).

12.6 Historical and Contemporary Audio Formats and Restoration

This section reviews historical audio formats and restoration methods for the study and research of audio archives. Basically, historical formats are analog, while contemporary data formats are digital, as further discussed below. There are, of course a few exceptions to this rule. For instance, at the beginning of the 20th century, piano music by Mahler, Strauss and other composers was recorded using punch cards and these *digital* recordings can still be reproduced on the so-called pianola.

12.6.1 Historical Audio Formats

Wax cylinder, shellac and vinyl records share the same idea of grooving the analog signal into material, a mechanical representation of the signal. Formats and technologies have changed over the course of 100 years (Table 12.7). Admittedly, these formats are not widely used today but they retain their importance due to the heritage of historical recordings.

Recording on steel wires began in 1889 with wires 1 mm thick at a recording speed of $v = 20\,\mathrm{m/s}$. From

Table 12.7 Important historical storage formats

Substrate	Year	Size	Speed	Channel/groove
Wax cylinder	1902–1929	10.5× Ø 5.1 cm	160 rpm	1 Vertical
Shellac	1904–1960	Ø 10″/12″	78 rpm	Vert. or horiz.
Vinyl	Since 1930	Ø 12″	33 1/3 rpm	1 Horizontal
Vinyl	Since 1949	Ø 7″	45 rpm	1 Horizontal
Vinyl	Since 1958	Ø 7″/12″	45/33 1/3 rpm	2 Diagonal 'x'

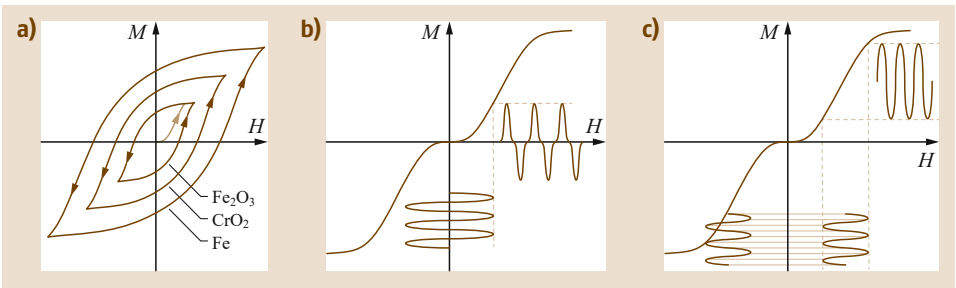


Fig. 12.22 (a) Magnetization M versus magnetic field intensity H for different tape coatings, (b) remanent magnetization M versus magnetic field intensity H and signal modulation, (c) id., but with a biased modulating signal to shift the modulation into the linear region

1930 to 1954 the BBC worked with Poulsen's steel band machine. The 3000 m long steel band was 3 mm wide and 0.08 mm thick. The speed was 1.5 m/s, delivering 4 kHz bandwidth and a dynamic range of 25–30 dB.

Storage of analog signals on magnetic tape promised improved sound quality. However, magnetization implies the problem of nonlinearity. The magnetization of a tape in relation to the magnetic field is nonlinear. The magnetization follows a hysteresis function when excited by a bipolar magnetic field (Fig. 12.22).

Important for the nonlinear effects of recording in combination with playback is the remanence of the magnetic tape. The remanence is the magnetic strength that remains in the tape after releasing the magnetic field (values at $H = 0$ in Fig. 12.22a). An input signal that drives the magnetic field will result in a recorded signal that is distorted due to nonlinearity (Fig. 12.22b). Typically, a third harmonic will develop. To avoid such nonlinearity, a bias signal with a frequency well above the audio frequency band is added to the input signal during recording. The operating range is effectively shifted towards the linear region of the remanence curve (Fig. 12.22c). Nonlinearity cannot be fully avoided but a properly adjusted bias can reduce the distortion factor significantly. Adjusting the bias for minimum distortion, however, reduces the output signal during playback and so the achievable dynamic range, especially for higher frequencies of the audio signal. Therefore, a sound engineer in those days had to compromise between a high dynamic range and low distortion. For this purpose, professional recording systems facilitated manual control of the bias amplitude.

The high-frequency content of an audio signal already has a similar effect to the bias signal and the two work together in terms of an effective bias. On the other hand, high-frequency components of an audio signal suffer from too much bias. The HX Pro System was one of the approaches to adjust the necessary bias in relation to the recorded input signal. Another important impact factor to bias adjustment is the type and the quality of the tape used. Professional recording machines can measure the necessary bias. In summary, analog tape recordings reveal individual quality in terms of dynamic range and distortion.

Magnetic tape playback suffers from several losses. First, the air gap between the tape and playback head is approximately 1 µm. The physical size of one wave of a 15 kHz signal recorded on a tape at $v = 4.75$ cm/s is only 3.17 µm. An air gap of 3 µm would dampen the signal by roughly 50 dB. Second, the size of the gap in the playback head will define losses. The damping follows a $\sin(x)/x$ curve with the zeros at multiples of b_{eff}/λ, where b_{eff} = effective head gap, λ = geometric size of the wavelength on the tape, typically 2.5–12 µm in the recording head and 1.5–6 µm in the playback head for professional systems. Third, a mismatched position or angle of the head gap will introduce further losses. Therefore, working with old recordings may require adjustments on the playback head. Fourth, tapes are rolled on a reel and the tape sections that are layered upon each other will mutually magnetize each other. The resulting pre- and postecho can be heard. This copy effect depends on the frequency and is strongest for frequencies at about 400–1000 Hz. Mag-

netic density grows with frequency therefore the copy effect also grows with frequency. However, the distance between the layers in relation to the physical size of a recorded wavelength raises with frequency and therefore the copy effect again declines with higher frequencies.

Playback of tapes of unknown origin implies further uncertainties. Usually, high-frequency components are amplified for recording up to +20 dB. Such pre-emphasis and the complementary de-emphasis during playback should match each other. However, the pre-emphasis selection is subject to the tape quality and the tape speed v, according to different standards, and it will be aligned to an assumed amplitude statistics of the input signal (Table 12.9). As there are different filter time constants defined for de-emphasis, there should be clarity about the tape type used and the standard followed during recording to match the de-emphasis. The often used dynamic compressor Dolby B is yet another recording feature that adds uncertainty since it asks for complementary expansion during playback. In summary, without knowledge of the specific recording parameters, the playback may differ from the intended recording in terms of spectral balance and dynamic.

There are tapes from 1/4 to 2 in width for tape speeds v from 1.875 to 30 in/s (Tables 12.8 and 12.9).

Table 12.8 Magnetic tape widths and track numbers

Tape	Tracks
Compact cassette 3.81 mm	Consumer: 2 plus 2 (reverse) Philips: 1 plus 1 for timecode Tascam: 2 plus 1 for timecode /4/8
1/4″	1 track NAB: 2/2 plus 1 for timecode DIN: 2 4
1/2″	2/4/8
1″	4/8
2″	16/24

Table 12.9 Tape speeds and time constants for bias modulation

Tape speed v	Standard	τ (µs) treble	τ (µs) bass
15 in/s	NAB	50	3180
38.1 cm/s	DIN	35	–
7.5 in/s	NAB	50	3180
19.05 cm/s	DIN	70	–
3.75 in/s	NAB	90	3180
9.53 cm/s	DIN	90	3180
1.875 in/s	DIN (Fe_2O_3)	120	3180
4.75 cm/s	DIN (CrO_2)	70	3180

In the 1980s digital audio tape (DAT) formats were also used. R-DAT is based on video recorders with rotating heads, with the following operating modes:

1. 2 channels 16-bit linear at 48 kHz sampling rate
2. 2 channels 16-bit linear at 32 kHz
3. 12-bit nonlinear at 32 kHz
4. 4 channels 12-bit nonlinear at 32 kHz
5. 2 channels 16-bit linear at 44.1 kHz playback only.

The format extensions DA-88 and ADAT facilitate 8-channel recording. The parity protected 30 ms frames host supplementary data such as search IDs. S-DAT stands for digital audio recording with stationary heads. The digital compact casette (DCC) is the consumer format and is based on the compact cassette at $v = 1.875$ in/s. It uses MPEG 1 coding to reduce the linear 16-bit/48 kHz format down to 384 kbps for two channels. The tape hosts 8 channels plus 1 supplementary data channel, and another 8 plus 1 channels in the reverse direction. The professional extensions digital audio stationary head (DASH) and ProDigi work at high tape speeds to record 2 to 64 linear channels on 1/4, 1/2 and 1 in tapes. The data format works with interleaving and channel coding, so that the digital tapes can be manually edited in the same way as analog tapes.

12.6.2 Restoration

Restoration may be necessary when working with historical recordings. The reasons for derogated sound quality are damaged media, unprofessional storage or limitations or problems with the recording or playback equipment. This section gives a brief overview of typical problems and methods, although practical work in this field is usually assigned to professionals. Workbenches or software packages for restoration are often based on audio workbenches and comprise manual or automatic detection of distorted signals and tools for repair. A monitoring function allows control of the restored signals but also of discarded components to ensure that not too much of the desired sound gets lost.

The analysis identifies the type and duration of impairments, e.g., noise, tonal components, transients or typical background noise. Noise originating from the signal chain or the medium can be spectrally analyzed and a predicted spectrum will be subtracted from the signal. Passages of silence allow extraction of a fingerprint of the noise, which is more precise than a prediction. Subtraction always affects the sound, and there are specific denoisers for speech and for music. The typical downside of denoising is a loss of brilliance. Artefacts such as ringing or tonal components can be avoided by carefully selecting thresholds and parameters.

Scratches, clicks and cracks are classified by level and duration. Scratches have a large impulse level and are several milliseconds long in duration. They result, for instance, from mechanical damage on cylinders, shellac or vinyl. Noise from dust in the grooves of these media can be reduced by wet playback. Clicks are of lower level and shorter in duration. Crackles have low levels but come at a high rate. Impulse shapes are already characterized for common problems and can be automatically identified. Descratchers, declickers and decrackers cut out the disturbing impulse and restore the resulting gap by interpolation, polynomial fit or using the B-spline method.

Pop noise resulting from close-up speech or singing is often simply reduced by high-pass filtering. Clipping artefacts may result from poor input adjustment and a declipper will restore the waveform by lowering the level and interpolating the impacted section. Overemphasized voiced dental fricatives or sibilant noise are reduced by a de-esser which consists of an amplifier in parallel with a limiter, where the threshold and frequency can be adjusted.

The power supply often causes noise in amplifiers and recording equipment. Poor ground loops cause humming and power dimmers cause buzzing. A debuzzer will employ comb filters to reduce the tonal components of 50 or 60 Hz supply systems and respective multiples of these frequency components.

Tape-recording azimuth errors are caused when the recording head and the playback head are not well aligned to each other. Such differences on the order of tens to hundreds of microseconds between tracks adversely affect the stereo image and the treble range. These differences can be analyzed by correlation. Modern deazimuth tools can even detect differences as low as a small fraction of one sample and can compensate for the difference by internal oversampling. Preferably, the mismatched alignment is compensated for during playback by individually adapted adjustment of playback heads, even if the setting is abnormal.

Temporary dropouts may result from erosion of the magnetic coat on a tape, from kinks or from poor splicing. Such impairment will be compensated for by manual levelling.

12.6.3 Contemporary Digital Formats

Contemporary digital formats for storage on hard disc, optical disc or flash memory are strongly related to specific coding technologies. The overwiev of these is organized along three issues: audio quality, data type and data embedding. In terms of quality, coding reduces the necessary storage volume and is likely to reduce the sound quality as well. Second, there are different types of formats, such as raw sound data, sound data with fully embedded description for archives and formats for streaming or broadcast. Third, there are container formats that link between several audio components or between audio, video, presentations and project data. Other context formats are, for instance, proprietary formats to capture sessions in specific production environments.

To begin with the first issue, sound quality in digital formats, the prominent parameters to mention are the word size and sampling rate. As outlined above, each bit contributes to potentially 6 dB dynamic range. Linear data formats, or linear pulse code modulation (PCM), hold such data sample by sample. These formats can be edited sample-wise. Linear PCM is often hosted in WAV files and it is also the format of the compact disc. Lossless coding reduces the data rate with virtually no loss of audio quality. Only some procedures use calculations that ask for rounding, which is equal to introducing an additional quantization step, including quantization noise. An example of lossless coding is the predictive coding that uses signal differentiation, complemented by signal integration in the decoder. The format is called differential PCM (DPCM). A derivative of this method additionally adapts differentiaton increments to the type of input signal and is called adaptive DPCM (ADPCM) (ITU-T G.726) [12.11]. Another example is redundancy reduction by entropy, or variable-length coding, e.g., Huffman coding. The word size is not fixed, but varies with the probability of occurrence. Frequently appearing values are coded by short words, effectively reducing the data volume. Examples of respective formats are Free Lossless Audio Codec (FLAC), Apple lossless, Windows Media Audio Lossless and MPEG4-ALS.

Lossy coding discards components that seem irrelevant to human listening. Decoded data will therefore reveal technically observable differences to the original audio sample, but hopefully no perceptual differences. While the linear format captures the dynamic range according to the word size across the entire band from 0 Hz to half of the sampling rate, human listening is limited and the dynamic range depends on frequency. Figure 12.23 indicates the threshold of human listening according to ISO R226 [12.12]. The word size can be reduced for a wide range of frequency bands because the quantization noise cannot be heard. Furthermore, pure tones in a signal mask spectrally adjacent bands (Fig. 12.23, lines). In these adjacent bands, the word size can be even further reduced. Additionally, there are dynamical masking effects in the range of 10 to 20 ms that encourage segment-wise scaling of segments. Scaling is equivalent to achieving full-scale levelling for each segment, effectively reducing quantization noise.

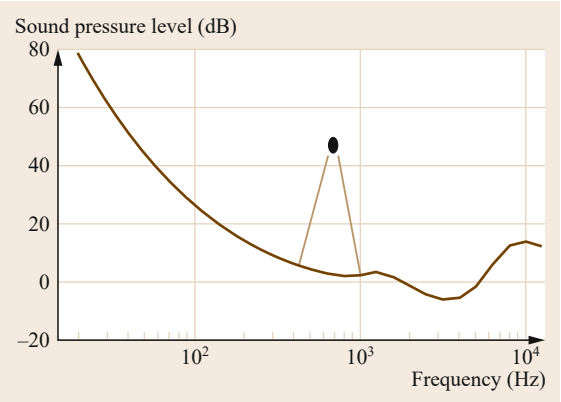


Fig. 12.23 Hearing threshold according to ISO R226:2003 (*brown*) and masking effect (*yellow*) of a pure tone (*black dot*)

Of course, such scaling will be reversed in the decoder. To facilitate word size reduction in spectral bands, the input signal is segmented into 32 (MPEG-1 layer 1) or 576 (MPEG-1 layer 3) bands by filter banks. These processing steps are supplemented by nonlinear quantization (dynamic compression), reducing the redundancy between stereo channels and final Huffman coding.

AAC and AC-3 coding follow the same method of tolerating psychoacoustically irrelevant quantization noise to minimize data volume. Instead of a filter bank, the discrete cosine transformation (DCT) is used, the spectral resolution of which will be dynamically matched to the input audio characteristics. Additional features of AC-3 coding are: dithering, spectral band expansion, karaoke mode and more. HE-ACC (or ACC+) and HE-ACC v2 (or ACC+ v2) use spectral band replication (SBR) to achieve very low data rates, namely 32 to 80 kbps and 16 to 40 kbps, respectively. Instead of encoding all tonal components of a sound, only tonal components of low frequency are encoded plus a rule set on how the tonal components of high frequency can be replicated from the encoded components.

The digital theatre system (DTS) format is defined for use in cinemas and home theatres. The code comes along with a timecode to allow for synchronization with the film, even when played from a separate CD or DVD player. The advanced DTS-HD can host linear lossless formats of 5.1 channels at 192 kHz or 7.1 at 96 kHz, with data rates up to 24 Mbps.

The second issue is the type of format. Basically, there are formats that are composed in a way to facilitate the sharing or archiving of audio data, and formats that facilitate broadcasting or streaming. Formats for data sharing and archiving typically embed associated metadata and only limited means of error protection. The amount of embedded metadata and the size of the data file are limited in some formats. Streaming formats follow the idea of an endless stream. They are equipped with powerful error correction, or channel coding, to facilitate forward-error correction (FEC) upon reception over long-distances and transport over channels of vague transmission quality. Often, the streaming formats are able to host more than just one stereo stream or just one 5.1 audio stream. They might contain an entire program of a broadcaster. Finally, metadata is usually not part of the stream but rather linked to the stream and hosted in separate files, or separate streams. The audio data stream, however, will embed enough supplementary information on a regular basis to allow a decoder to jump in at any time and start its decoding process after a negligible initialization period. Examples of audio streaming formats are briefly summarized in Table 12.10, audio file formats are listed in Table 12.11.

In terms of sharing or archiving audio data, many audio file formats emerged from the 1985 Interchange File Format (IFF) and the later Resource Interchange

Table 12.10 Digital audio streaming formats

Format	File extension	Coding	Audio channel	Sample rates (kHz)	Bit rates (kbps)	Channel coding
MPEG-1	.mpeg .mpg	MPEG-1 layer 1	1–2	32/44.1/48	Fixed 32 . . . 448	
MPEG-1	.mpeg .mpg	MPEG-1 layer 2	1–2	32/44.1/48	Fixed 32 . . . 384	
MP3	.mp3	MPEG-1 layer 3	1–2	16 . . . 48	Fixed/var. 32 . . . 320	
MPEG-2	.mp2	MPEG-2 BC	1–5	16 . . . 48	8 . . . 160	
MPEG-2 ATDS/ADIF	.aac .adif .3gp .m4a	ACC	1–5 (48)	8 . . . 96		Wide area/local
AC3 (Dolby Digital)	.ac3	ATSC/5 2	1–5	32/44.1/48	Fixed/scalable 32 . . . 640	Wide area/local
DTS (Digital Theatre System)	.dts	dts (APT-X100)	5.1	48/96	Fixed 255 . . . 1510	Local

Table 12.11 Digital audio file formats

Format	File extension	Coding	Application
WAVE	.wav	Linear PCM, DPCM	Studio, consumer
BWF	.wav	Linear PCM, MPEG plus metadata	Studio, broadcast
RF64	.wav	Linear PCM plus metadata	Studio, broadcast
AVI	.avi	Diverse audio and video codecs	Consumer, studio
AIFF	.aiff .aifc .afc	Linear PCM	Studio

File Format (RIFF). The common idea is that a preceding header provides information about the type of data (audio, video, pictures or text), then about the format, and then details of the data, which follows in a larger data field. The widely-used WAVE format is a specific RIFF format. It holds linear PCM but sometimes also DPCM and is limited to 4 GB of data.

The Broadcast Wave Format (BWF) was developed for exchange between broadcasters. The format identifies the production and informs about the digital rights of use. Extensions and supplements of this format now facilitate the recording of a coding history together with the data, but also the linking of formats for concatenating files of limited size. RF64 is a standardized format to overcome the limitation of 4 GB [12.13]. RF64 is not limited to just one audio mix and just PCM coding. It may host up to 18 channels and non-PCM coded data. Audio Video Interleaved (AVI) is also a specific RIFF format, but seldom used for just audio alone. The Audio Interchange File Format (AIFF) Is based on IFF and only admits linear PCM, while the extension AIFF-C also admits lossy coding. It also features the support of markers, loop-identifiers, MIDI-data and commentary for handy editing tasks.

The third issue is the scope of the data format and how data is embeded. The formats described above all have in common that they describe one single presentation, independent of the type – file or stream – even if they are already composed of audio and video and other content. So-called container formats not only host media of different type but also independent presentations, supplementary timecodes, structural data, text and markers. Container formats more easily link to existing technology platforms, such as players or audio production environments, and facilitate compatibility, see Table 12.12.

Quicktime, introduced by Apple in 1991, was one of the earliest formats. The package includes the data format and the encoder, but also the player, authoring and streaming technology. It is hierarchically organized to facilitate any desired link between granular data entities. The encoding of Quicktime 7 supports linear PCM, MP3, AAC, ADPCM and Apple lossless.

MPEG-4 is based on Quicktime but follows an object-oriented approach to govern independent or synchronized streams of media, or to conduct media syntheses or content analyses. The associated Delivery Multimedia Integration Framework (DMIF) covers the channel coding and other transport issues for broadcast or streaming, so there is a modular approach for the tasks of data organization and of transport. MPEG-4 audio supports ACC coding with options for low delay and spectral band replication as well as for parametric stereo to further compress data (HE-AACv2). With parametric stereo, the redundancy between channels is analyzed and modelled. Determined model parameters will conduct the resynthesis of the stereo image for the incoming audio data. Synthesis is supported by the codefined Structured Audio Orchestra Language (SAOL), the Structured Audio Sample Bank Format (ASBF) and a text-to-speech algorithm. Data streams and synthesized entities together compose a complex scene, for instance for films, in a compact way while still keeping the individual entities apart and controllable during play.

Open Media Framework Interchange (OMF or OMFI) is a proprietary format and widely used in post-

Table 12.12 Digital audio container formats

Format	File extension	Objects	Application
Quicktime	.mov .qt	Multimedia	Consumer
MPEG-4	.mp4	Multimedia	Consumer
MPEG-4 Audio	.m4a .m4b .m4p	ACC audio	Consumer
OMFI	.omf	Video, audio, project data	Studio
AAF, MXF	.aaf .mxf	Multimedia	Studio
AES-31	Fat32 files	Audio and project data	Studio
WMA ASF	.wma .asf	Multimedia	Consumer
Real Media	.ra .rm	Multimedia	Consumer
OGG, Matroska	.ogg .ogm .mka .mkv	Audio, video, text	Consumer
OpenMg	.oma .omg	ATRAC audio, DRM	Consumer

production environments by AVID and Digidesign. The further evolved Advanced Authoring Format (AAF) and Material Exchange Format (MXF) target the interchange of data for video production. The AES31 format extends the features towards production issues even further. Examples are track numbers, cuts and overlays, or control information for automation.

Windows Media Audio (WMA) covers a family of audio encoders and describes file and stream properties, content description and marker objects. Real Media came as a platform covering format, encoding, streaming server and media-client technology; it was widely used between 1995 and 2000. OGG is an open-source format specified in RFC3533, capsuling audio, video and text. Likewise, Matroska is an open-source format. OpenMG is proprietary and supports digital rights management (DRM) and encryption. SDIF was defined for composition, production and research. There are descriptors for pitch, energy, spectral coefficients, envelopes, synthesis. It supports matrices and the choice between fixed and floating-point numbers. Entities are IFF-like organized together with time-tags to guarantee synchronicity.

12.7 Signals, Connectors, Cables and Audio Networks

The transport of audio signals between devices in a production chain has its own history and complexity. First of all, there are usually control signals and also supply signals next to or superimposed to the audio signals for remote operation. The benefits, but also the problems, of using as few cables as possible created a variety of technical solutions. Additionally, the analog audio signal is more and more replaced by digital signals on a variety of alternative copper or fiber links, with a strong trend towards Internet Protocol (IP) technology.

12.7.1 Cables, Fibers, and Wireless Local Connections

The physical link has an impact on the forwarded signal. A cable is, physically speaking, a circuit of a spatially expanded capacity and a likewise expanded inductance. Therefore, cables have intrinsic impedance that depends on frequency. The spectral deficiencies of cables usually do not emerge due to the low frequency of audio.

However, cables are susceptible to external noise and distortion. Any electromagnetic field in close proximity will induce currents that will be superimposed on the wanted audio signal. Sources of distortion are radio broadcast, mobile devices, power transformers, switched electrical loads or local thunderbolts. Twisted cables are more robust against such unwanted electromagnetic interference (EMI) than parallel cables, since the twist flips the orientation of the wire pair within a given electromagnetic field and therefore flips the polarity of induced levels every few inches along the cable. A homogenous twist will therefore compensate locally.

Shielding is another protective measure. The braids usually applied around the signal wires protect against the electrical component, but not against the magnetic component of a field. Protection against the electrical field is only in effect if the braid is properly terminated by grounding. Grounding, in almost all examples of studio technology, is a simple low-impedance connection between the cable shielding and local ground. Broadband shielding requires a termination resistance that matches the impedance of the shielding, otherwise most of the signal energy will be reflected at the point of termination and will remain on the shielding, where it feeds back to the wires.

Symmetric signals are another means of protection. For any audio signal to be transmitted, its polarity-reversed representation will be generated, either by high-quality transformers or by differential amplifiers. The symmetric pair of signals will be transmitted and in the receiver the difference of these signals will bring about the audio signal and at the same time suppress the common-mode signals resulting from external distortion. Efficient common-mode suppression requires low-impedance drivers and high-impedance receivers to reach 30–40 dB in standard technology and it additionally requires adaptive circuits to reach 100 dB in professional equipment.

Digitalization aggravates distortion artefacts, because time-discrete sampling potentially shifts high-frequency distortion into the audible band. The principle of spectrally periodic perpetuation not only applies to the wanted audio component in a signal but also to all components irrespective of their spectral origin (Figs. 12.15 and 12.16). So even if the monitoring of an arriving analog audio signal seems to be fine, originally nonaudible components will be contained in the digital representation after sampling. The antialiasing filter will only attenuate such distortion for high frequencies. Digitalization therefore asks for awareness of potential sources of distortion, especially in the frequency range of 20–150 kHz, which is well within the operating range of professional audio preamps, and in which the antialiasing filter does not have a strong attenuation.

Digital signal transmission on cables is less susceptive to distortion. However, noise always causes some

residual bit errors. In local installations the bit error rate (BER) is typically between 10^{-12} and 10^{-8}. This seems small, but on a gigabit Ethernet connection with BER $= 10^{-9}$ this is equivalent to one bit error every second. Channel coding technologies allow for compensation to achieve quasi-error-free (QEF) transmission. However, coding implies significant delay and contradicts requirements of real-time applications.

Optical fibers are not susceptible to electromagnetic interference, but to damping. Glass fibers only have about 1 dB attenuation per km and 1 dB loss per connector, but they are not practical in nonstationary environments. For instance, the popular 50/125 micron multimode fibers (50 µm core with 125 µm cladding) are used in MAD optical, but also in LAN computer networks (Sect. 12.7.9). Polymer optical fibers (POF) are more practical in nonstationary and in rugged environments and are used in semiprofessional applications, but they have a considerable attenuation of about 0.2 dB/m. They are typically not used across a building or between buildings. For an overview on copper and optical links see [12.14].

FM-modulated radio links in the ultra high frequency (UHF) band (300–3000 MHz) have been popular for a long time to connect microphones with a mixing console. The typically used frequency range of 790–814 MHz is now dedicated to LTE networks and cannot be used anymore. Frequency reallocations within the broadcast market and rededication of bandwidth from broadcast to mobile communications are heterogeneous across Europe. This reorganization will last beyond 2017. Wireless applications will eventually relocate into diverse gaps. A license-less band across Europe is defined at 863–865 MHz, while the wider bands in the former TV channels 61–63 and 67–69 for professional users ask for licenses. At the same time, the wireless link for microphones has become digital. For instance, some microphones now use modern digital modulation techniques, such as frequency-hopping spread spectrum, and these operate in the unlicensed 2.4 and 6 GHz bands.

12.7.2 Signals and Grounding

Analog audio signals are referenced to a specific level to be compatible between components. Power adjustments refer to 0 dBm, which is 1 mW power or 0.775 V at an impedance of 600 Ω. The commonly used voltage reference reuses the 0.775 V level to define 0 dBu. The so-called line level of inputs and outputs is usually defined at 0 dB = +4 dBu = 1.23 V, or at +6 dBu = 1.55 V.

The phantom supply for microphones and external equipment is symmetrically fed into the signal wire pair and is defined at 48 V (Fig. 12.24).

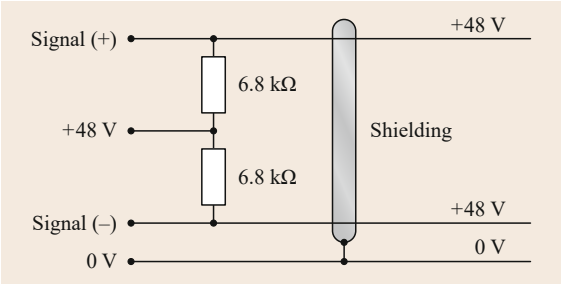


Fig. 12.24 Phantom supply to remote microphones and devices

Digital signal levels are defined by the respective standards of the corresponding digital links, such as Ethernet and others.

An example of such external equipment is the direct injection box (DI-box). These are widely used to convert asymmetric signals, such as those from a guitar pickup, into symmetrical signals (Fig. 12.25). If this conversion is done by differential amplifiers, an extra power supply is required and is conveniently cotransmitted on the symmetrical line that links with the mixing console. DI boxes also address the problem of grounding. The quality of grounding depends on matching impedances and the thoughtful selection of a sparse set of grounding points to avoid ground loops. Loops of grounded shielding should be prevented since the area enclosed in such loops relates to the level of immission. Star configurations have no such enclosed area, but a central grounding point. Therefore, the option of omitting the grounding in DI boxes is helpful in practice.

12.7.3 Digital Connections

In terms of digital connections, there is a wide variety of technical solutions. Some systems target studio-only applications, basically replacing signal lines by digital lines. Some systems target digital workflows and transmissions beyond a studio. Some systems are open and flexible to allow for optional replacements of modules; some proprietary systems are not.

12.7.4 The OSI Model

The open standard interconnection (OSI) structure for communications is useful to understand the scope and potential of solutions. The reader is briefly introduced to its seven layers. The physical layer represents the physics of transmission, which covers cables, antennas, fibers, modulation, connectors and clock regeneration. It works on a bit level, without any comprehension of meaning. The data link layer works with frames or

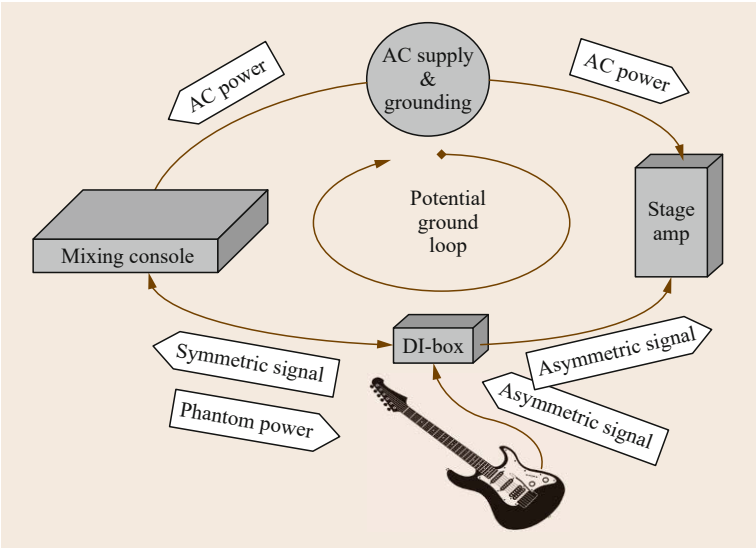


Fig. 12.25 DI-box application and potential ground loop

blocks of data ensuring the proper delivery between two connected points. This implies the error-detection or error-correction, flow control, monitoring of sequences of frames and losses. On this layer, a lost or erroneous frame might be automatically repeated by request (ARQ) which raises the matters of memory in the transmission device, and of delay introduced by the retransmission and its protocol.

The network layer is responsible for establishing the route across the multitude of point-to-point connections within smaller networks but also across the world. In the world of connected computers, or productivity, this is called routing, while in telecommunications it is called circuit switching. In broadcast, data sinks and sources are rather stationary and the routes are still kind of hard wired by installation. The transport layer covers issues of quality of service (QoS) while restoring sequences, managing alternative routes and fees. The session layer might link more than two parties together. It organizes transmission jobs over time and quarantine services in case of absence. Data sharing is organized in larger segments and associated protocols silently ensure reasonable connectivity in times of temporary disconnections.

The presentation layer covers everything concerning the meaning or presentation of data, such as the form factor, data compression, encryption and keyboard mapping. In terms of audio, the form factor answers the question of audio channels involved and the type of data and stream (Tables 12.10 and 12.11). The application layer is not the application software but the entity that manages resources in a system, such as memory and I/O entities. It handles rights of access, data constancy, and operational tasks, such as remote jobs or data bank services.

Table 12.13 shows the layers and the well-established protocols associated with these layers. Clearly, digital workflows in modern audio productions demand professional and seamless systems at all levels.

12.7.5 Stereo Digital Audio Links

In 1985 the AES3 standard evolved on the basis of Sony's SDIF link. The family of two-channel standards

Table 12.13 OSI layered communication model and well-established protocols

	OSI layer	Established data networks protocols
7	Application	Simple Mail Transfer Protocol (SMTP), File Transfer Protocol (FTP), virtual terminal (TELNET), Network File System (NFS)
6	Presentation	
5	Session	
4	Transport	Transmission Control Protocol (TCP), User Data Protocol (UDP), Real-Time Protocol (RTP), Real-Time Transport Control Protocol (RTCP)
3	Network	Internet Protocol (IP)
2	Data link	Logical link control with medium access frames (MAC) for Ethernet, token-ring, token-bus etc.
1	Physical	Token-ring, token-bus, wireless LAN (WLAN), fiber-distributed digital interface (FDDI), Ethernet

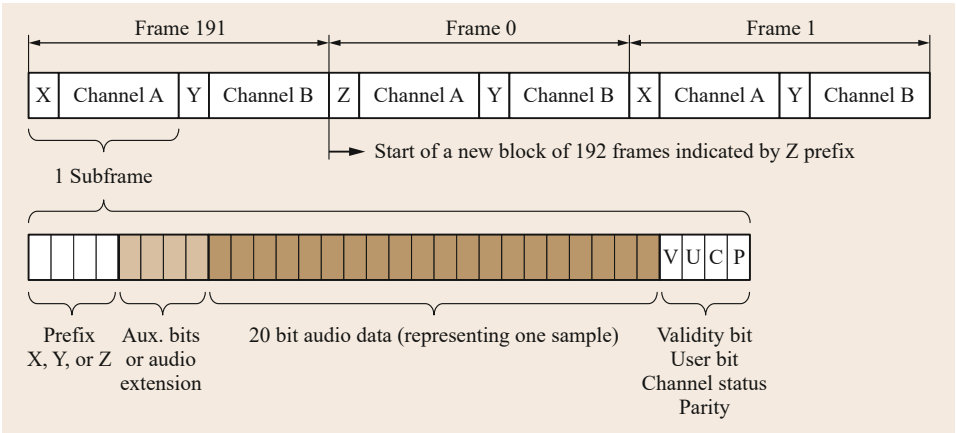


Fig. 12.26 AES3 data frame structure

based on AES3 is ANSI S4,40, EBU Tech. 325-E, ITU-R BS.64, BS 7239, DIN 60958 and IAJ CP-120 [12.15]. They have differential, symmetrical signal transmission at 2–7 V level on 110 Ω shielded cables with XLR connectors in common. One of the differences is that the EBU standard asks for galvanic isolation with the help of transformers. The physical layer therefore specifically defines levels for digital transmission while reusing existing cables and connectors. The data link layer is also specific. Blocks of 192 frames are transmitted in a specifically defined structure (Fig. 12.26). The data frame structure and data rate directly correspond to transmitting uncompressed native audio in linear PCM format. Data rate R and sampling rate f_s are therefore strictly related: $R = f_s\, C\, 32$, for the number of channels C. The use of auxiliary bits and interpretation of amended check bits vary from standard to standard.

Extensions and variations to this standard relate to the evolution of supplementary control signals and to migration into other applications. For instance, AES18 facilitates the user bit to carry radio data system (RDS) data or MIDI signals organized in frames of 192 bits under control of the widely applied HDLC protocol. AES3-id promoted 75 Ω coax cables and video-compatible signal levels to migrate into video applications. While IEC 60958 Type I (professional) is fully compatible with AES3, IEC 60958 Type II is the consumer version known as S/PDIF which differs in terms of electrical specifications and the use of amended check bits. This widely applied digital interface uses 0.5 V on 75 Ω cables and cinch connectors. TOSLink or S/PDIF optical work on plastic optical fibers (POF) at maximum distances of 10 m, while using the same data frame. AES3 family links are also used to transmit multichannel compressed audio instead of uncompressed stereo, following IEC 61937. SDIF-2 and SDIF-3 use separate links for two channels and for the system clock. These are not derivatives of AES3.

12.7.6 Multichannel Digital Audio Links

The multichannel audio digital interface (MADI) is specified in AES10 [12.16]. It reuses the AES3 subframe structure for individual channels but organizes 56 or 64 of these within each MADI frame to transport 64 audio channels at $f_s = 48$ kHz, or 32 audio channels at $f_s = 96$ kHz, with 16, 20, or 24 bit word size. The resulting data rate of $R = 125$ Mbps restricts the MADI coax to quite narrow electrical specifications to facilitate links of up to 50 m. MADI optical works on 50 µm glass fibers to bridge up to 2000 m.

There is a variety of proprietary digital serial and parallel links to carry multichannel audio. The ADAT lightpipe (ODI), Mitsubishi digital interface (PD, ProDigi), Roland R-bus, and Tascam digital interface TDIF-2 are merely specified to link multichannel recording equipment over short links.

Multichannel audio can also be embedded within video signals. The serial digital interface (SDI) and HD-SDI take up to 16 channels that are organized in the vertical blanking interval of the video signal while using the same AES3 subframe format.

12.7.7 High-Speed Digital General Purpose Links

Another class of links is distinguished by a high data rate, by support of various formats of video and audio, by connecting many systems and by synchronous transmission capability.

The IEEE Firewire interface connects up to 63 devices at data rates of 400, 800, 1600 or 3200 Mbps [12.17]. Links are physically limited to a few meters, suggesting usage within racks. Only low data rates would allow usage across a studio. With its synchronous transmission mode, Firewire can serve as a digital signal line for audio or video. While i.Link is

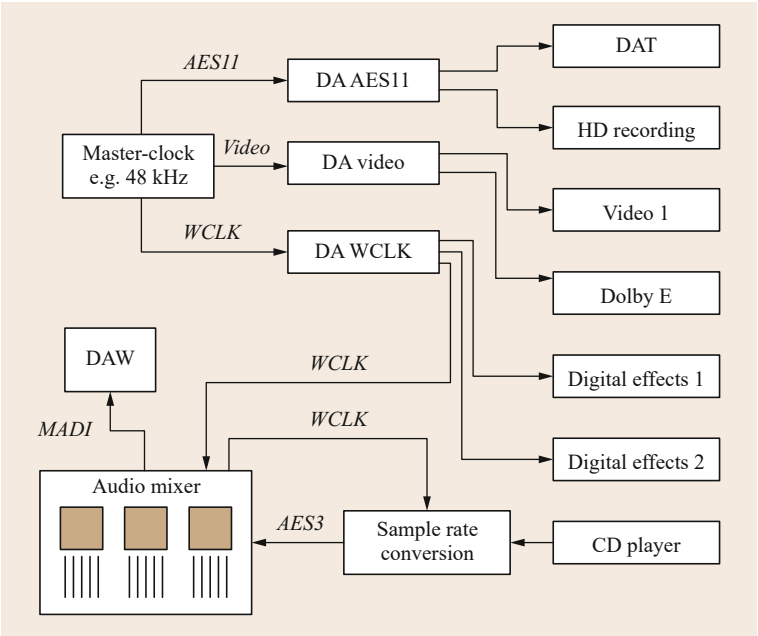


Fig. 12.27 Word clock distribution in a professional production environment with video components. Only clock signals are shown (no video or audio signals). Two concepts are combined: distribution by DA, and daisy chaining from device to device

Firewire plus Sony's video transport layer, mLAN is Firewire plus Yamaha's audio transport layer. The related audio and music transmission protocol supports 16 channels of 96 kHz audio at 24 bit word size, as well as compressed formats and metadata. Devices with mLAN connectivity require Firewire ports with an open host controller interface (OHCI).

The high-definition multimedia interface (HDMI) is a general purpose video and audio link that allows synchronous transmission between devices. It transports a wide range of compressed and uncompressed video and audio formats. In terms of audio links, it can host 8 channels at 192 kHz at 24-bit word size, or 8 channels SACD. HDMI can carry streams such as MPEG or other digital video broadcast (DVB) streams. It is preferably employed in presentation context rather than in production environments.

In summary, the AES3 family and the other families of audio links have similar data link layers that are combined with a variety of physical layers. From the OSI perspective, the data link layer and the presentation layer are unfavorably related with each other, since the data rate is directly derived from the sampling rate. Such strict relation between link and presentation is unusual in data networks. Furthermore, transport, routing and session layers are not specified for any of the mentioned AES3 family members. The transport layer is only marginally defined in an application-specific way for HDMI and Firewire. The OSI perspective makes it clear that these digital links are much more signal line replacements than general purpose data connections.

12.7.8 Synchronization

Mixing consoles, recording equipment, analog-to-digital (ADU) converters and digital effects must be synchronized to the same sampling frequency not only nominally, but also physically. With only the nominal accuracy of a quartz crystal in the range of ± 300 ppm, the systems would gradually drift apart and sampled words would get lost. Such coherent operation is mandatory in live performance, live broadcast and recording. In order to synchronize studio equipment, there are several types of clock sources and concepts. A professional clock source is a device that also generates a timecode and that may even be linked with a global time domain, such as DCF77 or GPS. In video productions, video black & burst is a preferred clock source that is derived from the video line frequency and image frequency.

Conceptually, if only a few components are to be synchronized, the ADU might be used for direct control of the mixing console and the recording device if the three components are not already integrated within one audio workbench (AWB). Professional productions have a wider scope in terms of functionality and workflow and require a synchronization strategy for each production. Figure 12.27 illustrates an example using both the concepts of word clock distribution amplifiers (DA) and word clock daisy chaining, while employing different formats at the same time, AES11, video format and an explicit word clock format.

12.7.9 Ethernet and IP-Based Links

Postproduction and playback do not require the coherent operation of systems, since the recording/production and the consumption are not committed to the same progression of time. Audio can be played back, listened to and processed without any problem if the presentation rate slightly deviates from the original sample rate of the recoding session. Second, parallel workflows encourage file sharing, rather than sharing of signals. Third, multisite productions and broadcast encourage the use of established data networks. These three reasons drive the use of IP-based links in audio production.

Employing general data network technology in audio implies both inheriting the mass market advantage but also the QoS limitations. Data packet networks work on the basis of probabilistic access, where the availability might be high but the latency times and loss rates are typically high as well. Capacity is not guaranteed in data networks. This is one of the reasons why some systems reuse parts of the general data network technology but still define a specific data link or transport layer to follow the concept of signal distribution with defined QoS. These systems take advantage of economies of scale while reusing Ethernet technology.

AudioRail reuses the physical layer of Ethernet and a specific link protocol to host 32 channels at 48 kHz or 16 channels at 92 kHz sampling rate, with 24-bit word size. It only carries native audio data but not any supplementary control data or MIDI.

Rocknet is another proprietary system that uses the Ethernet physical layer, however with XLR connectors. It supports up to 160 channels at 48 kHz and at 24 bits word size over a single CAT-5 cable. Additionally, the devices support redundant cabling and redundant power supply.

Cobranet reuses both, Ethernet's physical and data link layer. The proprietary protocol specified on top of this facilitates unicast and broadcast transmissions of up to 64 channels. Cobranet facilitates sample clock extraction from data. MIDI and control data can be transmitted together with the audio data, for instance for remote control tasks.

Ethersound also reuses an Ethernet physical layer and data frame. It is organized as a strict synchronous network with fixed capacities for 64 channels at 48 kHz or 16 channels at 192 kHz. Likewise, Ethersound has no potential of carrying general purpose data together with the audio data.

Dante differs from the three Ethernet-based systems in that it works on top of the network layer. It works on any IP based system and can be switched across wider networks. Audio traffic can be mixed with other traffic, but this must be operationally managed in terms of performance. Latency is in the sub-ms range and deterministic, because Dante uses voice over IP (VoIP) QoS, which is already provided in modern switches. WLAN is not recommended, and typically, IP will locally operate on top of Ethernet.

Ravenna, like Dante, works on top of the IP protocol. So while in most cases of local area networks Ethernet will be the option of choice, Ravenna is not so restricted. It can be established together with already existing IP networks in a building, and it even works across wide-area networks (WAN). Quality of service must be cared for by statistics in operation and maintenance. This is possible since today's routing IP devices support traffic shaping and traffic-specific control of QoS.

Ravenna directly supports AES3 subframes, but also any other format, such as a 32-bit floating point. It also provides synchronization across devices through the [12.18]. Precision Time Protocol (PTPv2) is a separate protocol devoted for real-time IP traffic. PTPv2 provides means for synchronizing local clocks to a precision in the lower nanoseconds range provided that all participating switches natively support PTPv2.

The European Broadcasting Union promotes a general approach of audio over IP (AoIP). The framework is defined by EBU Tech 3326 [12.19] and recommends

Table 12.14 ISO perspective on existing Ethernet- and IP-based audio links

	OSI layer	Audiorail Rocknet	Cobranet Ethersound	Dante	Ravenna	Audio over IP
7	Application					
6	Presentation					
5	Session					
4	Transport				RTP over UDP plus PTPv2	AoIP over RTP over UDP
3	Network			IP	IP	IP
2	Data link		Ethernet	Any LAN	Any LAN/WAN	Any LAN/WAN
1	Physical	Ethernet	Ethernet	Any LAN	Any LAN/WAN	Any LAN/WAN

Table 12.15 Descriptions of analog and digital links for a typical mixing console (Fig. 12.28)

A	Analog symmetric audio input, XLR-3-F
B	Analog symmetric audio output, XLR-3-M
C	AES3 digital audio I/O on 25-Sub-D
D	TDIF-1 digital audio I/O on 25-Sub-D
E	Timecode input on XLR-3-F
F	MIDI out on 5-pin DIN connector
G	MIDI in on 5-pin DIN connector
H	USB link to computer
I	Word clock (WCLK) output on BNC
J	Word clock (WCLK) input on BNC
K	AES3 Digital audio output on XLR-3-M
L	IEC 60958 Type II output on cinch
M	IEC 60958 Type II input on cinch
N	AES3 Digital audio input on XLR-3-F

Fig. 12.28 Rear side of a digital mixing console and its typical connectivity (© Yamaha, for descriptions see Table 12.15)

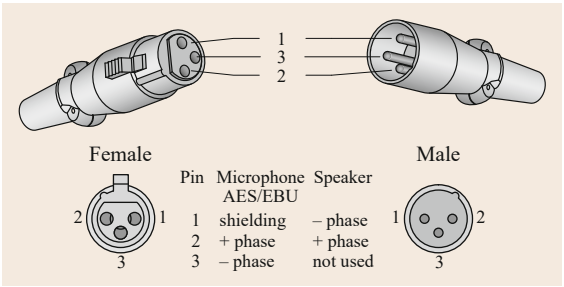


Fig. 12.29 Male and female XLR-3 connectors and pinning for symmetrical microphone, speaker and digital AES/EBU links

jitter buffering, FEC and error concealment to compensate for the unpredictable QoS in unmanaged IP networks. For a tutorial see EBU Tech 3329 [12.20].

The OSI perspective finally allows one to distinguish between classes of systems. Table 12.14 gives an overview of the structured layers. Without specifications for layers 5 to 7, all of these systems come without session management, coding for presentation or format, access rights management, digital rights management (DRM) or data consistency management. Future systems will incorporate these aspects to facilitate seamless workflows between live performance/recording, postproduction and distribution/broadcast.

12.7.10 Connectors

Connectors form part of the physical layer, since they codetermine the mounted type and amount does of wires or fibers. The identified connector not always directly determine the type of link associated with it, because widely spread connectors have been reused for other links. For instance XLR, the classical microphone connector, is also used for the digital AES/EBU link (Fig. 12.28 for an overview).

Figure 12.29 sketches the popular 3-pin XLR (XLR-3) connector and its pinning. The XLR-5 connectors carry stereo signals (Canon's X-series latched connectors (XL), with resilient connector filling (XLR)).

12.8 Loudspeakers, Reference Listening and Reinforcement

The fidelity of sound reproduction is defined by the quality of speakers, while the quality of the spaciousness and envelopment is defined by the speaker arrangement within a given room and the match of this arrangement with the underlying principle of the microphone arrangement used.

12.8.1 Loudspeakers

Similar to microphones, there are various principles of conversion. First, classical electrodynamics is used in speakers to convert electrical signals into mechanical displacement, or to sound. A current-carrying coil within a magnetic field $\boldsymbol{B}$ will experience the Lorentz force

$$\boldsymbol{F} = I\,(\boldsymbol{l} \times \boldsymbol{B}) \qquad (12.13)$$

given the length $\boldsymbol{l}$ of the coil-wire. In speakers, the coil is conducted in a gap of a strong magnetic field such that the effect is maximized and that the coil experiences a degree of freedom in the direction of force. This force will cause displacement of the coil, which in return drives a membrane. Large displacements are possible, which is of particular importance for low frequencies. However, large displacements also constitute

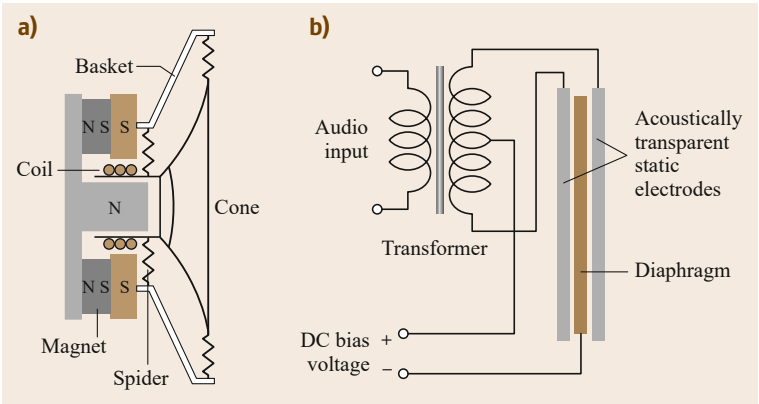


Fig. 12.30a,b Operational principle of **(a)** electrodynamic and **(b)** electrostatic speakers

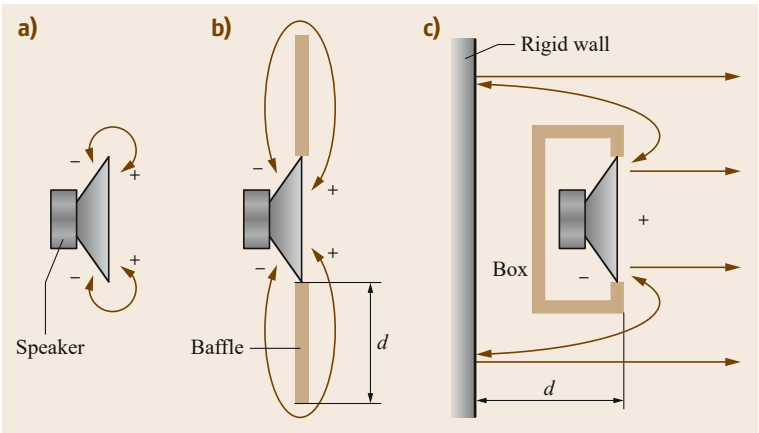


Fig. 12.31 **(a)** Acoustically short-circuited loudspeaker, **(b)** loudspeaker in a baffle, **(c)** loudspeaker in a closed box in front of a rigid wall

inhomogeneous conversion since the coil might leave the magnetic field, at least partially. This is one of the reasons for nonlinearity in speakers. Second, ribbon speakers use the same electrodynamic principle but the current directly flows through a plane membrane. Third, electrostatic speakers employ the force that is in effect on an electrical charge Q

$$|\boldsymbol{F}| = \frac{QU}{d} \qquad (12.14)$$

with the voltage U and the physical distance d between electrodes. Flipping the polarity of the charge will not flip the force direction (see the magnitude for force in (12.14)). Therefore, for bipolar audio signals, a high DC voltage is used to bias the system (Fig. 12.30). The electrostatic principle will typically facilitate only small displacements of the membrane or diaphragm.

Another conversion principle uses the piezoelectric effect, however reversely operated when compared with the piezomicrophone (Sect. 12.1). The voltage across the crystalline structure will cause deformation and therefore displacement. The displacement is very small and therefore these converters are used for high-frequency applications or for buzzing in low-end applications. Other principles use ionized air and again others use structure-borne sound across large plane panels, so-called distributed-mode loudspeakers (DML).

All the conversion principles have in common that the typical arrangement implies only one degree of freedom and therefore the speakers typically radiate a figure of eight, potentially short-circuiting acoustically (Fig. 12.31a). The air that is pushed away at the front side of the speaker is simply pulled in at the back side and reversed. The effect of short-circuiting is frequency-dependent and can be prevented by using an extended baffle (Fig. 12.31b). Now the pressure wave has to travel around the baffle to be compensated by a matching sink. Travel time is twice the distance d at the speed of sound $c \approx 340\,\mathrm{m/s}$. The critical frequency below which the radiation suffers by 6 dB/octave is

$$f_0 = \frac{c}{4d}\,. \qquad (12.15)$$

A closed box has the effect of an infinitely extended baffle, since the direct acoustical short-circuit is prevented (Fig. 12.31c). However, loudspeakers are

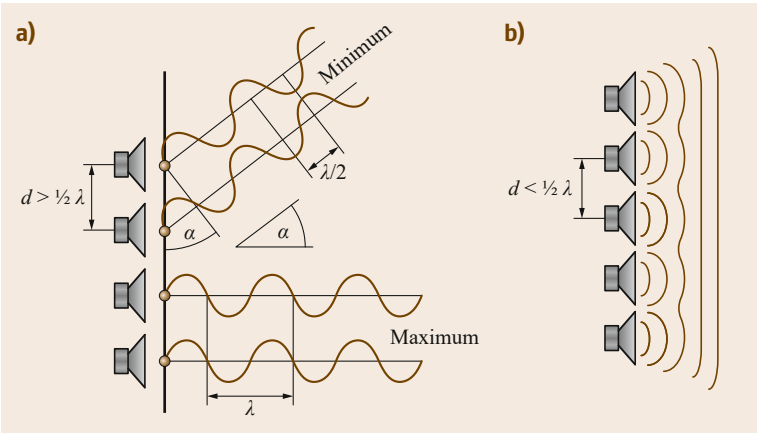


Fig. 12.32a,b Soundfields from lines of loudspeakers separated by distance d with **(a)** interference ($d > \lambda/2$) and **(b)** plane wavefronts ($d < \lambda/2$)

usually placed within buildings and in many cases in front of walls. The travel time of a pressure wave to the wall behind and back to the plane of the speaker cone is twice the distance d at speed c and defines frequencies of compensation.

$$f_n = \frac{c\,(1+2n)}{4d}\,, \quad n = 1, 2, 3, \ldots \tag{12.16}$$

At these frequencies, the radiated power is much lower, at least for listeners on axis. When listening off-axis, the geometry constitutes derivatives of the simple form in (12.16).

In most electrodynamic speakers, the membrane is a cone, further defining nonlinearity but also the radiation beam. The membrane is not completely stiff and the cone geometry will increase stiffness only to a certain degree. An impulse generated at the center will need time to travel across the membrane. The vibrational modes of the membrane therefore codetermine the general frequency response of a speaker. The radiation pattern of speakers constitutes narrow beams for high frequencies and wide beams for low frequencies. Sound pressure measurements are usually only done on axis and are no reference to the spaciously integrated total radiation power.

The electrodynamic principle is also used in horns, which are usually designed to radiate sound of higher frequencies. The horn geometry is typically designed such that the cross sectional area grows exponentially along the horn in order to gradually adapt the driver impedance to air impedance.

If several speakers are driven by the same signal, the geometrical arrangement again constitutes cases of attenuation and amplification, of interference and plain wavefronts. In cases where the distance d between speakers is more than half of the wavelength λ the superposition of pressure waves gets a destructive component. The relation of d to λ defines the direction α with respect to the normal axis in which individual waves have a phase shift of $\lambda/2$, or π, effectively compensating for each other (see the minimum in Fig. 12.32a). This is true also for phase shifts of $3/2\lambda$, $5/2\lambda$ and so on. The general relation for the off-normal-axis angle reads,

$$\alpha_n = \arcsin\left(\frac{(1+2n)\,\lambda}{2d}\right) d > \frac{\lambda}{2}\,; \quad n = 1, 2, 3, \ldots \tag{12.17}$$

In cases of d smaller than half of the wavelength λ, (12.16) cannot be solved and circular wavefronts of the elementary sound sources jointly establish a plane wavefront according to Huygens' principle (Fig. 12.32b). For a given loudspeaker arrangement with a distance d between loudspeakers, sound components below a critical frequency will be radiated in the form of a plane wave.

$$f_{\text{crit}} = \frac{c}{\lambda_{\text{crit}}} = \frac{c}{2d}\,, \quad d = \frac{\lambda_{\text{crit}}}{2} \tag{12.18}$$

At the same time, sound components above the critical frequency will suffer from interference. The principle of establishing plane wavefronts is followed in line arrays (LA) and in wave field synthesis (WFS). But the principle also applies to two-way or three-way speakers, where speakers are distant to each other but the spectral range they serve is not separated without a crossover. Certainly, any listening room raises further and even more complex scenarios of interference due to multiple reflections.

12.8.2 Reference Listening

Reference listening addresses several issues beyond the sound quality of the speakers. How well does the

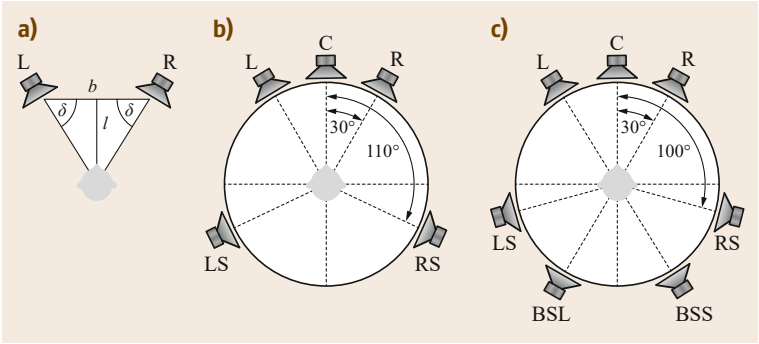


Fig. 12.33a–c Recommended loudspeaker arrangements according to (a) DIN 15996 and (b,c) ITU-R BS.775-1

speaker arrangement match the underlying principle of microphone arrangement? How well does the generated soundfield represent the fictional soundfield that was intended by an audio designer? While for stereo and DTS there are reference listening arrangements and standards, for many of the three-dimensional and application-specific formats standards currently evolve.

12.8.3 Two-Dimensional Loudspeaker Arrangements

According to DIN 15996, stereo playback is recommended with the two speakers separated by $b = 3{-}4.5\,\mathrm{m}$ and the listener position at a distance of $l = (0.9{\pm}0.3)b$. The ideal arrangement is equilateral, however there is some tolerance defined, $\delta = 60° \pm 15°$ (Fig. 12.33a). Reference arrangements for 5.1 and 7.1 according to ITU-R BS.775-1 incorporate the original stereo triangle and all speakers are on a circle (Fig. 12.33b and 12.33c). Dolby Digital is the consumer market format to capture stereo and multichannel audio with and without the low-frequency effects channel (LFE), respectively 5.0 and 5.1 (Table 12.10).

Another reference for stereo is ITU-R BS.1116-1. This standard also specifies listening conditions beyond the loudspeaker geometrical setup, such as operational ranges for loudspeakers and for room-acoustical parameters. In terms of sound quality assessments there are specific recommendations, such as ITU-R BS.1284-1 [12.21] and handbooks on general methodology for sound quality evaluation [12.22]. More specifically, the EBU Tech. 3286 [12.23] specifies parameters for the quality assessment of stereo sound. Among these are spatial impression (subparameters: homogeneity of spatial sound, reverberance, acoustical balance, apparent room size, depth perspective), stereo impression (subparameters: directional balance, stability, sound image width, location accuracy), transparency, sound balance, timbre and more. So there is nuanced expert knowledge shared by audio engineers and equipment OEMs if it comes to stereo.

Wave field synthesis (WFS) employs the outlined planar wave propagation principle to control a two-dimensional soundfield. Loudspeakers are densely spaced, for instance at $d = 17\,\mathrm{cm}$ to achieve plane waves for frequencies below $f_{\mathrm{crit}} = 1000\,\mathrm{Hz}$. A listening arrangement can easily ask for several hundred speakers. With WFS, a pressure wave front can be formed to simulate source positions that lay behind the row of loudspeakers (see source 1 in Fig. 12.34). This is achieved by appropriately delayed loudspeaker signals. The operational technology even allows to model sound sources that are in front of the line (see source 2 in Fig. 12.34). Likewise, the acoustical room and its real reflections can be modelled as well by treating reflections as additional sound sources.

WFS modelling is a complex task even for stationary sources and listener positions, since the directivity of the speakers codefine the synthesis operators [12.24] and directional room-impulse responses must be generated for each pair of listening position and sound source, which can be done by parameterized recordings from circular microphone arrays [12.25].

The two-dimensional sound representation in WFS and the modelled source distance are convincingly perceived when high-fidelity impulse responses are used. Even more complex is the model for sources in motion.

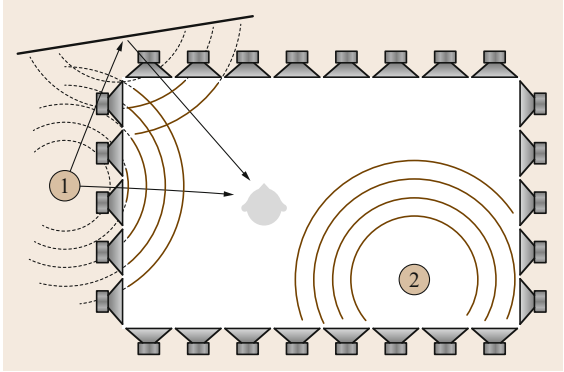


Fig. 12.34 Wave field synthesis

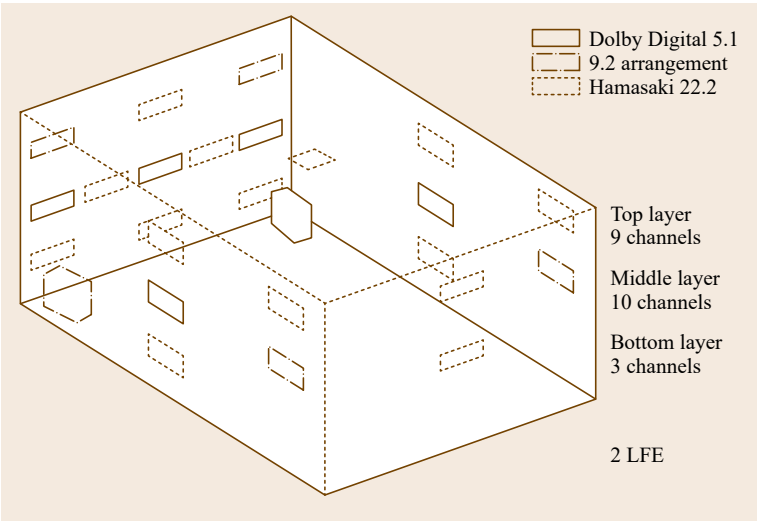


Fig. 12.35 22.2 Loudspeaker arrangement according to Hamasaki

Moving sound sources are so impressive nowadays that only little effort has been made to achieve high-fidelity input parameters. WFS today is often used for fictional sound and audio drama, while relatively large PA speakers are spaced at $d = 1$ m. The resulting $f_{crit} = 170$ Hz does not really facilitate plane waves for speech and music. WFS therefore is more commonly operated as a multichannel surround system for applications without preferred listening direction.

12.8.4 Three-Dimensional Loudspeaker Arrangements

There are various loudspeaker arrangements that consider elevation in sound reproduction. The intentions differ from application to application. In music recording, the envelopment is strongly enhanced by reproducing ceiling reflections, while in cinema, discrete sound sources fictionally move above, but also possibly below the audience.

Auro 3D addresses elevation in two versions. The Auro 3D 9.1 proposal targets the home theatre and is identical with the 5.1 arrangement (Fig. 12.33b) for the base level and it uses four more speakers for the upper level, one above each of the side channels. When standard formats such as 4.0, 5.1, 6.1, or 22.2 are used in cinemas, the arrangement of speakers is adapted to the much larger listening area. In particular, the side channels LS and RS are distributed over multiple loudspeakers along the side, and even from the rear. In the Auro 3D 11.1 for cinema applications, this is also the case, and the side is even extended towards the back side. The upper level has the same format as the lower level. In addition, there are speakers in the ceiling (voice of God). Auro 3D is a complete system with its own theory to support direct upmixing of stereo or 5.1 to the 9 or 11 spatial channels. The conversion engine creates multiple decorrelated channels from the few inputs. It can also convert to three-dimensional (3-D) headphone representations.

The Hamasaki 22.2 arrangement [12.26] was developed in the context of high-definition television (HD-TV) and is standardized [12.27]. It facilitates elevation, both above and below the audience, but also an improved resolution in the front (Fig. 12.35).

The 5.1 multichannel file and streaming format has also been used to accommodate elevation in the form of 2+2+2 channels. *Dabringhaus* promoted this format of 2 front, 2 rear, and 2 elevated front channels, working for a maximum of room ambience and auditor envelopment with a minimum number of channels [12.28]. This format reuses the 5.1 technology both in production and in consumer electronics. It is not widely used but the concept is convincing, because the multichannel recording can directly be projected to the loudspeaker arrangement at a nuanced expert level. Sound engineers familiar with stereo recording can directly project the outcome of their microphone arrangement. Listening to 2+2+2 is a convincing experience and there is a selection of fine recordings of classical music.

Ambisonics is the keyword of 3-D spatial representation. The loudspeaker arrangement complements the 3-D coincident microphone arrangement, or A-format and B-format, as outlined in Sect. 12.1.3. Coincident loudspeakers are not possible and near-coincident speakers are typically approximated by dodecahedrons. Ambisonics arrangements are typically equally spaced on a spherical grid. The soundfield must be encoded for the specific number of channels on the grid. Technologies in this field facilitate conversion of standard

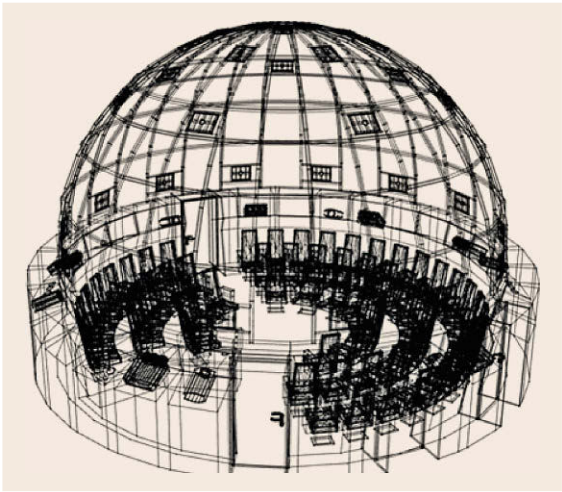


Fig. 12.36 Planetarium in Kiel, Germany – an example of individual 3-D auralization

formats to ambisonics, but also the creation of virtual acoustical spaces.

There is also a wide range of application-specific solutions for exhibitions, fairs, museums and shows. For instance, in the Kiel planetarium, the position of speakers is codetermined by building construction. There are three levels, with 14 speakers on the ground level and 7 speakers on each of the upper layers (Fig. 12.36). Input to the system is a matrix of sound sources and time-dependent vectors. Standard 5.1 and stereo panning technologies have been combined to seamlessly move individual sound sources across the sky.

Headphones are also widely used for reference listening. The advantage is that the soundfield is free of disturbing reflections from walls, or more generally speaking, there is no acoustical interference from the listening room that would degrade the room acoustics contained in the presented sound. However, there are also problems and disadvantages. First, for stereo, the geometry strongly differs from the intended 60° arrangement. While in reference stereo both channels reach both ears, under headphone conditions the channels and ears are strictly separated. The stereo image is much wider and strongly deviates from the intended scene. This does not mean that such wide stereo is disliked. Second, the headphone is in many cases equivalent to a pressure chamber, introducing additional rules for equalization. Third, for many samples and for many people, the so-called in-head localization is very annoying. The sound source might well be spatially discriminated, but it is inside rather than outside the head. Interaural time difference (ITD) and interaural intensity difference (IID) which are usually the main factors for spatial discrimination do not necessarily translate in a standardized form for every shape of the head. Microphone arrangements that imply an equivalent ITD have an advantage for headphone listening (e.g., ORTF in Fig. 12.7), even though they are problematic in mono downmixes. Reflections from the torso are likewise important to translate the soundfield into a real-world experience, but these reflections are missing in headphone listening, too.

Dummy head recordings simulate the real-world listening scenario since ITD and IID apply, but also the shadow of the head and the form of the pinna and concha. But even with the dummy head, in-head localization remains a problem for many people, because dummy head parameters only approximate the average listener. Based on anechoic dummy head recordings, there also exist direction-specific head-related transfer functions (HRTF). *Algazi* et al. provided an extended reference library of spatial high-resolution HRTFs [12.29]. There is also a general model derived from this data [12.30]. However, again it is only by chance that the perceived direction of a sound source will match the true direction of the dummy head in the original recording setup. Given these restrictions, there are technologies that translate 5.1 formats to the two-channel format for headphone listening. Some of these even allow the head to be moved, while the virtually created reference 5.1 monitoring remains statically fixed with respect to the room.

Perceived elevation in binaural listening is strongly determined by the pinna and cochna shape. Reference

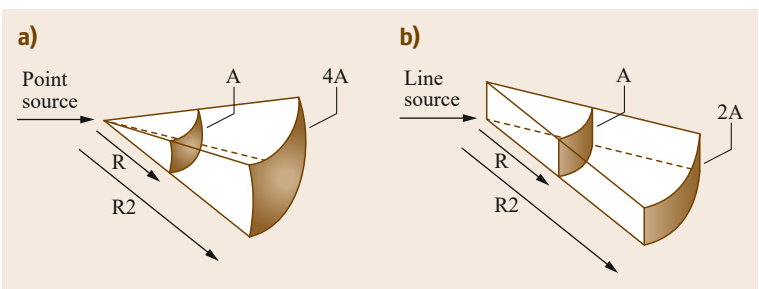


Fig. 12.37a,b Spherical and cylindrical sound radiation

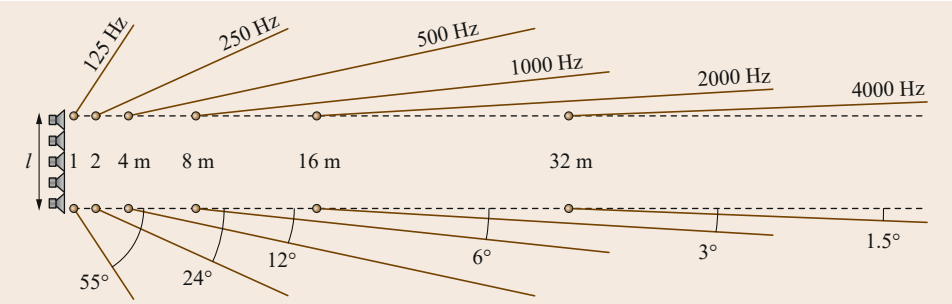


Fig. 12.38 Frequency-dependent transition between near-field and far-field for a $l = 2.3$ m long line array

HRTF recordings inherently incorporate this and this know-how is used to encode 3-D audio in such a way that it can be represented on the two channels for headphone listening. The pinna and cochna shape can also be used to directly model elevation in individual HRTFs [12.31].

12.8.5 Reinforcement

Line arrays of speakers are another example of establishing plane wavefronts and are widely applied in reinforcement systems or PA systems. Solitary speakers radiate spherically, with a loss of −6 dB sound pressure level (SPL) for each doubling of distance R, beginning with the frequency-dependent far-field. Line arrays, as they approximate ideal line sources, radiate cylindrically with a loss of only −3 dB SPL (Fig. 12.37 for a comparison of geometry).

The transition from cylindrical to spherical radiation, or from near-field to far-field, is given at the distance

$$r_{\text{transition}} = \frac{3fl^2}{2}\sqrt{1-\frac{1}{(3fl)^2}} \qquad (12.19)$$

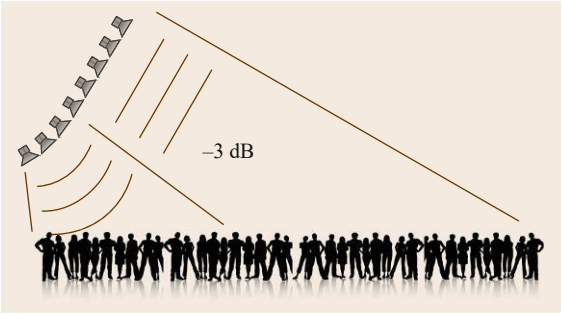


Fig. 12.39 J-curved line array with the straight section creating cylindrical waves to reach far listeners

for line arrays of limited length l and frequency f in kHz [12.32]. Figure 12.38 illustrates this frequency dependency for a 2.3 m long line array.

The reduced loss of SLP for line sources can also be expressed by a focused beam. This advantage is used in PA systems where a wide auditory space has to be covered at possibly comparable SPL for listeners in the front and in the rear. Figure 12.39 illustrates a J-curved line array where far listeners are reached by cylindrical waves, while nearby listeners are not.

References

12.1 M. Williams: *Microphone Arrays for Stereo and Multichannel Sound Recording*, Vol. 1 (Editrice Il Rostro, Segrate 2004)

12.2 U. Herrmann, V. Henkels, D. Braun: Comparison of 5 surround microphone methods. In: *20. Tonmeistertagung* (Bildungswerk des Verbandes Deutscher Tonmeister, Verlag K. G. Saur, München 1999) pp. 508–517

12.3 M. Williams, G. Le Dû: Microphone array analysis for multichannel sound recording. In: *107th AES Convention Preprint 4997* (1999)

12.4 G. Theile: Natural 5.1 music recording based on psychoacoustic principals. In: *AES 19th Int. Conf.: Surround Sound – Techniques, Technol. Percept., Elmenau* (2001)

12.5 M.A. Gerzon: The design of precisely coincident microphone arrays for stereo and surround sound. In: *50th Convention of the Audio Engineering Society* (Mathematical Institute, University of Oxford, Oxford 1975), pp. Preprint L–20

12.6 M. Vorländer: *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality* (Springer, Berlin 2008) pp. 1–335

12.7 R. Bader: Reconstruction of radiating soundfields using minimum energy method, J. Acoust. Soc. Am. **127**, 300 (2010)

12.8 M.R. Schroeder: Natural sounding artificial reverberation, J. Audio Eng. Soc. **10**(3), 219–223 (1962)

12.9 W.G. Gardner: Efficient convolution without input-output delay, J. Audio Eng. Soc. **43**(3), 127–136 (1995)

12.10 G. de Poli (Ed.): *Representations of Musical Signals. Proc. Int. Workshop, Sorrento* 1991) pp. 1–478

12.11 CITT G.726, Recommendation: *General Aspects of Digital Transmission Systems; Terminal Equipments 40, 32, 24, 16 kbit/s adaptive differential pulse code modulation (ADPCM)* (The International Telegraph and Telephone Consultative Committee, Geneva 1990)

12.12 DIN EN ISO 226:2006-04 (E): Acoustics – Normal Equal-Loudness-Level Contours (Beuth, Berlin 2006)

12.13 EBU – TECH 3306, Technical Specification: *MBWF/RF64:An Extended File Format for Audi* (European Broadcasting Union, Geneva 2009)

12.14 ISO/IEC 11801, International standard: *Information Technology – Generic Cabling for Customer Premises*, 2nd edn. (International Organization for Standardization, Geneva 2002)

12.15 AES3, AES standard for digital audio engineering: *Serial transmission format for two-channel linearly represented digital audio data* (Audio Engineering Society, New York 2009)

12.16 AES10-2008 (r2014): *AES Recommended Practice for Digital Audio Engineering – Serial Multichannel Audio Digital Interface (MADI)* (Audio Engineering Society, New York 2008

12.17 IEEE 1394, IEEE Sandard: *IEEE Standard for a High-Performance Serial Bus* (Institute of Electrical and Electronics Engineers, Piscataway 2008)

12.18 IEEE 1588, IEEE Standard: *IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems* (Institute of Electrical and Electronics Engineers, Piscataway 2008)

12.19 EBU – TECH 3326, Technical Specification: *Audio Contribution over IP*, Rev. 4 (European Broadcasting Union, Geneva 2014)

12.20 EBU – TECH 3329, Tutorial: *A Tutorial on Audio Contribution over IP* (European Broadcasting Union, Geneva 2008)

12.21 ITU-R BS.1284-1, Recommendation: *General Methods for the Subjective Assessment of Sound Quality* (International Telecommunications Union—Radiocommunication, Geneva 2003)

12.22 S. Bech, N. Zacharov: *Perceptual Audio Evaluation – Theory, Method and Application* (Wiley, Chichester 2006)

12.23 EBU document Tech. 3286: *Assessment Methods for the Subjective Evaluation of the Quality of Sound Programme Material – Music* (European Broadcasting Union, Geneva 1997)

12.24 D. de Vries: Sound reinforcement by wavefield synthesis: adaptation of the synthesis operator to the loudspeaker directivity characteristics, J. Audio Eng. Soc. **44**, 1120–1131 (1996)

12.25 E.M. Hulsebos, D. de Vries: Parameterization and reproduction of concert hall acoustics measured with a circular microphone array. In: *AES Convention, Munich* (2002), Paper 5579

12.26 K. Hamasaki, T. Nishiguchi, R. Okumura, Y. Nakayama, A. Ando: A 22.2 multichannel sound system for ultra-high-definition TV (UHDTV), SMPTE Motion Imaging J. **117**(3), 40–49 (2008)

12.27 ST 2036-2, Standard: *Ultra High Definition Television – Audio Characteristics and Audio Channel Mapping for Program Production*. (Soc. of Motion Picture and Television Eng., White Plains, New York 2008)

12.28 W. Dabringhaus: 2+2+2 – kompatible Nutzung des 5.1 Übertragungsweges für ein System dreidimensionaler Klangwiedergabe klassischer Musik mit drei stereophonen Kanälen, (The 5.1 reproduction chain gives us the chance to be used as a true threedimensional sound-reproduction system for classical music with three pairs of loudspeakrers). In: *21. Tonmeistertagung* (MM-Musik-Media-Verlag, Hannover 2000)

12.29 V.R. Algazi, R.O. Duda, D.M. Thompson, C. Avendano: The CIPIC HRTF Database. In: *Proc. 2001 IEEE Workshop Appl. Signal Process. Audio Electroacoust* (Mohonk Mountain House, New Paltz 2001) pp. 99–102

12.30 V.R. Algazi, R.O. Duda, R. Duraiswami, N.A. Gumerov, Z. Tang: Approximating the head-related transfer function using simple geometric models of the head and torso, J. Acoust. Soc. Am. **112**, 2053 (2002)

12.31 V.C. Raykar, R. Durais, B. Yegnanarayana: Extracting the frequencies of the pinna spectral notches in measured head related impulse responses, J. Acoust. Soc. Am. **118**, 364 (2005)

12.32 M. Urban, C. Heil, P. Bauman: Wavefront sculpture technology, J. Audio Eng. Soc. **51**(10), 912–932 (2003)

13. Delay-Lines and Digital Waveguides

Gary Scavone

A digital delay line is a particular type of finite impulse response (FIR) filter that has many useful applications in audio signal processing. Simply put, signals that are input to a delay line reappear at the output after a specified time period (in samples). Delay lines are often implemented to support delay times that can vary dynamically. As well, delay times corresponding to noninteger sample lengths can be approximated.

Time delay of signals is fundamental to signal processing systems. In this chapter, we focus on applications in digital audio signal processing and in particular, the modeling of wave propagation in air and in strings. The fundamentals of delay lines will be introduced and their implementation detailed, including common fractional-delay filtering techniques. Feedforward and feedback comb filters are simple signal processing structures built with delay lines and they exhibit characteristics that not only make them interesting for delay-based audio effects algorithms, but also as simple models of acoustic wave propagation. Finally, the use of delay lines to simulate wave propagation in one-dimensional waveguides will be introduced with a focus on the synthesis of plucked string instrument sounds.

13.1 Digital Delay Lines

For an input signal given by the time series $x[n]$ with sample indices $n = 0, 1, 2, \ldots$ and a delay-line length of M samples, the delay-line output is defined by the difference equation

$$y[n] = x[n - M], \quad n = 0, 1, 2, \ldots ,$$

where $x[n]$ is equal to zero for sampled time step indices of $n < 0$.

The functioning of a delay line can be visualized as in Fig. 13.1, assuming $M = 4$ and a discrete-time unit impulse input signal defined as

$$\delta[n] = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \, . \end{cases}$$

For every time step n, the signal value in the last (rightmost) memory location is output from the delay line, the remaining stored values are propagated to adjacent memory locations (to the right), and a new input sample

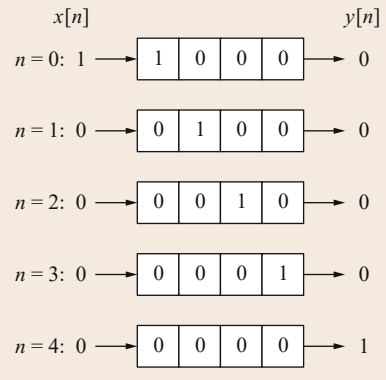


Fig. 13.1 Signal flow in a digital delay line of length $M = 4$ samples

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_13

is written to the first (left-most) memory location. Thus, the sample value of 1 that is input at time $n = 0$ will appear at the delay-line output at time step $n = 4$. For this particular example, all subsequent inputs and outputs to the delay line are zeroes. In this way, the finite length impulse response of the length $M = 4$ digital delay line is given by $h = \{0, 0, 0, 0, 1\}$.

It is straightforward to apply standard filter analysis methods to investigate the behavior of digital delay lines in the frequency domain. For example, application of the z-transform to the difference equation above results in a transfer function for this system of

$$H(z) = Y(z)/X(z) = z^{-M} .$$

The frequency magnitude response, determined by replacing z with $e^{j\omega}$ and evaluating the magnitude for $0 \le \omega \le \pi$ (corresponding to frequencies from 0 to $f_s/2$ Hz at the sample rate f_s, where $\omega = 2\pi f/f_s$ is the normalized discrete-time radian frequency), is equal to one across all frequencies. This should not be surprising considering that the delay line does not change the input values in any way aside from delaying them in time. The phase response for the example $M = 4$ delay-line system is shown in Fig. 13.2, plotted both in radians (a) and as *phase delay* in samples (b), from which its linear phase response is clear.

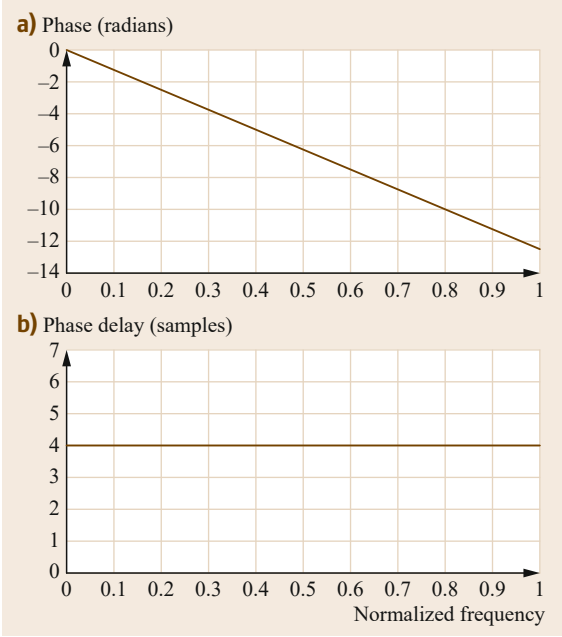


Fig. 13.2a,b Phase response of a digital delay line of length $M = 4$ samples: **(a)** phase in radians; **(b)** phase delay in samples

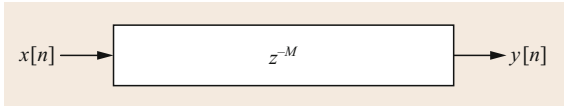


Fig. 13.3 The block diagram representation of a digital delay line of length M samples

In digital filter block diagrams, the digital delay line is typically represented as shown in Fig. 13.3, where the length of the delay is given by the exponent of z.

13.1.1 Delay Line Implementation

The delay-line visualization of Fig. 13.1 implies a significant computation burden associated with its implementation. That is, from the description above it might seem that every signal value stored in the delay line must be copied over to an adjacent memory location for every new input/output value. However, digital delay lines can be very efficiently implemented in computer programming environments, requiring only sufficient memory and one or two memory pointers. For example, the fixed-length $M = 4$ digital delay line can be implemented with just four samples of memory storage and a single pointer as demonstrated in the following Matlab example:

```
N = 60; % time steps to compute
x = [1, zeros(1, N-1)]; % unit impulse
M = 4; % delay line length
delayline = zeros(1, 4);
ptr = 1;
y = zeros(1, N); % output signal

for n = 1:N,
  y(n) = delayline(ptr); % read output
  delayline(ptr) = x(n); % write input
  ptr = ptr + 1;  % increment pointer
  if ptr > M,     % check pointer limit
    ptr = 1;
  end
end
```

The approach demonstrated above also supports dynamically variable delay-line lengths. The example will work for any value of M in the range $1 \le M \le 4$. It is only necessary to allocate enough memory to support the maximum desired delay-line length. The example above does not support a delay-line length $M = 0$. For that to work, it is necessary to use and increment two pointers, one each for the input and output memory locations. Memory allocation can be a relatively time consuming operation in a real-time synthesis environment. Thus, in situations where the delay length may change over time, a large buffer of some maximum size

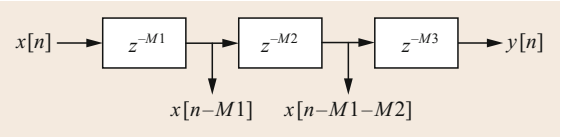


Fig. 13.4 The block diagram representation of a tapped digital delay line with two internal output taps

is usually created during program initialization (for example, the fourth line of the Matlab example could be changed to `delayline = zeros(1, 50)` to support values of M in the range $1 \leq M \leq 50$).

13.1.2 Tapped Delay Lines

There are often contexts where it is desirable to access the contents of a delay line at one or more intermediate delay length values. Such a system is referred to as a tapped delay line. The block diagram of Fig. 13.4 illustrates a tapped delay line of total length $M1 + M2 + M3$ samples and two taps located at delay lengths of $M1$ and $M1 + M2$ samples. In general, a tap delay line is implemented as a single buffer in memory but with as many *read* pointers as taps. Such functionality could be added to the Matlab example above using a new pointer for each intermediate tap length.

13.1.3 Delay-Line Interpolation

The delay lines discussed thus far have had lengths given by an integer number of samples. In many contexts, especially when the delay length can change dynamically, it is necessary to compute the output of a delay line for lengths that correspond to some fraction of a sample. For example, if a delay-line length was to be slowly increased from 30 to 32 samples over a time of one second, it would be necessary to perform this increase at small increments (perhaps 0.1 sample increments or less) in order that the change be smooth and free of audible artifacts.

While signal values in a delay line are stored at integer multiples of the sample period $T_s = 1/f_s$, it is possible to apply interpolation techniques to approximate values between those in memory. A large body of literature exists on such interpolation techniques [13.1]. The most common interpolation techniques used in conjunction with delay lines are first-order linear interpolation and allpass interpolation.

Linear Interpolation

Linear interpolation is an efficient technique for determining sample values at fractional delay-line lengths. Intermediate values between two neighboring samples are found by effectively *drawing* a straight line connecting those two samples and returning values at desired positions along that line.

Let Δ be a number between 0 and 1 that indicates how far to interpolate a signal y between time steps n and $n-1$. The linearly interpolated value $y[n-\Delta]$ between known values $y[n]$ and $y[n-1]$ is given by

$$\begin{aligned} y[n-\Delta] &= (1-\Delta)\cdot y[n] + \Delta \cdot y[n-1] \\ &= y[n] + \Delta \cdot (y[n-1] - y[n]) \,. \end{aligned}$$

This operation is equivalent to a first-order FIR filter. The use of linear interpolation at the output of a delay line is diagrammed in Fig. 13.5. The frequency magnitude response and phase delay for a linear interpolation filter is shown in Fig. 13.6 for fractional delays between 0 and 1. The *ideal* interpolation filter would have a constant magnitude of one and a constant phase delay exactly equal to Δ across all frequencies. In other words, the filter would time-shift its input by the desired amount without modifying the gain, irrespective of the signal frequency content. From Fig. 13.6, it is apparent that linear interpolation becomes less accurate at higher frequencies, though this nonideal behavior varies with Δ. As an example, when $\Delta = 0.5$ the phase delay is exact across all frequencies but higher frequency components are attenuated (their magnitude response falls well below 0 dB).

Linear interpolation is also known as first-order Lagrange interpolation. Higher order Lagrange interpolators can be more accurate but are less efficient to implement.

Allpass Interpolation

Allpass filters have a unity magnitude response but variable phase delay properties. This makes them potentially useful for fractional-delay filtering, as they are guaranteed to pass signals without attenuating their

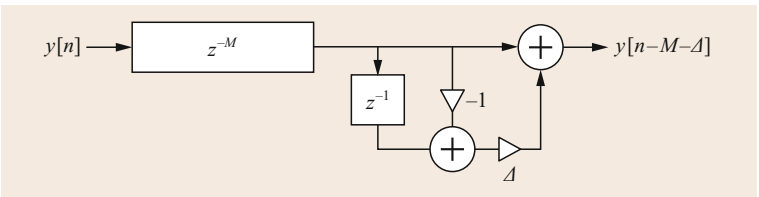


Fig. 13.5 A linearly interpolated delay line

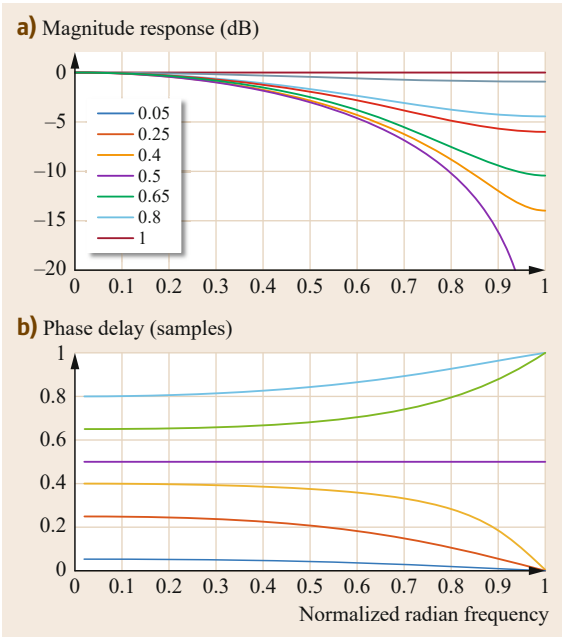


Fig. 13.6 **(a)** Frequency magnitude and **(b)** phase delay for linear interpolation with fractional delays between 0 and 1

magnitude. The problem then becomes one of finding an allpass filter with the desired phase delay property.

The difference equation for a general first-order allpass filter is given by

$$\begin{aligned} y[n] &= a \cdot x[n] + x[n-1] - a \cdot y[n-1] \\ &= a \cdot (x[n] - y[n-1]) + x[n-1] \,. \end{aligned}$$

Its transfer function is

$$H(z) = \frac{a + z^{-1}}{1 + az^{-1}} \,,$$

from which the phase delay can be estimated for low frequencies as [13.2]

$$-\frac{\angle H(\mathrm{e}^{-\mathrm{j}\omega})}{\omega} \approx \frac{1-a}{1+a} \quad \text{as} \quad \omega \to 0 \,,$$

where $\angle H(\mathrm{e}^{-\mathrm{j}\omega})$ is the phase response of $H(z)$ at normalized discrete-time radian frequencies $\omega = 2\pi f / f_\mathrm{s}$, with sample rate f_s. Thus, the allpass coefficient a can be determined for a given desired fractional delay Δ as

$$a = \frac{1-\Delta}{1+\Delta} \,.$$

The phase delay of the first-order allpass filter for fractional delay values between 0.1 and 2 is plotted in Fig. 13.7. From the plot, it is clear that the phase delay of the filter is not constant across all frequencies (other than for $\Delta = 1$), with the phase delay converging to one at $f_\mathrm{s}/2$ for all fractional delay values. A general rule of thumb is to choose values in the range $0.3 \le \Delta \le 1.3$ to obtain more constant phase delay response at lower frequencies, together with the fastest decaying impulse response, which is desirable to minimize transient effects when dynamically changing the fractional delay length. The transient responses of first-order allpass filters used to implement several different fractional delay values are shown in Fig. 13.8. Note how values of Δ closer to 0.0 have significantly longer transient *tails*.

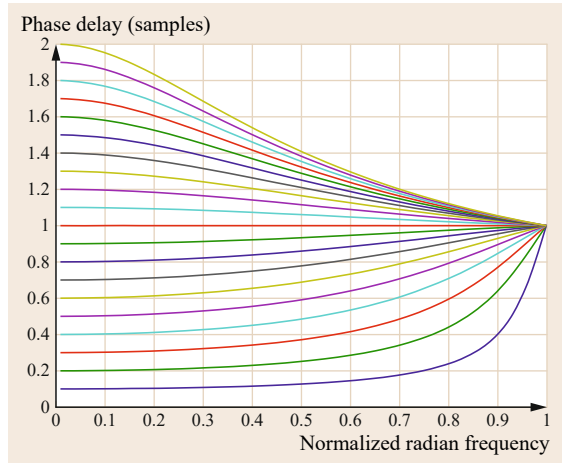


Fig. 13.7 Phase delay for first-order allpass filters with fractional delay settings between 0.1 and 2

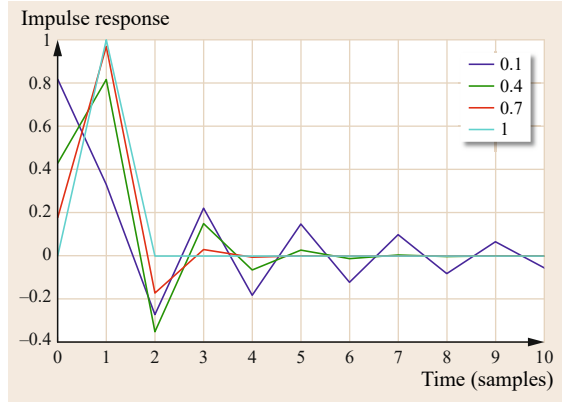


Fig. 13.8 Impulse response of first-order allpass filters with fractional delay settings of $\Delta = 0.1$, 0.4, 0.7, and 1.0

A block diagram for a first-order allpass interpolation filter at the output of a delay line is shown in Fig. 13.9. For values of $\Delta > 1$, a unit of delay can be subtracted from the adjoining delay line.

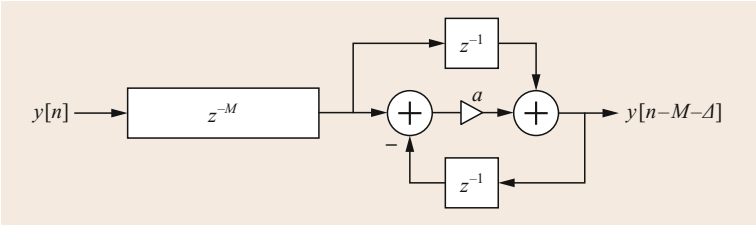


Fig. 13.9 A first-order allpass interpolated delay line

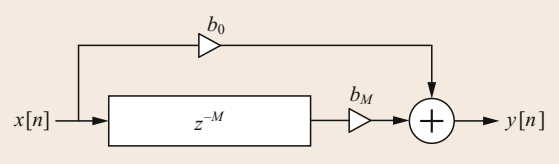


Fig. 13.10 A feedforward comb filter block diagram

First-order allpass and linear interpolation both involve the same level of computational complexity, though allpass interpolation does not introduce higher frequency attenuation. That said, linear interpolation provides the shortest transient response when changing Δ values and its phase delay is generally flatter out to higher frequencies. As well, linear interpolation can be used to arbitrarily interpolate between sample values in a delay line at any instant in time, which is not possible with allpass interpolation filters because of their transient response behavior (they must be *warmed up* before reaching their steady-state response). Thus, there are trade-offs and the best choice will depend on particular algorithm constraints.

13.1.4 Comb Filters

Comb filters are simple signal processing structures, made with delay lines, that have applications in many digital audio effects algorithms, including flanging and artificial reverberation. They get their name from the shape of their frequency magnitude responses, as will be shown below.

The signal processing block diagram for a feedforward comb filter is shown in Fig. 13.10. The difference equation for the filter is

$$y[n] = b_0 x[n] + b_M x[n-M] \,,$$

from which its transfer function is determined as

$$H(z) = b_0 + b_M z^{-M} \,.$$

The frequency magnitude response of the feedforward comb filter is thus given by

$$\left|H(\mathrm{e}^{\mathrm{j}\omega})\right| = \left|b_0 + b_M \mathrm{e}^{-\mathrm{j}\omega M}\right| \,, \quad -\pi \le \omega \le \pi \,,$$

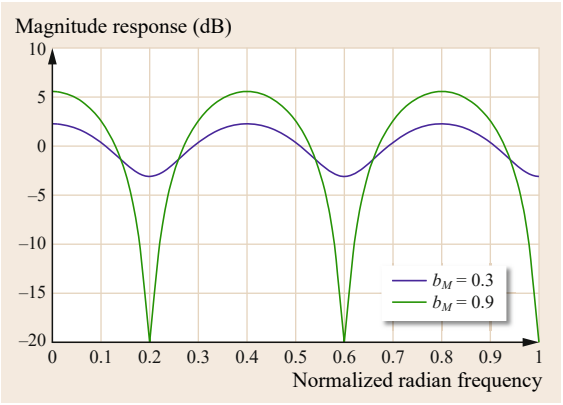


Fig. 13.11 Magnitude response of a feedforward comb filter with $M = 5$, $b_0 = 1$, and $b_M = 0.3$ and 0.9

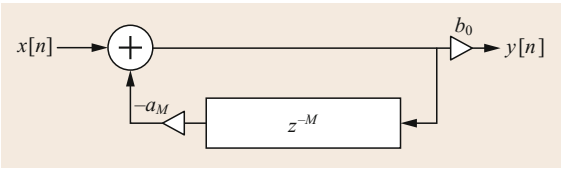


Fig. 13.12 A feedback comb filter block diagram

which is plotted in Fig. 13.11 for $M = 5$, $b_0 = 1$, and $b_M = 0.3$ and 0.9. The maximum gain is 6 dB (or a linear gain of 2), which occurs at frequencies for which the output from the delay line is in phase with the direct path. In between these frequency values, the two signals are out of phase and thus destructively interfere with each other, to an extent controlled by the coefficient b_M.

The frequencies of the notches in Fig. 13.11 occur at the roots of the transfer function $H(z)$. For b_0 and b_M positive, there are M notches evenly spaced in frequency from $f_s/(2M)$ to f_s Hz at increments of f_s/M Hz. If either b_0 or b_M is negative, the notches are shifted in frequency to start at 0 Hz and increment by f_s/M Hz. Note that the frequency axis of Fig. 13.11 is normalized from 0 to $f_s/2$ Hz.

The block diagram of a feedback comb filter is illustrated in Fig. 13.12. The difference equation for this filter is given by

$$y[n] = b_0 x[n] - a_M y[n-M] \,.$$

For stability, the coefficient a_M must have a magnitude < 1. The transfer function of the feedback comb filter is

$$H(z) = \frac{b_0}{1 + a_M z^{-M}} ,$$

from which the amplitude response is found as

$$G(\omega) = \left|H(e^{j\omega})\right| = \left|\frac{b_0}{1 + a_M e^{-j\omega M}}\right| ,$$
$$-\pi \le \omega \le \pi .$$

The magnitude response of a feedback comb filter is shown in Fig. 13.13 for $M = 5$, $b_0 = 1$, and $a_M = -0.3$ and -0.9. The frequencies of constructive feedback can be found from the roots of the denominator of the transfer function. The maximum gain is $b_0/(1 + a_M)$, which involves constructive feedback at the same frequencies where the notches were found in the feedforward comb filter.

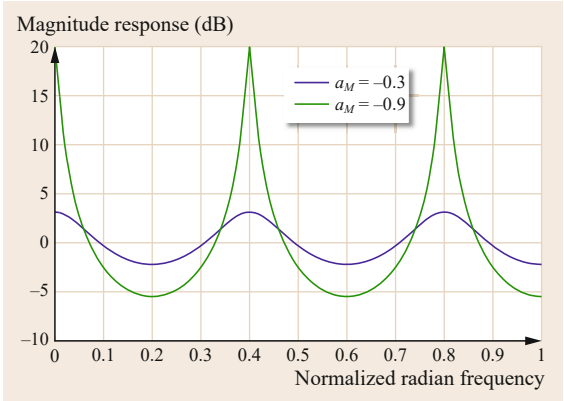


Fig. 13.13 Magnitude response of a feedback comb filter with $M = 5$, $b_0 = 1$, and $a_M = -0.3$ and -0.9

13.2 Simulating Sound Wave Propagation

Sound is transported through air via waves that travel at an approximate speed of 341 meters per second (at room temperature). As a result, there is a time (or propagation) delay for sound to travel from an emitting source to a listener some distance away. This is most obvious when the distance between a sound source and listener is large, for example when observing fireworks from a few kilometers away. But these inherent time delays are equally important for various acoustic phenomena over shorter distances as well. For example, the propagation delay for sound waves traveling inside a clarinet controls the resulting frequency of reed oscillations, or the sounding frequency of the instrument. When the effective length of the clarinet is shortened by opening holes along its length, the distance traveled by the waves inside the air column decreases and the sounding frequency increases.

Given what has previously been said about delay lines, it should come as no surprise to learn that delay lines are commonly applied in audio signal processing contexts to simulate sound wave propagation. A given distance d between source and listener will result in a time delay of d/c seconds (where c is the speed of sound propagation). In a discrete-time signal processing context, this time delay must be related to the sample period $T_s = 1/f_s$, where f_s is the sample rate in samples per second. Thus, the length of a delay line in samples M needed to simulate sound propagation over a distance d is given by $M = d/(cT_s)$. Note that the quantity cT_s represents the distance traveled by sound in a single sample period, which is about 7 mm at a sample rate of 48 000 Hz.

Sound waves emitted by a source will generally decrease in pressure level (and thus perceived loudness) as the distance from the source increases. For waves traveling in an open space or a large room, this decay or attenuation is due in part to the spreading of wavefront energy over a larger and larger spherical surface area as it propagates away from the source, as well as losses due to molecular interactions in air. For waves traveling in a confined space, such as a uniform pipe, there will be losses due to interactions with the surrounding walls, though not necessarily any attenuation due to wavefront spreading (if the wavefronts are assumed to be planar, which is typically the case in uniform pipes). Thus, an accurate simulation of sound travel must incorporate some distance-dependent gain control.

A distributed simulation approach would make use of gain factors g to represent the loss experienced over the distance traveled per unit delay, as diagrammed at the top of Fig. 13.14. For efficiency, these factors can be *commuted* (assuming linearity) and implemented at a single (or just a few) discrete locations in the system (as shown in the bottom of Fig. 13.14). In reality, these losses will be frequency dependent (typically more losses at higher frequencies) and thus more accurately represented with appropriately designed digital filters.

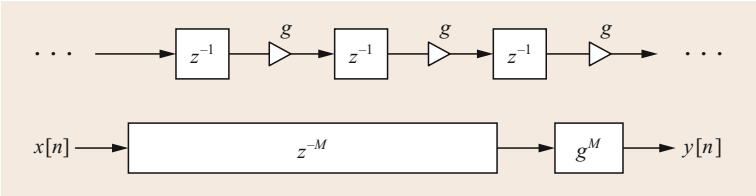


Fig. 13.14 Damped traveling-wave simulators with losses per unit sample (*top*) and with commuted losses (*bottom*)

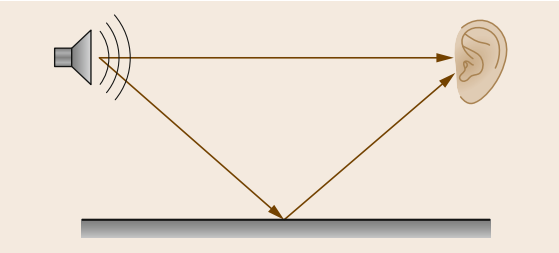


Fig. 13.15 Floor reflection illustration

13.2.1 Wave Reflections

If a traveling wave encounters a change in the physical properties of the medium through which it propagates, the wave will be perturbed where the change occurs. This perturbation generally involves some level of reflection, absorption, and transmission at the boundary. For example, if a wavefront impinges on a perfectly rigid surface, all of the wave energy will be reflected from the surface. However, if the surface is instead covered with a layer of absorbing material, only a portion of the wave energy will be reflected, with the remainder being trapped and damped within the material. The extent of such reflection can be characterized by a reflection coefficient (R), which specifies the ratio of reflected to incident wave energy. Materials that are very reflective will have a value of R close to 1, while R will be close to zero for materials that are very absorptive. In general, the reflection coefficient will be frequency dependent. In other words, materials will normally reflect or absorb waves of differing frequencies by different amounts.

Wave reflection from surfaces will also depend on the shape of the surface. If an acoustic wave encounters a rigid wall that is flat over at least several wavelengths in all directions, the wavefront will be reflected from that surface at an angle equal to its angle of incidence (referred to as *specular* reflection). Conversely, a wall that is very uneven will reflect a wavefront in many directions (referred to as *diffuse* scattering).

Waves can be *guided* within structures of uniform geometry, a common musical example being the roughly cylindrical air column of a clarinet. Wave reflection will occur in such *waveguides* at any geometrical discontinuity, such as the location of an open hole. Wave components will be partly reflected at such a discontinuity and partly transmitted through it into the surrounding air, with lower frequency components generally being more strongly reflected than higher frequency components. The size of the hole will control the frequency-dependence of this behavior.

Figure 13.15 illustrates a source–listener arrangement with sound wave reflection from an assumed rigid floor of infinite extent. In this case, two propagation paths exist for sound wave travel between the source and listener. The system of Fig. 13.16 (*top*) provides a signal processing block diagram to simulate the sound wave propagation using digital delay lines. The scale factors g_d and g_r account for losses over the respective direct and reflected paths due to air absorption and spherical spreading. If the floor had a reflection coefficient less than one, this could also be included in the g_r factor. The height of the source and listener will control the time delay between the arrival of the direct and reflected waves. The delay common to the two

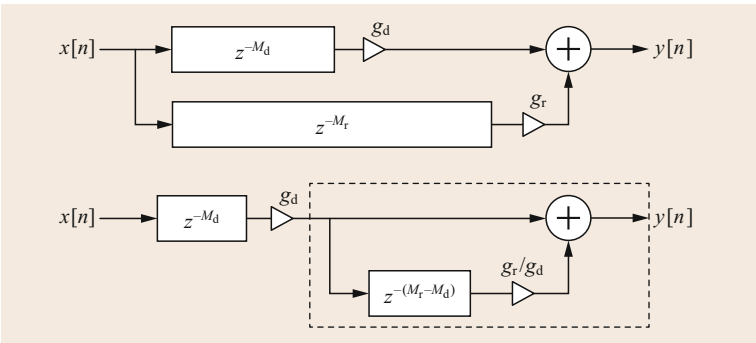


Fig. 13.16 Floor reflection block diagrams

paths can be pulled out and implemented separately, as illustrated in the lower part of Fig. 13.16. In this case, the length of the delay line for the reflected path must be adjusted by subtracting from it the common delay length and its attenuation factor appropriately scaled.

The portion of the signal processing block diagram within the dashed lines at the bottom of Fig. 13.16 is a feedforward comb filter. This simulated system results in a feedforward filter structure because none of the propagated sounds return to the listener. Thus, depending on the distance between the direct and reflected paths, certain frequency components in the sound will be destructively canceled at the listener position, which would correspond to the notches in the frequency response of the feedforward comb filter. In this way, the feedforward comb filter is a computational physical model of a source–listener arrangement involving a direct and single reflected path.

Another simulated source–listener arrangement is shown in Fig. 13.17, where the emitted sound is assumed to propagate back-and-forth between two parallel rigid walls of infinite extent. The corresponding signal processing block diagram is shown in Fig. 13.18. In this case, the sound is reflected back to the listener in a repeated fashion, thus the resulting filter structure has a feedback path. The portion of the signal processing structure within the dashed lines in Fig. 13.18 is a feedback comb filter. If the distance between the two walls is d and the listener is located at the wall opposite the source, the length of the initial delay line is $M1 = d/(cT)$, while the delay line in the feedback loop will be of length $M2 = 2d/(cT)$ (because the sound will return to the listener after propagating to the opposite wall and back). The gain factors would have to be determined

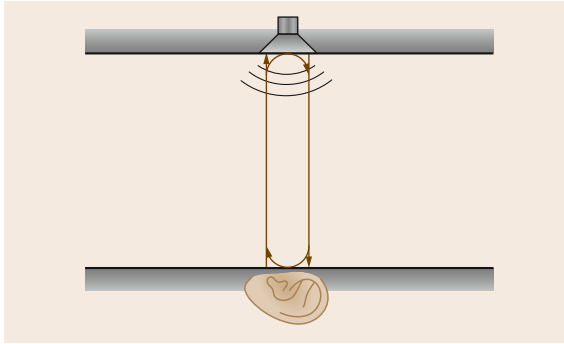


Fig. 13.17 Illustration of reflections between parallel walls

based on losses corresponding to air attenuation. If the walls were not assumed rigid, their reflection coefficient would also need to be included. As seen with the feedback comb filter, certain frequency components will feedback in phase with the emitted sound, producing *room* resonances at frequencies that are evenly spaced across the spectrum based on the distance between the walls. Such a phenomena is referred to as *flutter echo*. Thus, the feedback comb filter can be regarded as a computational physical model of a series of echoes, exponentially decaying and uniformly spaced in time.

Signal processing structures using delay lines have been widely explored to simulate much more complex room acoustics scenarios and artificial reverberation algorithms [13.3–7]. The fields of room acoustics modeling and artificial reverberation are well beyond the scope of this chapter but the underlying concepts are the same. In general, most artificial reverberation work has focused more on achieving flexible, efficient, and good sounding results without concern for accurately modeling particular room geometries.

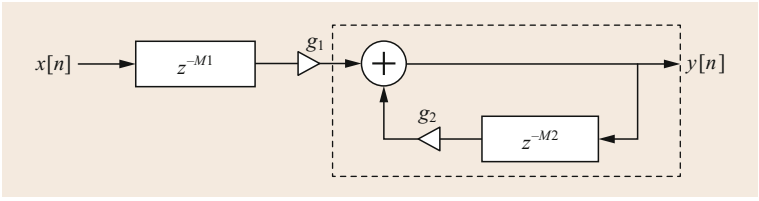


Fig. 13.18 Block diagram for simulation of reflections between parallel walls

13.3 Digital Waveguides

It is clear that delay lines provide an efficient way to simulate traveling-wave propagation. The concept of a *digital waveguide* derives more specifically from the discrete-time solution to the general one-dimensional (1-D) wave equation. Digital waveguides have been explored with great success over the past 30 years to achieve high-quality sound synthesis of musical instruments [13.8–17]. This brief treatment will focus on applications to plucked string instruments, though much research has been conducted on the use of digital waveguides in wind instrument modeling and synthesis as well [13.18–24].

13.3.1 1-D Traveling Waves

To first approximation, wave motion on a stretched string can be reasonably well approximated as 1-D, such that the string is considered to move only within an x–y plane. Lossless one-dimensional wave propagation along a stretched string is described by the 1-D wave equation as

$$\frac{\partial^2 y}{\partial t^2} = c^2 \frac{\partial^2 y}{\partial x^2} ,$$

where $c = \sqrt{\tau/\epsilon}$ is the speed of wave motion on the string, τ is the string tension, and ϵ is the mass density of the string. A particular solution to this equation was published by d'Alembert in 1747 having the general form

$$y(t,x) = y_\mathrm{r}\left(t - \frac{x}{c}\right) + y_\mathrm{l}\left(t + \frac{x}{c}\right) ,$$

for arbitrary functions $y_\mathrm{r}(\cdot)$ and $y_\mathrm{l}(\cdot)$. The subscripts r and l refer to right and left, respectively, because a function of $(t - x/c)$ can be interpreted as a fixed waveshape traveling to the right (in the positive x direction) over time and a function $(t + x/c)$ can be interpreted as a fixed waveshape traveling to the left (in the negative x direction) over time, both with speed c. As such, this solution describes *traveling waves* moving in opposite directions along a one-dimensional string, the superposition of which results in the observed physical displacement of the string. Note that it is theoretically possible to have only a single traveling-wave component present on the string, though a real string must be of finite length and thus fixed at one or both ends. A wave component that propagates into a fixed or free end will be reflected back in the opposite direction (as further described below).

To develop a discrete-time model or simulation of traveling-wave motion, it is necessary to sample the traveling-wave amplitudes in both time and space. The temporal sampling interval is T_s seconds, which corresponds to a sample rate $f_\mathrm{s} = 1/T_\mathrm{s}$ samples per second. The spatial sampling interval is given most naturally by $X = cT_\mathrm{s}$, or the distance traveled by a wave in one temporal sampling interval. In this way, each traveling-wave component moves left or right one spatial sample for each time sample.

The time and spatial sampling is accomplished with the following change of variables

$$\begin{aligned} x &\to x_m = mX \\ t &\to t_n = nT_\mathrm{s} , \end{aligned}$$

where n and m are integer indices of time and space, respectively. The traveling-wave solution then becomes

$$\begin{aligned} y(t_n, x_m) &= y_\mathrm{r}\left(t_n - \frac{x_m}{c}\right) + y_\mathrm{l}\left(t_n + \frac{x_m}{c}\right) \\ &= y_\mathrm{r}\left(nT_\mathrm{s} - \frac{mX}{c}\right) + y_\mathrm{l}\left(nT_\mathrm{s} + \frac{mX}{c}\right) \\ &= y_\mathrm{r}[(n-m)T_\mathrm{s}] + y_\mathrm{l}[(n+m)T_\mathrm{s}] . \end{aligned}$$

This representation can be further simplified by suppressing explicit reference to T_s and defining

$$\begin{aligned} y^+(n,m) &= y_\mathrm{r}(nT_\mathrm{s}, mT_\mathrm{s}) , \\ y^-(n,m) &= y_\mathrm{l}(nT_\mathrm{s}, mT_\mathrm{s}) , \end{aligned}$$

where the superscripts '+' and '−' denote wave travel to the right and left (or $+x$ and $-x$ directions), respectively. The resulting expression for physical displacement at time index n and location index m is given as the sum of the two traveling-wave components

$$y(t_n, x_m) = y^+(n-m) + y^-(n+m) . \qquad (13.1)$$

The term $y_\mathrm{r}[(n-m)T_\mathrm{s}] = y^+(n-m)$ can be interpreted as the output of an m-sample delay line of input $y^+(n)$. Similarly, the term $y_\mathrm{l}[(n+m)T_\mathrm{s}] = y^-(n+m)$ can be interpreted as the input to an m-sample delay line with output $y^-(n)$. The physical wave variable (string displacement in this case) is given by the superposition of the two traveling-wave components. Thus, we can accurately model one-dimensional motion on a string using two systems of unit delays, to model left- and right-going traveling waves, with delay-line values summed at corresponding *spatial* locations to obtain physical outputs, as depicted in Fig. 13.19.

In most modeling contexts, the calculation of physical output values can be limited to just one or two

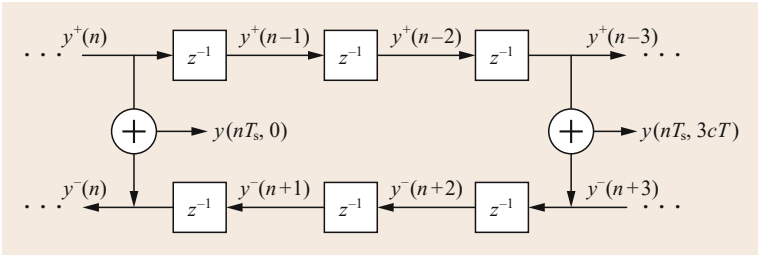


Fig. 13.19 Discrete-time simulation of ideal, lossless wave propagation with observation points at $x = 0$ and $x = 3X = 3cT_s$

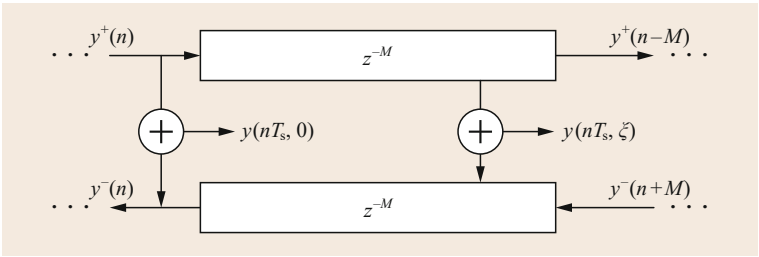


Fig. 13.20 Digital waveguide simulation of ideal, lossless wave propagation using delay lines

discrete spatial locations. Individual unit delays are more typically combined and represented by digital delay lines, as shown in Fig. 13.20.

Any ideal, lossless, one-dimensional waveguide can be simulated in this way. The model is exact at the sampling instants to within the numerical precision of the processing system. The delay lines can be initialized with displacement data corresponding to an initial pluck waveshape, though the waveshapes must be bandlimited to less than half the sampling frequency to avoid aliasing (as in all discrete-time systems).

13.3.2 Lossy Wave Propagation

Real wave propagation is never lossless. Sound waves in air lose energy via molecular frictional forces. Mechanical vibrations in strings are dissipated through yielding terminations, the viscosity of the surrounding air, and via internal frictional forces. In general, these losses vary with frequency.

Losses are often well approximated by the addition of one or more terms to the wave equation. In the simplest case, we can add a frequency-independent force term that is proportional to the transverse string velocity ($\partial y/\partial t$),

$$\tau \frac{\partial^2 y}{\partial x^2} = \epsilon \frac{\partial^2 y}{\partial t^2} + \mu \frac{\partial y}{\partial t} ,$$

where μ is a resistive proportionality constant. That is, the force lost to internal damping in the string is assumed to scale linearly with velocity. Assuming the resistive coefficient is relatively small, the following general class of solutions to this equation can be found

$$y(t,x) = \mathrm{e}^{-(\mu/2\epsilon)x/c} y_\mathrm{r}(t - x/c) + \mathrm{e}^{(\mu/2\epsilon)x/c} y_\mathrm{l}(t + x/c) .$$

The sampled solution is then

$$y(t_n, x_m) = g^m y^+(n-m) + g^{-m} y^-(n+m) ,$$

where

$$g = \mathrm{e}^{-\mu T_s/2\epsilon} .$$

Thus, a small decay factor should be inserted between each unit delay, as illustrated in Fig. 13.21.

Because the system is linear and time-invariant, the loss terms can be commuted and implemented at discrete points for efficiency as previously discussed in Sect. 13.2. In the more realistic situation where losses are frequency dependent (and typically of *lowpass* characteristic), the g factors are replaced with frequency responses of the form $G(\omega)$. These responses can likewise be commuted and implemented at discrete spatial locations within the system.

13.3.3 Reflections

Thus far, we have only considered wave propagation along a uniform, one-dimensional string of undefined length. In an anechoic or nonreflecting waveguide, waves traveling in only one direction may exist and can thus be simulated with just a single delay line. But in real situations, the media in which waves travel are of finite length and reflections occur at the boundaries that

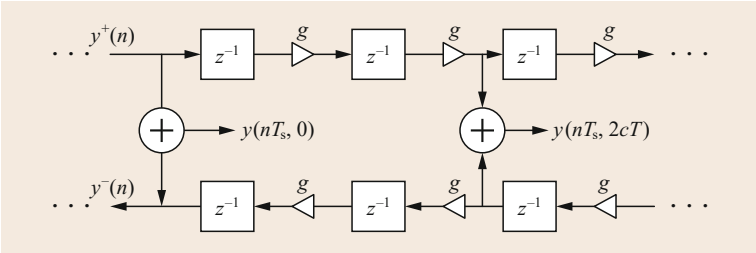


Fig. 13.21 Discrete-time simulation of lossy wave propagation

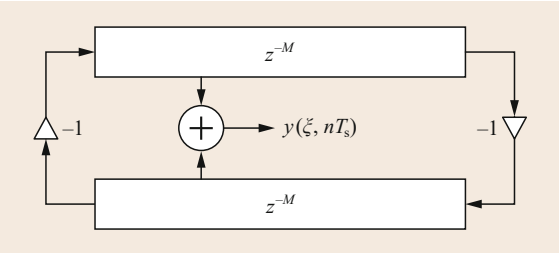


Fig. 13.22 Digital waveguide simulation of 1-D wave propagation on a string fixed at both ends

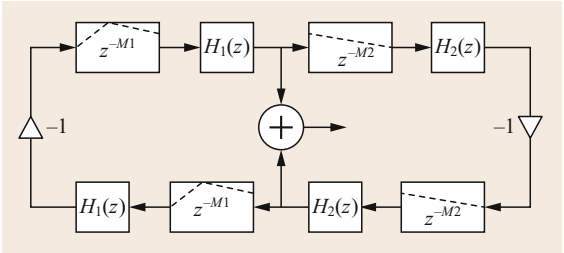


Fig. 13.23 Digital waveguide simulation of lossy wave propagation on a string fixed at both ends (with initial displacement)

give rise to waves traveling in two directions per dimension.

If we consider a string to be rigidly fixed at a position L, the boundary condition at that point is $y(t, L) = 0$ for all time. From the traveling-wave solution to the wave equation, we then have $y_r(t-L/c) = -y_l(t+L/c)$, which implies that displacement traveling waves reflect from a fixed end with a change of sign (or a reflection coefficient $R = -1$). Thus, a displacement wave component traveling into a rigidly fixed boundary at $x = L$ will be completely reflected back from that boundary, though it must be inverted when reflected back so that the boundary condition $y(t, L) = 0$ is satisfied. The simulation of displacement wave motion in a string rigidly terminated at both its ends (and without losses) is illustrated in Fig. 13.22.

In general, any change in the physical properties of the string (such as a diameter change) will cause wave scattering, which involves partial reflection and partial transmission at the boundary in such a way that energy is conserved.

13.3.4 The Plucked String Model

Finally, a full plucked string model is obtained by adding loss filters to the rigidly terminated string model from above and providing an initial displacement condition before the model is computed, as illustrated in Fig. 13.23. For example, if we assume the string is initially plucked by pulling it upward at a position $x = L/4$ and then releasing it, this condition would be simulated by initializing the internal elements of the upper and lower delay lines with values corresponding to that shape (1/2 of the displacement for each delay line, since the physical displacement of the string is given by the sum of the two traveling-wave components at any given position). For the moment, we will not be concerned with the details of the loss filters, though they would have a lowpass characteristic with a cutoff frequency that depends on the number of samples in each delay line (thus, the length of the simulated string).

The plucked string digital waveguide model above is typically modified in several ways to better correspond to the physical reality of plucked string instruments. In general, the majority of the sound produced by string instruments does not come directly from the string itself because strings do not transfer their vibrations well into the surrounding air (one might simply say that they cannot *push* the air in an efficient way). Rather, some of the string energy is transferred through one of the string terminations (the bridge) into a soundboard that can better transform the mechanical vibrations of the string into sound pressure fluctuations in air. Thus, the bridge termination of the string is not perfectly rigid as originally assumed but in fact can vibrate in a way that allows some of the string energy to *leak* into the body and then be radiated into the air. Rather than be represented by a reflection coefficient of -1, the bridge termination should be modeled by a frequency-dependent filter, $H_b(z)$, that simulates this behavior.

The output of the model is taken from both sides of the bridge filter as shown in Fig. 13.24a, which corre-

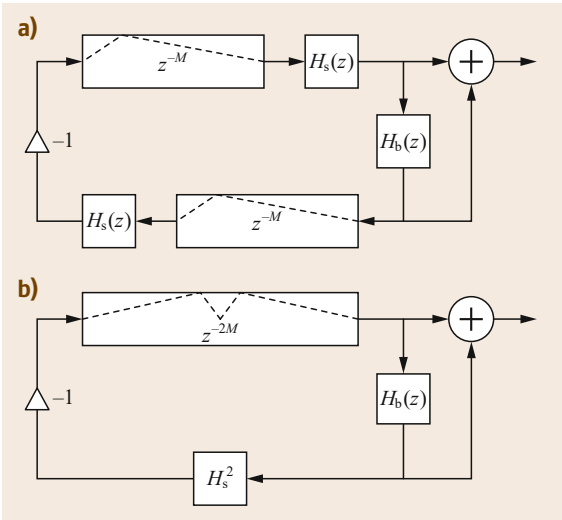


Fig. 13.24a,b Digital waveguide simulation of lossy wave propagation on a string with output through a nonrigid bridge. **(a)** Two delay line implementation. **(b)** Simplified (commuted) and more efficient implementation

sponds to the physical motion of the bridge. Because the system is linear, this structure can subsequently be simplified by rearranging and combining the two delay lines and the two string loss filters as illustrated in Fig. 13.24b.

The model output as calculated in Fig. 13.24 represents the motion of the bridge in response to vibrations of a plucked string. To accurately represent the sound from a string instrument, this output should then be input to a filter representing the vibrational response of the instrument body and its subsequent radiation of sound into the air. The body response is typically given by its time-domain impulse response, which can be measured or simulated. For example, one can tap lightly on the bridge of an instrument and record the resulting sound at some position away from the instrument in an anechoic room. The time samples of this recording, normally several tens of thousands of samples long, become the coefficients of an FIR bridge-to-air filter. Every output sample from the digital waveguide string model is then filtered by this body filter. As all parts of the model are assumed to be linear, however, it is possible to rearrange the body and string responses using an approach referred to as *commuted synthesis* [13.10]. In linear systems, the order of operations can be changed without affecting the overall output. For the plucked string model, commutivity thus allows the body impulse response to be used as the input to the delay line representing the string, with some prefiltering to account for pluck position. This results in significant computational savings because implementation of the long FIR body filter is avoided.

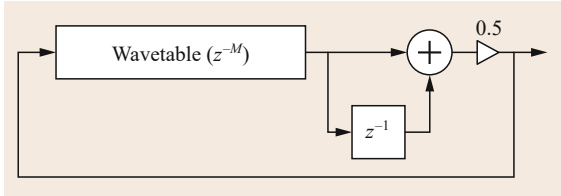


Fig. 13.25 The Karplus–Strong algorithm (reproduced from [13.25])

A variety of further simplifications are often exploited for computational efficiency. For example, the bridge, loss filters, and *nut* multiplier (the −1 scalar) are typically commuted together and the output taken from only one *side* of this filter. The resulting single loss filter can be designed based on a desired string decay rate [13.2]. Also, the resulting output of the system continues to sound like a plucked string for almost any sort of initialization of the delay line values. That is, any sort of wide-bandwidth input (including a single impulse in only one memory location of the delay line) will produce the sensation of a plucked string sound. These simplifications do not necessarily have physical justifications but are made for efficiency and because the resulting auditory result remains convincing.

At this point, it is interesting to consider the signal processing block diagram of Fig. 13.25 proposed by *Karplus* and *Strong* [13.25] to produce *surprisingly rich and natural* plucked string sounds. Karplus and Strong had been experimenting in the late 1970s with variations on wavetable synthesis in an effort to find efficient ways to achieve more interesting sounds. Wavetable synthesis generally involves the storage and subsequent playback, at varying rates (or frequencies), of a single period of a periodic sound. Perfectly periodic sounds do not sound natural and thus, Karplus and Strong added an operation at the output of the wavetable (Fig. 13.25) involving a simple averaging of the current and previous outputs, to change the content of the table over time. This operation is defined by the difference equation $y[n] = 0.5(x[n] + x[n-1])$, a first-order FIR filter that is equivalent to a linear interpolation filter with $\Delta = 0.5$ (with a frequency response as shown in Fig. 13.6). Karplus and Strong suggested initializing the wavetable with noise. By comparing Figs. 13.24 and 13.25, we see that the Karplus–Strong model can be interpreted as a simplified computational physical model of a plucked string, a result that was noted by *Jaffe* and *Smith* [13.2].

The Matlab code below, a minimally altered version of the delay line implementation provided in Sect. 13.1.1, implements the Karplus–Strong plucked string algorithm.

```
fs = 44100;
N = fs * 4; % compute 4 seconds
M = 60; % fixed delay line length
delayline = 2 * rand(1, M) - 1;
ptr = 1;
y = zeros(1, N); % output signal
xm1 = 0; % previous delay output

for n = 1:N,
  x = delayline(ptr); % delay output
  y(n) = 0.5 * (x + xm1); % do average
  xm1 = x; % save last output
  delayline(ptr) = y(n); % feedback
  ptr = ptr + 1; % increment pointer
  if ptr > M, % check pointer limit
    ptr = 1;
  end
end
soundsc( y, fs );
```

This section only touched upon the basics of musical instrument modeling and digital waveguide synthesis. Many more extensions have been proposed to account for numerous physical phenomena. For example, bending stiffness in strings results in dispersion, or a frequency-dependent speed of wave propagation, and this effect can be modeling using allpass filters [13.26]. The analysis and synthesis of transverse, longitudinal, and torsional string vibrations in multiple dimensions coupling at the bridge is reported in [13.27, 28]. The bowed string interaction has been studied in depth and efficient models developed in [13.29, 30]. Finally, digital waveguide modeling of wave propagation in two-dimensional (2-D) and three-dimensional (3-D) systems, such as in drum membranes, was first proposed by [13.31, 32]. The interested reader should consult the references provided in this chapter, as well as the comprehensive overview of [13.17].

References

13.1 T.I. Laakso, V. Välimäki, M. Karjalainen, U.K. Laine: Splitting the unit delay – Tools for fractional delay filter design, IEEE Signal Process. Mag. **13**, 30–60 (1996)

13.2 D.A. Jaffe, J.O. Smith: Extensions of the Karplus-Strong plucked string algorithm, Comput. Music J. **7**(2), 56–69 (1983)

13.3 M.R. Schroeder: Natural-sounding artificial reverberation, J. Audio Eng. Soc. **10**(3), 219–223 (1962)

13.4 M.R. Schroeder: Digital simulation of sound transmission in reverberant spaces (Part 1), J. Acoust. Soc. Ame. **47**(2), 424–431 (1970)

13.5 J.A. Moorer: About this reverberation business, Compt. Music J. **3**(2), 13–18 (1979)

13.6 J. Stautner, M. Puckette: Designing multichannel reverberators, Compt. Music J. **6**(1), 52–65 (1982)

13.7 J.M. Jot: An analysis/synthesis approach to real-time artificial reverberation. In: *Proc. Int. Conf. Acoust., Speech, Signal Process., San Francisco* (1992) pp. II.221–II.224

13.8 J.O. Smith: A new approach to digital reverberation using closed waveguide networks. In: *Proc. Int. Comput. Music Conf., Vancouver* (1985) pp. 47–53

13.9 J.O. Smith: Physical modeling using digital waveguides, Comput. Music J. **16**(4), 74–91 (1992)

13.10 J.O. Smith: Efficient synthesis of stringed musical instruments. In: *Proc. Int. Comput. Music Conf., Tokyo* (1993) pp. 64–71

13.11 J.O. Smith: Physical modeling synthesis update, Comput. Music J. **20**(2), 44–56 (1996)

13.12 M. Karjalainen, J.O. Smith: Body modeling techniques for string instrument synthesis. In: *Proc. Int. Comput. Music Conf., Hong Kong* (1996) pp. 232–239

13.13 V. Välimäki, J. Huopaniemi, M. Karjalainen, Z. Jánosy: Physical modeling of plucked string instruments with application to real-time sound synthesis, J. Audio Eng. Soc. **44**, 331–353 (1996)

13.14 J.O. Smith: Principles of digital waveguide models of musical instruments. In: *Applications of Digital Signal Processing to Audio and Acoustics*, ed. by M. Kahrs, K. Brandenburg (Kluwer, Boston, Dordrecht, London 1998) pp. 417–466

13.15 M. Karjalainen, V. Välimäki, T. Tolonen: Plucked string models: From the Karplus-Strong algorithm to digital waveguides and beyond, Comput. Music J. **22**(3), 17–32 (1998)

13.16 V. Välimäki, J. Pakarinen, C. Erkhut, M. Karjalainen: Discrete-time modelling of musical instruments, Rep. Prog. Phys. **69**(1), 1–78 (2005)

13.17 J.O. Smith: *Physical Audio Signal Processing* (W3K Publishing, Denver 2010)

13.18 V. Välimäki, M. Karjalainen, T.I. Laakso: Modeling of woodwind bores with finger holes. In: *Proc. Int. Comput. Music Conf., Tokyo* (1993) pp. 32–39

13.19 V. Välimäki, M. Karjalainen: Digital waveguide modeling of wind instrument bores constructed of truncated cones. In: *Proc. Int. Comput. Music Conf., Århus* (1994) pp. 423–430

13.20 G.P. Scavone, P.R. Cook: Real-time computer modeling of woodwind instruments. In: *Int. Symp. Musical Acoust., Leavenworth* (1998) pp. 197–202

13.21 M. van Walstijn, G.P. Scavone: The wave digital tonehole model. In: *Proc. Int. Comput. Music Conf., Berlin* (2000) pp. 465–468

13.22 G.P. Scavone: Time-domain synthesis of conical bore instrument sounds. In: *Proc. Int. Comput. Music Conf., Göteborg* (2002) pp. 9–15

13.23 M. van Walstijn, M. Campbell: Discrete-time modeling of woodwind instrument bores using wave variables, J. Acoust. Soc. Am. **113**(1), 575–585 (2003)

13.24 G.P. Scavone, J.O. Smith: A stable acoustic impedance model of the clarinet using digital waveguides. In: *Proc. Int. Conf. Digit. Audio Effects (DAFx-06), Montreal* (2006) pp. 89–94
13.25 K. Karplus, A. Strong: Digital synthesis of plucked-string and drum timbres, Comput. Music J. **7**(2), 43–55 (1983)
13.26 J. Abel, V. Välimäki, J.O. Smith: Robust, efficient design of allpass filters for dispersive string sound synthesis, IEEE Signal Process. Lett. **17**(4), 406–409 (2010)
13.27 G. Weinreich: Coupled piano strings, J. Acoust. Soc. Amer. **62**, 1474–1484 (1977)
13.28 B. Bank, M. Karjalainen: Passive admittance matrix modeling for guitar synthesis. In: *Proc. 2010 Int. Conf. Digital Audio Effects (DAFx-10), Graz* (2010)
13.29 M.E. McIntyre, J. Woodhouse: On the fundamentals of bowed-string dynamics, Acustica **43**(2), 93–108 (1979)
13.30 J.O. Smith: Efficient simulation of the reed-bore and bow-string mechanisms. In: *Proc. 1986 Int. Comput. Music Conf., The Hague* (1986) pp. 275–280
13.31 S. Van Duyne, J.O. Smith: Physical modeling with the 2-D digital waveguide mesh. In: *Proc. 1993 Int. Comput. Music Conf., Tokyo* (1993) pp. 40–47
13.32 S. Van Duyne, J.O. Smith: The 3D tetrahedral digital waveguide mesh with musical applications. In: *Proc. 1996 Int. Comput. Music Conf., Hong Kong* (1996) pp. 411–418

14. Convolution, Fourier Analysis, Cross-Correlation and Their Interrelationship

Jonas Braasch

The scope of this chapter is to derive and explain three fundamental concepts of acoustical signal analysis and synthesis: convolution, Fourier transformation, and cross-correlation. Convolution is an important process in acoustics to determine how a signal is transformed by an acoustical system that can be described through an impulse response, a room, for example. Fourier analysis enables us to analyze a signal's properties at different frequencies. This method is then extended to Fourier transformation to convert signals from the time domain to the frequency domain and vice versa. Further, the method of cross-correlation is introduced by extending the orthogonality relations for trigonometric functions that were used to derive Fourier analysis. The cross-correlation method is a fundamental concept to compare two signals. We will use this method to extract the impulse response of a room by comparing a signal measured with a microphone after being transformed by a room with the original, measurement signal emitted into the room using a loudspeaker. Based on this and other examples, the mathematical relationships between convolution, Fourier transformation, and correlation are explained to facilitate deeper understanding of these fundamental concepts.

In this chapter, we introduce and discuss three related methods to analyze and process time-based functions and systems: convolution, Fourier analysis, and cross-correlation. All three methods are of fundamental importance when examining acoustical systems such as musical instruments, concert venues, and electroacoustic systems, among many other acoustical applications. These techniques are also essential to understand how the auditory system processes sound.

14.1 Convolution

D'Alembert [14.1] was the first to publish an equation representing the convolution integral. A few years later, *Laplace* [14.2] used convolution to calculate the mean of a distribution – see [14.3] and [14.4]. *Doetsch* [14.5] and others laid the mathematical foundation to use the convolution operation to process signals. In the 1980s, computers were available to convolve audio signals with longer impulse responses in the frequency domain [14.6].

Convolution is a mathematical process in the time domain that describes a system with a defined impulse response as shown in Fig. 14.1. A signal passing through the system is transformed by convolution with the system's impulse response. A good example for such a system is an acoustical enclosure, such as the music performance space shown on the left in Fig. 14.2. A room has a characteristic impulse response, as shown on the right in Fig. 14.2, which determines much of what we hear when we speak or play an instrument in the aforementioned space. To hear this impulse response, we can excite the room with a very short impulsive signal such as a handclap, a balloon pop or a starter pistol.

An important property of the type of system we are investigating is that the duration of the impulse response

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_14

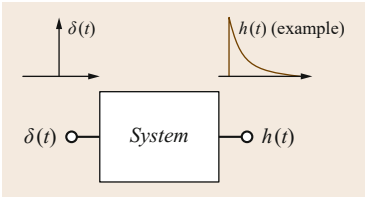


Fig. 14.1 Impulse response $h(t)$ of a system as the system's output to an impulsive input signal $s(t)$

is usually longer than the brief impulse. The duration of the impulse response of a typical church, for example, is in the order of several seconds. The central point that will allow us to grasp the concept of convolution is to understand how the system will respond to input signals of longer duration and how we can describe this mathematically.

To illustrate the general problem, let us next assume that we have two consecutive handclaps that generate an impulse response in a room (Fig. 14.3a) where the two resulting impulse responses overlap in time (Fig. 14.3b). Now we need to find a method describing how to overlap the impulse response of the first signal with that of the second. For this purpose, we mathematically define the handclap as an impulse. We will then see that we can describe any complex signal as a series of such impulses over time. The Dirac delta is strictly mathematically speaking not a function, but to intuitively understand the Dirac delta it is often represented as

$$\delta(t) = \begin{cases} \infty & \text{for } t = 0 \\ 0 & \text{for } t \neq 0 \end{cases} . \tag{14.1}$$

It is an impulse of infinite amplitude. Mathematically, only the distribution of the Dirac delta is defined, setting the area over the Dirac delta to be one

$$A = \int_{-\infty}^{\infty} \delta(t)\mathrm{d}t = 1 . \tag{14.2}$$

(The other commonly used delta function is the Kronecker delta, defined as

$$\delta(t) = \begin{cases} 1 & \text{for } t = 0 \\ 0 & \text{for } t \neq 0 \end{cases} , \tag{14.3}$$

with amplitude of 1.)

According to Fig. 14.1, the response of the room to the impulse signal $\delta(t)$ is

$$h(t) = \psi\left\{\delta(t)\right\} , \tag{14.4}$$

where ψ is the system response.

If we assume that the response of a room can be treated as time invariant, we will get the same impulse response $h(t)$ for a time-shifted impulse $\delta(t-\tau)$; the impulse response is simply shifted by the same time interval τ

$$h(t-\tau) = \psi\left\{\delta(t-\tau)\right\} . \tag{14.5}$$

Since the room is a sufficiently linear system, a change of the amplitude a of the input impulse $\delta(t-\tau)$ will change the amplitude of the impulse response in the same way

$$ah(t-\tau) = \psi\left\{a\delta(t-\tau)\right\} . \tag{14.6}$$

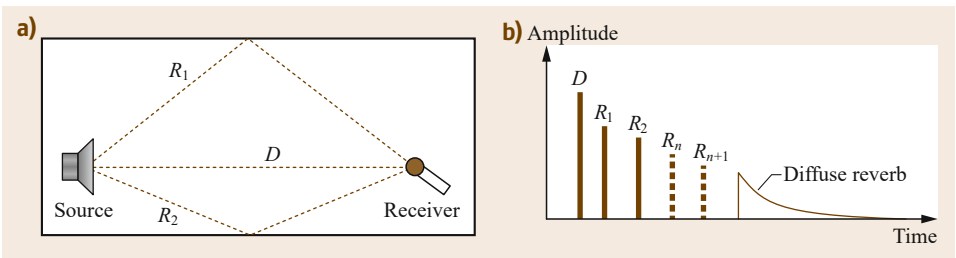


Fig. 14.2 **(a)** Example sound pathways, including one direct D and two wall reflections R_1 and R_2, between source and receiver in a room, forming the room impulse response shown in **(b)**

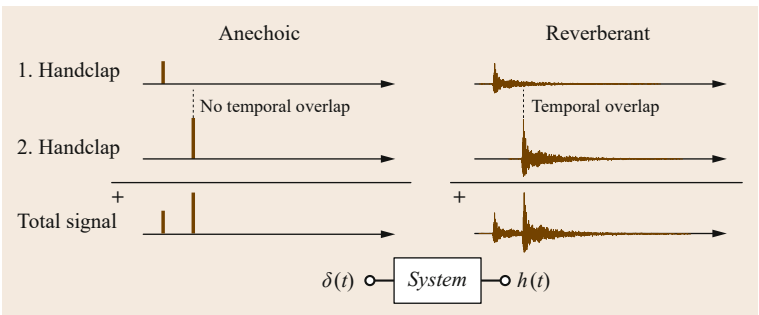


Fig. 14.3 Example of acoustically exciting a room with two successive handclaps. The *left graphs* show the time courses of the two handclap source signals individually (*two top graphs*) and as a total signal (*bottom graph*). The *right graphs* show the same signals but measured at a distance where the signals are affected by the room impulse. In this case, the two handclaps overlap in sound (*bottom-right graph*)

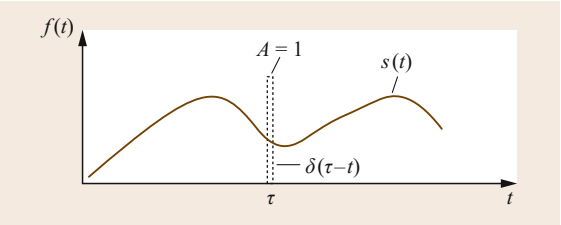


Fig. 14.4 Determining the value of a signal $s(t)$ using a Dirac delta function $\delta(t)$ with area A of 1

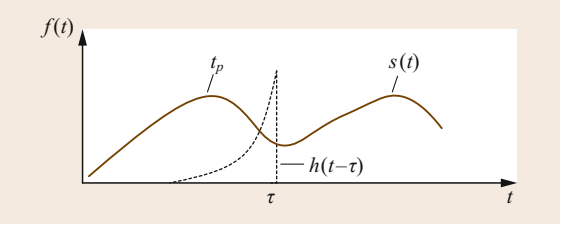


Fig. 14.5 Relationship between the signal $s(t)$ and the flipped impulse response $h(t)$ in the convolution process

The example of two consecutive impulses from Fig. 14.3 can now be written as

$$\begin{aligned} a_1 h(t-\tau_1) + a_2 h(t-\tau_2) &= \sum_{n=1}^{2} a_n h(t-\tau_n) \\ &= \sum_{n=1}^{2} a_n \psi\{\delta(t-\tau_n)\} = \sum_{n=1}^{2} \psi\{a_n \delta(t-\tau_n)\} \\ &= \psi\left\{\sum_{n=1}^{2} a_n \delta(t-\tau_n)\right\} \\ &= \psi\{a_1\delta(t-\tau_1) + a_2\delta(t-\tau_2)\}\,, \end{aligned} \tag{14.7}$$

by using the principle of superposition. Superposition is defined as

$$\psi\left\{\sum_{n=1}^{N} a_n s_n(t)\right\} = \sum_{n=1}^{N} a_n \psi\,\{s_n(t)\} = \sum_{n=1}^{N} a_n g_n(t) \tag{14.8}$$

and can be generally applied to *linear time-invariant* (LTI) systems. As the name indicates, an LTI system $\Psi\{s(t)\} = g(t)$ must meet the criteria of *linearity* toward an amplitude change a, i. e., $ag(t) = \Psi\{as(t)\}$, and *time invariance* toward a time shift τ, i. e., $g(t-\tau) = \Psi\{s(t-\tau)\}$.

Next, we analyze the system response $\Psi\{s(t)\}$ for a more complex input signal $s(t)$ by taking advantage of the fact that we can represent any signal as a series of adjacent amplitude-weighted Dirac delta functions. Because the Dirac delta has a defined area under the integral, we can use it to determine the value of a signal $s(t)$ at any time instance t we want

$$s(\tau) = \int_{-\infty}^{\infty} s(t)\delta(\tau - t)\mathrm{d}t\,. \tag{14.9}$$

The signal $s(\tau)$ represents the value of the signal $s(t)$ at time τ (Fig. 14.4).

We can also exchange t and τ and write

$$s(t) = \int_{-\infty}^{\infty} s(\tau)\delta(t - \tau)\mathrm{d}\tau\,. \tag{14.10}$$

This new equation basically describes how we resynthesize the signal $s(t)$ from many delta impulses $\delta(t-\tau)$ that are amplitude weighted with $s(\tau)$. Since we know the response of the system to every delta function in time, we can substitute $\delta(t)$ with $h(t)$ to obtain the system response to a complex time signal $s(t)$ using a process defined as *convolution*

$$g(t) = \int_{-\infty}^{\infty} s(\tau)h(t - \tau)\mathrm{d}\tau\,. \quad \textit{Convolution} \tag{14.11}$$

Convolution is often denoted as $g(t) = s(t) * h(t)$ (where the asterisk should not be mistaken for the multiplication sign). Figure 14.5 depicts the idea of the convolution process. At any time instance τ, we not only see the current signal $s(\tau)$ but also signal components back in time t_p, all weighted by the time-reversed impulse response $h(t-\tau)$, and this time reversal lends its name to convolution.

14.2 Fourier Frequency Analysis and Transformation

In 1822, *Jean-Baptiste Joseph Fourier* [14.7] described his analysis method, now known as Fourier analysis, to understand heat flow. A modern overview of the discrete Fourier transformation, which is commonly used for digital signal processing applications, can be found in [14.8]. This book also discusses the fast Fourier transformation (FFT). The FFT is a discrete Fourier transformation (DFT) method to save computational costs by using sparse matrices that contain mostly zero elements to reduce the computational load. The FFT method requires that the signal length be a power of two.

Convolution has a very interesting property: it becomes simple multiplication in the frequency domain. Since multiplication and conversion to the frequency domain are computationally less expensive than convolution, computer-based systems often take this approach. Before we can demonstrate how convolution in the time domain converts to multiplication in the frequency domain, we need to review general frequency analysis and conversion of signals based on the Fourier theorem (often referred to as *synthesis*).

Let us start with frequency analysis based on the Fourier series for periodic signals. A periodic signal $s(t)$

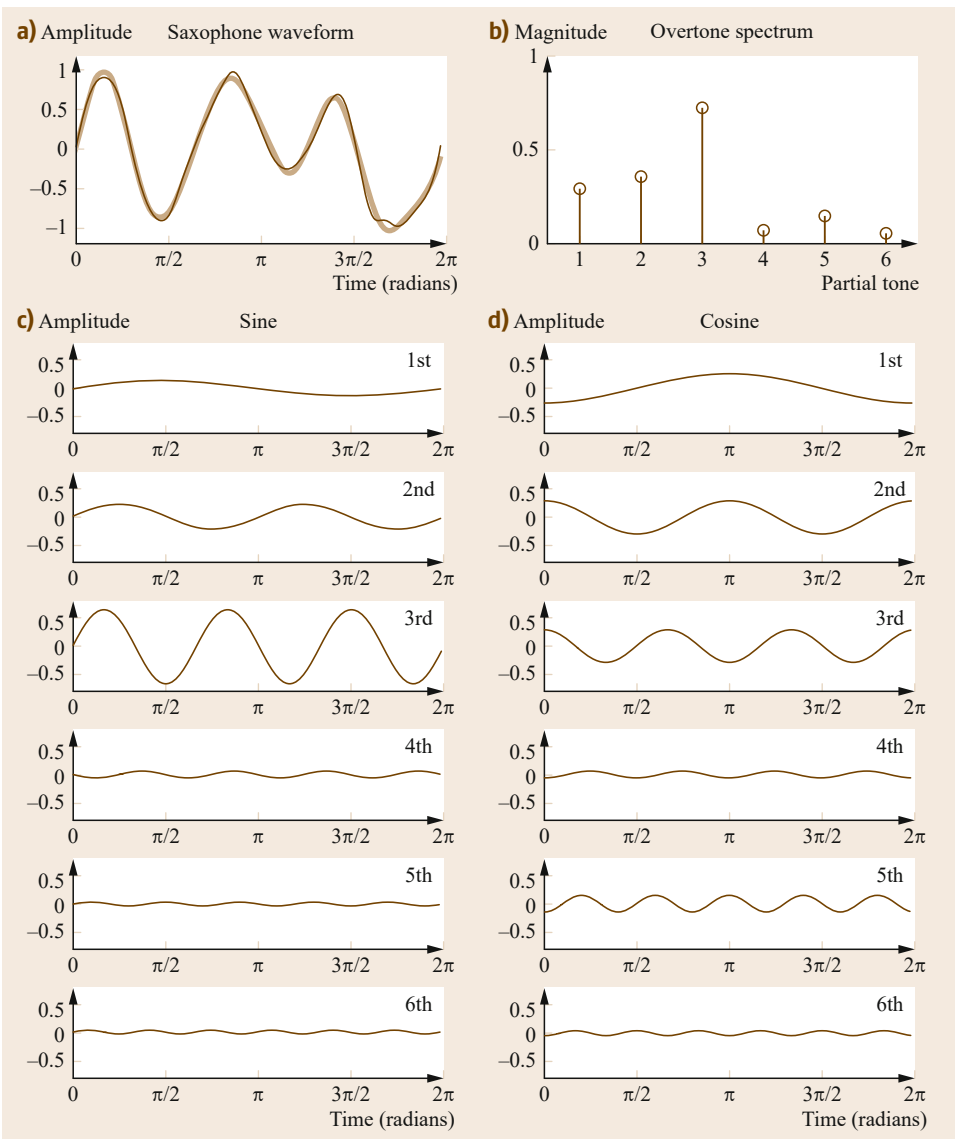


Fig. 14.6a–d The *solid black line* in **(a)** shows an example of a saxophone waveform ($f_0 = 440\,\text{Hz}$) with the individual magnitudes $c_{0n} = \sqrt{a_n^2 + b_n^2}$ of the partial tones shown in **(b)**. **(c,d)** show the periodic functions for each partial tone separated into the sine and cosine components. The *gray line* in **(a)** shows the reproduction of the saxophone waveform as the sum of the partial-tone components

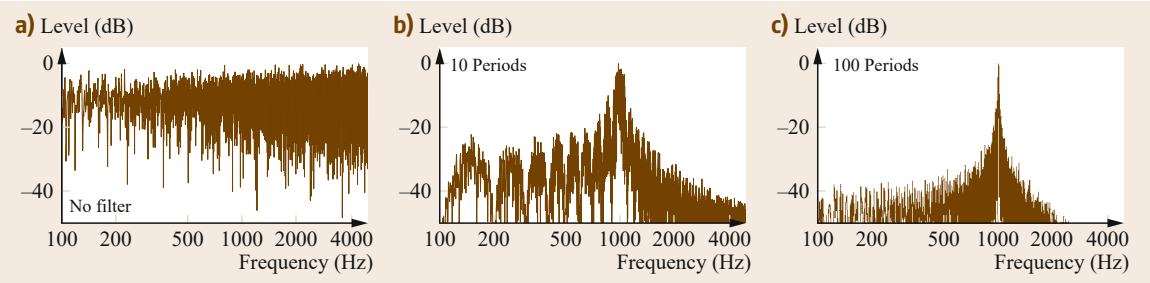


Fig. 14.7a–c Filter properties of a sine function. **(a)** shows the frequency spectrum of a Gaussian broadband noise signal. **(b)** depicts the spectrum of the same noise signal after it has been convolved with a 10-period-long 1 kHz sine tone. **(c)** shows the same condition, but this time the 100-period-long 1 kHz sine tone was used as the filter

can be defined in the time domain as

$$s(t+T) = s(t) \;\forall t \in \mathrm{Re}\ . \tag{14.12}$$

The variable T is the period of the signal, which indicates that the signal repeats itself every T seconds. For practical purposes, let us take the sample of a saxophone tone $s(t)$ shown in Fig. 14.6a (black solid line) as a periodic signal and determine the amplitudes of the tone's individual harmonics shown in Fig. 14.6b. Let us assume that we know the fundamental frequency $f_0 = 1/T$ of the tone. Since it is a wind instrument, we can also assume that the frequencies of the higher harmonics f_n are integer multiples of the fundamental tone nf_0. According to Fourier's theorem, we can now represent each harmonic s_n using a combined sine/cosine term

$$s_n(t) = a_n \cos(2\pi f_n t) + b_n \sin(2\pi f_n t)\ . \tag{14.13}$$

Remember that the sine and cosine functions are essentially the same except for a 90° phase shift. Consequently, the sine function represents the signal component for a 0° phase shift and the cosine function for a 90° phase shift with respect to the sine function. The variables a_n and b_n represent the amplitudes of each of the two trigonometric functions. Often the angular frequency ω, which is defined as $\omega = 2\pi f$, is used instead of the linear frequency f to avoid having to carry the term 2π along. We can represent the whole periodic signal as the sum over all partial tones – see gray line in Fig. 14.6a. The Fourier series is defined as

$$\begin{aligned} f(t) = a_0 &+ (a_1 \cos(\omega t) + b_1 \sin(\omega t)) \\ &+ (a_2 \cos(2\omega t) + b_2 \sin(2\omega t)) \\ &+ (a_3 \cos(3\omega t) + b_3 \sin(3\omega t)) + \ldots \end{aligned} \tag{14.14}$$

The a_0 term is the DC bias, the constant, average-valued offset of the function. Figure 14.6c,d shows the individual partial tone components of the equation above for the saxophone waveform.

14.2.1 Filter and Orthogonality Properties of Sine and Cosine Functions

The next question we must ask is how we can analyze a periodic signal $s(t)$ to determine the amplitude coefficients a_n and b_n. In this context we need to understand that a sine or cosine acts as a filter if we convolve it with a signal; For example, if we take a Gaussian noise signal and convolve it with a sine signal of frequency f_c, only the single component of the Gaussian noise at this particular frequency will remain. Figure 14.7 shows an example of a Gaussian noise signal (Fig. 14.7a) that is filtered with two sine signals of different duration but the same frequency $f_c = 1$ kHz. Figure 14.7b shows the result for the short filter (10 ms = 10 periods), which is noticeably wider in frequency than the signal for the long filter (100 ms = 100 periods) shown in Fig. 14.7c.

Based on this filter property, we can now use the sine and cosine functions to filter out each partial tone for the given fundamental frequency $f = f_0$; For example, if we multiply the saxophone sample $s(t)$ with the sine function of the fundamental frequency and integrate it over one period $T = 1/f_0$, we will obtain another sinusoidal function with an amplitude corresponding to the intensity of the saxophone tone at that frequency and phase corresponding to the sine signal. If we apply the cosine function at $f = f_0$, we also filter out the fundamental, but now obtain a sinusoidal function that corresponds with the phase of the cosine signal.

Mathematically, the filter properties of sine and cosine functions are described by their orthogonality to each other. For two sine signals, we find the orthogonality

$$\int_0^{2\pi} \sin(kx)\sin(lx)\mathrm{d}x = \pi\delta_{kl} \quad \text{for } k,l \in N^+, \tag{14.15}$$

with $x = \omega_0 t$. The variables k and l are harmonic indices in the form of positive integers, formerly denoted as n. The delta function here is the Kronecker delta, which is defined as

$$\delta_{kl} = \begin{cases} 1 & \text{for } k = l \\ 0 & \text{for } k \neq l \end{cases} . \tag{14.16}$$

Basically, two sine functions are orthogonal to each other, which means their product when integrated over one period 2π is zero, unless both signals have the same frequency. In the latter case the product is π. From a filter perspective, we can assume that the first sine function is the signal we would like to analyze and the second sine function is the filter. We will obtain a residual only if the filter is tuned to the same frequency as the signal. In all other cases, the filter is looking at frequencies where there is no energy in the signal and so there is no residual.

We also find a similar relationship for the cosine function

$$\int_0^{2\pi} \cos(kx)\cos(lx)\mathrm{d}x = \pi\delta_{kl} \quad \text{for } k, l \in N^+ . \tag{14.17}$$

For the product of a sine function with a cosine function, the two functions are always orthogonal to each other, even if they share the same frequency

$$\int_0^{2\pi} \cos(kx)\sin(lx)\mathrm{d}x = 0 \quad \text{for } k, l \in N^+ . \tag{14.18}$$

Using these three orthogonality relationships, we can determine the amplitude coefficients a_n and b_n

$$\begin{aligned} a_n &= \frac{1}{\pi}\int_0^{2\pi} s(x)\cos(nx)\mathrm{d}x \quad \text{for } n \geq 0 , \\ b_n &= \frac{1}{\pi}\int_0^{2\pi} s(x)\sin(nx)\mathrm{d}x \quad \text{for } n \geq 1 , \end{aligned} \tag{14.19}$$

with $x = \omega_0 t$. The cosine and sine functions act as phase-sensitive filters. The following relationship can be used to determine the DC component a_0

$$\int_0^{2\pi} \cos(0x)\mathrm{d}x = \int_0^{2\pi} \mathrm{d}x = 2\pi . \tag{14.20}$$

The calculated coefficients a_n and b_n can then be used to resynthesize the periodic signal $s(t)$

$$s(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos(nx) + b_n \sin(nx)). \tag{14.21}$$

The requirement for periodic signals to perform Fourier analysis should also be discussed. In theory, the Fourier series applies to periodic signals, but in practice aperiodic signals can be analyzed by approximating f_0 to zero, which of course prolongs the period $2\pi f_0$ toward ∞ ($x = \omega_0 t$, $T \to \infty$). In a practical application, of course, the integration window needs to be limited and the length of the sine and cosine filters then determines the frequency bandwidth of the extracted signal.

In the next section, we will explore how convolution in the time domain transforms into multiplication in the frequency domain. In order to demonstrate this, we need to represent the Fourier series using complex exponential functions

$$s(x) = \sum_{n=-\infty}^{\infty} c_n \mathrm{e}^{\mathrm{i}nx} . \tag{14.22}$$

The complex exponential function, also called the Euler identity, is defined as

$$\mathrm{e}^{\mathrm{i}nx} = \cos(nx) + \mathrm{i}\sin(nx) , \tag{14.23}$$

with i denoting the imaginary unit. Using the definition of the complex exponential function, we can express the cosine and sine functions in the following way

$$\begin{aligned} \cos(kx) &= \left(\frac{1}{2}\left(\mathrm{e}^{\mathrm{i}kx} + \mathrm{e}^{-\mathrm{i}kx}\right)\right) \quad \text{and} \\ \sin(kx) &= \left(\frac{1}{2\mathrm{i}}\left(\mathrm{e}^{\mathrm{i}kx} - \mathrm{e}^{-\mathrm{i}kx}\right)\right) . \end{aligned} \tag{14.24}$$

Now we can insert these two terms to derive the Fourier series representation using complex exponential functions

$$\begin{aligned} s(x) &= \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(kx) + b_k \sin(kx)) \\ s(x) &= \frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k \left(\frac{1}{2}\left(\mathrm{e}^{\mathrm{i}kx} + \mathrm{e}^{-\mathrm{i}kx}\right)\right) \right. \\ &\quad \left. + b_k \left(\frac{1}{2\mathrm{i}}\left(\mathrm{e}^{\mathrm{i}kx} - \mathrm{e}^{-\mathrm{i}kx}\right)\right)\right) \\ s(x) &= \sum_{n=-\infty}^{\infty} c_k \mathrm{e}^{\mathrm{i}kx} \\ &\quad \text{with } c_k = \frac{1}{2}(a_k - \mathrm{i}b_k) , \quad c_{-k} = \frac{1}{2}(a_k + \mathrm{i}b_k) , \\ &\quad \text{and } k \geq 0 . \end{aligned} \tag{14.25}$$

The complex amplitude coefficient c_k can then be determined using this formula

$$c_k = \frac{1}{\pi}\int_0^{2\pi} s(x)\mathrm{e}^{-\mathrm{i}kx}\mathrm{d}x\,,$$
$$s(x) = \sum_{k=-\infty}^{\infty} c_k\mathrm{e}^{\mathrm{i}kx}\,. \tag{14.26}$$

We can also describe the complex amplitude coefficient c_k using the magnitude c_{0k} and phase φ_k as

$$c_k = c_{0k}\mathrm{e}^{\mathrm{i}\varphi_k}\,, \tag{14.27}$$

with magnitude

$$c_{0k} = \sqrt{a_k^2 + b_k^2} \tag{14.28}$$

and phase

$$\varphi = \arctan\left(\frac{b_k}{a_k}\right). \tag{14.29}$$

For continuous frequency signals, the Fourier transformation is typically written as

$$S(f) = \int_{-\infty}^{\infty} s(t)\mathrm{e}^{-\mathrm{i}2\pi ft}\mathrm{d}t\,, \tag{14.30}$$

and the inverse Fourier Transformation is defined as

$$s(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(f)\mathrm{e}^{\mathrm{i}2\pi ft}\mathrm{d}f\,. \tag{14.31}$$

The term $1/2\pi$ is a normalization factor to ensure that a Fourier-transformed signal recovers its original amplitude when retransformed into the time domain using the inverse Fourier transformation.

14.2.2 Convolution in the Frequency Domain

In this section, we will demonstrate that convolution in the time domain becomes multiplication in the frequency domain. As mentioned above, the process of multiplication of Fourier transforms is a computationally less expensive operation than convolution. For this reason, computers typically perform convolution in the frequency domain.

Our proof starts with the convolution in the time domain

$$g(t) = s(t) * h(t) = \int_{-\infty}^{\infty} s(\tau)h(t-\tau)\mathrm{d}\tau\,. \tag{14.32}$$

In the next step, we replace the signal $h(t-\tau)$ with the inverse Fourier-transformed signal in the frequency domain $H(f)$. We will see later that this will help us to separate t and τ

$$g(t) = \int_{-\infty}^{\infty} s(\tau)\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} H(f)\mathrm{e}^{\mathrm{i}2\pi f(t-\tau)}\mathrm{d}f\right)\mathrm{d}\tau\,. \tag{14.33}$$

The exponential function provides the fortunate feature that the term $(t-\tau)$ is now in the exponent of the signal and we can separate the two variables using a mathematical property of the e-function $\mathrm{e}^{(x+y)} = \mathrm{e}^x\mathrm{e}^y$, in our case substituting $x = \mathrm{i}2\pi ft$ and $y = -\mathrm{i}2\pi f\tau$. We separate the t and τ terms using the relation $\mathrm{e}^{\mathrm{i}2\pi f(t-\tau)} = \mathrm{e}^{\mathrm{i}2\pi ft}\mathrm{e}^{-\mathrm{i}2\pi f\tau}$

$$g(t) = \int_{-\infty}^{\infty} s(\tau) \times\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} H(f)\mathrm{e}^{\mathrm{i}2\pi ft}\mathrm{e}^{-\mathrm{i}2\pi f\tau}\mathrm{d}f\right)\mathrm{d}\tau\,. \tag{14.34}$$

Next, we change the order of integration, which we can do as long as the double integral converges

$$g(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} s(\tau)H(f)\mathrm{e}^{\mathrm{i}2\pi ft}\mathrm{e}^{-\mathrm{i}2\pi f\tau}\,\mathrm{d}\tau\mathrm{d}f\,. \tag{14.35}$$

Then, we move out of the inner integral those terms that do not depend on τ to obtain

$$g(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} s(\tau)\mathrm{e}^{-\mathrm{i}2\pi f\tau}\,\mathrm{d}\tau H(f)\mathrm{e}^{\mathrm{i}2\pi ft}\mathrm{d}f\,. \tag{14.36}$$

It can now be seen that the inner integral is the Fourier transform of $s(\tau)$, and we substitute the term accordingly with $S(f)$

$$g(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(f)H(f)\mathrm{e}^{\mathrm{i}2\pi ft}\mathrm{d}f\,. \tag{14.37}$$

The inverse Fourier transform of $G(f)$ is written as

$$g(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(f)\mathrm{e}^{\mathrm{i}2\pi ft}\mathrm{d}f\,. \tag{14.38}$$

Since both integrals representing $g(t)$ have the same form, the integrand of the first equation $S(f)H(f)$ must equal the integrand of the second equation $G(f)$. Consequently, we have proved that

$$G(f) = S(f)H(f) \,. \tag{14.39}$$

We have shown that convolution in the time domain becomes multiplication in the frequency domain

$$g(t) = s(t) * h(t) \longleftrightarrow G(f) = S(f)H(f) \,. \tag{14.40}$$

This relationship is called the *convolution theorem for the Fourier transform.*

14.3 Cross-Correlation

The method of correlation was first introduced by *Bravais* in 1846 [14.9] to determine the error probability of a position in space. Forty years later, *Galton* [14.10] introduced the method of correlation to understand evolutionary processes, realizing that it could be applied to any set of variable pairs. *Pearson* [14.11] later proved that the method introduced by Bravais is the best fit to data with a linear relationship. Moreover, he introduced the Pearson product-moment correlation coefficient to measure the linear correlation between two variables, normalizing the degree of correlation between -1 (fully negative correlation) and $+1$ (fully positive correlation). *Pearson* [14.11] and *Stanton* [14.12] provide detailed overviews on the development of the correlation method. *Norbert Wiener* [14.13] laid the groundwork for the application of the cross- and autocorrelation functions to the analysis of time series, connecting the time-domain-based autocorrelation function to the frequency-domain-based power spectrum.

The cross-correlation method is a very powerful tool to compare the similarity of two signals. When we introduced Fourier analysis, we evaluated the similarity of two sine and cosine functions with different frequencies by multiplying them and integrating over their product. There we found that the sine and cosine functions (as well as sine and cosine functions of different frequencies) are orthogonal to each other

$$\int_0^T \cos(\omega t)\sin(\omega t)\mathrm{d}t = 0 \,, \tag{14.41}$$

with $\omega = 2\pi f$ and $f = 1/T$. However, at the same time, we agreed that cosine and sine functions are basically the same function, differing only by a phase shift of $\pi/2$

$$\cos(\omega t) = \sin\left(\omega\left(t + \frac{\pi}{2}\right)\right) . \tag{14.42}$$

In order to compensate for the phase shift, we now use a time delay τ to provide a tool to make our comparison function invulnerable to time and phase shifts between the two compared functions

$$\rho(\tau) = \int_0^T \cos(\omega t)\sin(\omega(t+\tau))\mathrm{d}t \,, \tag{14.43}$$

with the results $\rho(\tau = 0) = 0$ and $\rho(\tau = T/4 \to \varphi = \pi/2) = 1$. Now we generalize the equation above to any two input functions $x(t)$ and $y(t)$ and use the $f \to 0$ limit to define the cross-correlation

$$\rho_{x,y}(\tau) = \int_{-\infty}^{+\infty} x(t)y(t+\tau)\mathrm{d}t \,. \quad \textit{Cross-correlation} \tag{14.44}$$

For $\rho(\tau)|_{\max}$, the variable τ indicates for which time shift between the two functions they become most similar. The autocorrelation is a special case of the cross-correlation where both input signals x and y are identical

$$\rho_{x,x}(\tau) = \int_{-\infty}^{+\infty} x(t)x(t+\tau)\mathrm{d}t \,. \quad \textit{Autocorrelation} \tag{14.45}$$

The cross-correlation and the convolution are very similar functions that differ only in the sign of the second t. This becomes apparent if we use the variable convention from the previous equation for the convolution

$$g(\tau) = \int_{-\infty}^{+\infty} x(t)y(\tau - t)\mathrm{d}t = \int_{-\infty}^{+\infty} x(t)y(-t+\tau)\mathrm{d}t \,. \quad \textit{Convolution} \tag{14.46}$$

We can extend the cross-correlation function to the normalized cross-correlation function by dividing by the energies of both signals under analysis

$$\rho_{\mathrm{N}}(\tau) = \frac{\int_{-\infty}^{+\infty} x(t)y(t+\tau)\mathrm{d}t}{\sqrt{\int_{-\infty}^{+\infty} x^2(t)\mathrm{d}t \int_{-\infty}^{+\infty} y^2(t)\mathrm{d}t}} \,. \quad \textit{Normalized cross-correlation} \tag{14.47}$$

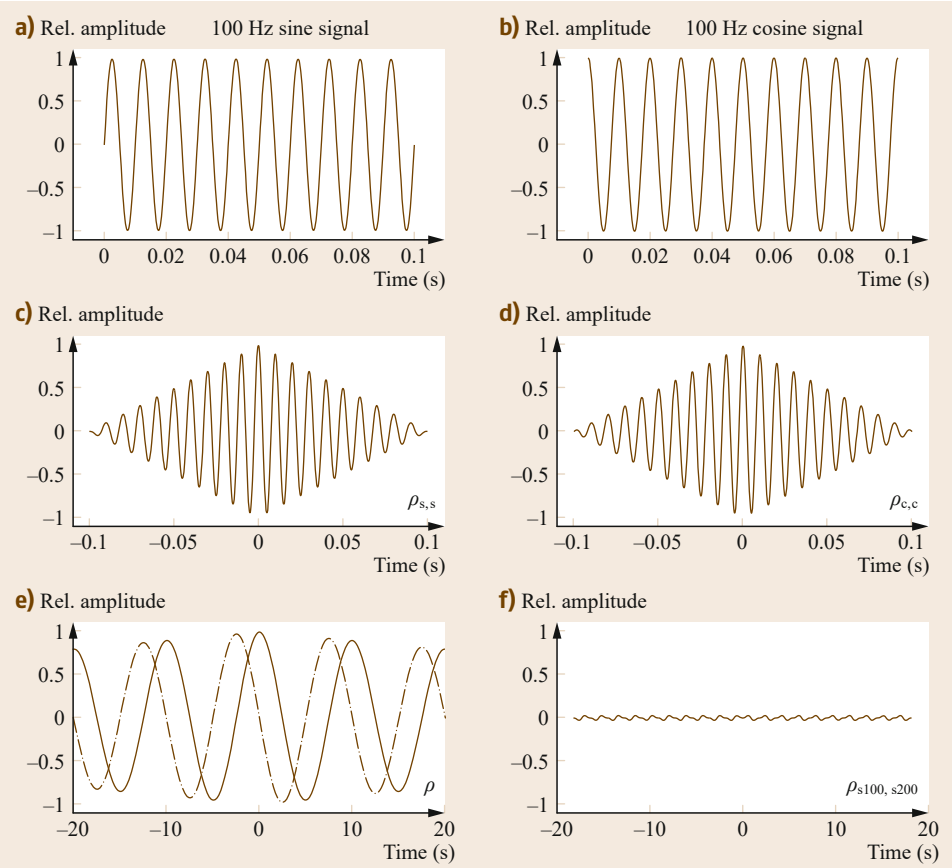


Fig. 14.8a–f Cross-correlation of two sinusoidal signals. The *top row* shows two 100 Hz signals (**(a)** sine signal; **(b)** cosine signal) that only differ by a 90° phase shift. The *center row* shows the autocorrelation function of both signals (**(c)** sine signal; **(d)** cosine signal), which turn out to be identical. Note that both functions have a peak value of 1 at an internal delay τ of 0 ms. **(e)** depicts the cross-correlation function between the sine and cosine signal (*dotted line*) versus the autocorrelation function of the sine signal. **(f)** shows the cross-correlation function of two sine signals with different frequencies (100 and 200 Hz)

For the normalized cross-correlation function, the possible range of $\rho(\tau)$ is restricted to values between -1 (fully inversely correlated) and 1 (fully correlated). A maximum value of 0 for $\rho(\tau)$ indicates that the two signals are fully uncorrelated. Consequently, a correlation value of 1 means the signals are identical at the time-shift delay τ, a value of -1 means they are identical but out of phase, and a value of 0 means they have no overlap in area.

Next, we use the normalized cross-correlation function to reexamine the orthogonality properties of trigonometric functions. Figure 14.8a,b show 100 Hz sine and cosine signals with duration of 0.1 s each. Both functions only differ in their phase. The autocorrelation function is depicted directly below each signal. Not surprisingly, the two sinusoidal signals are most identical to themselves without a temporal shift ($\tau = 0$). The reason the functions roll off to either side is because the signals no longer fully overlap in time. If the signals had been continuous, their autocorrelation functions would not roll off with τ. The autocorrelation goes to zero once one of the signals has been shifted beyond its duration ($\tau > 0.1$ s or $\tau < -0.1$ s). The local maxima in the autocorrelation functions occur each time the comparison signal is shifted by an integer multiple of the period T.

Now let us compare the sine and cosine signals with each other using the normalized cross-correlation function. The dotted line in Fig. 14.8e shows the result. Now the maximum is at 2.5 ms rather than at 0 ms. Compared with the previously discussed autocorrelation functions, the whole function is shifted by

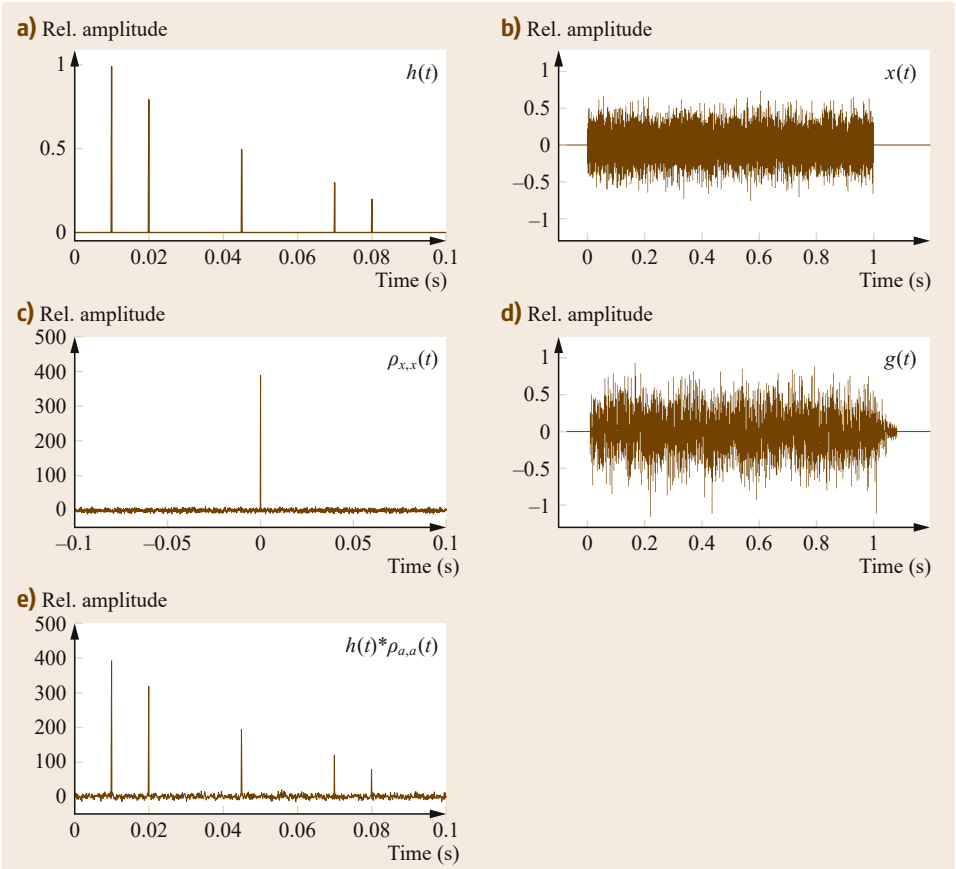


Fig. 14.9a–e Example of extracting an impulse response using the cross-correlation method. **(a)** shows an artificial impulse response with a direct sound at 0.01 s followed by four reflections. **(b)** shows the measurement signal, a 1 s Gaussian noise burst. The noise burst's autocorrelation function is basically a delta function with a peak at 0 ms and residual noise elsewhere **(c)**. **(d)** shows the measurement signal after it has been convolved with the impulse response. **(e)** shows the impulse response after it has been extracted from the convolved signals by means of cross-correlation with the original, nonconvolved measurement signal

2.5 ms, which is a quarter of the period $T = 10$ ms and corresponds to a phase shift of $\pi/2$ or 90°. Without the shift of the cosine function ($\tau = 0$ ms), the integrated product of the sine and cosine signals is 0, which explains why both signals were orthogonal to each other in Fourier analysis. However, if we consider sine signals with different frequencies (100 and 200 Hz), shifting one of the functions no longer works to obtain a significant cross-product. In fact, the latter would always remain zero if both functions were continuous.

14.3.1 Example: Extracting a Convolved Impulse Response

The main aim of this section is to understand the relationship between convolution and cross-correlation. *Stan* et al. [14.14] provide a detailed overview on standard methods to measure the impulse response of rooms and acoustic systems.

We end this chapter with a practical example of measuring an impulse response with a combined convolution/cross-correlation approach to shed more light on the properties of both methods. Let us assume that we would like to measure the hypothetical room impulse response shown in Fig. 14.9a. One way of measuring a room impulse response is to place a microphone at one position in the room as the receiver and then to excite the room with an impulse sound source (eg, a balloon pop or starting pistol) at a second position (Fig. 14.1). An alternative method that leads to better signal-to-noise ratios is to use a loudspeaker to play back a known, continuous signal such as the broadband noise signal $x(t)$ shown in Fig. 14.9a. A Gaussian noise signal has the advantage of having a delta function as its autocorrelation function for a continuous signal. Also, its power spectrum is constant, implying equal probabilistic energy at all frequencies (hence why its autocorrelation is a Dirac delta). This feature will become important for our method later on. The autocorrelation function of a Gaussian signal with restricted duration contains residual noise beside the delta peak, but is similar to the continuous signal case otherwise (Fig. 14.9c). When we record this signal coming from the loudspeaker position with the microphone at the receiver position, we obtain the source signal convolved

with the room impulse response

$$g(t) = \int_{-\infty}^{+\infty} x(\tau)h(t-\tau)\mathrm{d}\tau = \int_{-\infty}^{+\infty} h(\tau)x(t-\tau)\mathrm{d}\tau \,, \tag{14.48}$$

as shown in Fig. 14.9d. Now we cross-correlate the convolved signal with the original noise burst x that was played back from the loudspeaker

$$\rho_{x,g}(k) = \int_{-\infty}^{+\infty} g(n+k)x(n)\mathrm{d}n \,, \tag{14.49}$$

with the variables n and k also representing time, and insert the equation for $g(t)$ into this equation to substitute $g(t)$

$$\rho_{x,g}(k) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(\tau)x(n+k-\tau)\mathrm{d}\tau x(n)\mathrm{d}n \,. \tag{14.50}$$

We can now move $x(n)$ into the inner integral, because it does not depend on τ

$$\rho_{x,g}(k) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(\tau)x(n+k-\tau)x(n)\mathrm{d}\tau\mathrm{d}n \,. \tag{14.51}$$

Next, we will substitute $m = k-\tau$, which does not change the infinite boundaries of the integral. In the same step we also change the order of the integrals

$$\rho_{x,g}(k) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(k-m)x(n+m)x(n)\mathrm{d}n\mathrm{d}m \,. \tag{14.52}$$

Since $h(k-m)$ does not depend on n, we can move the variable out of the inner integral

$$\rho_{x,g}(k) = \int_{-\infty}^{+\infty} h(k-m)\int_{-\infty}^{+\infty} x(n+m)x(n)\mathrm{d}n\mathrm{d}m \,. \tag{14.53}$$

Now, it becomes clear that the inner integral is the autocorrelation function of our excitation signal x and we can replace it with the symbol for autocorrelation $\rho_{x,x}$

$$\rho_{x,g}(k) = \int_{-\infty}^{+\infty} h(k-m)\rho_{x,x}(m)\mathrm{d}m \,. \tag{14.54}$$

Our result shows that the cross-correlation between the convolved signal $g(t)$ and the excitation signal $x(t)$ results in the convolution of the autocorrelation function for the excitation signal x with the impulse response h. In our special case, the autocorrelation function of the excitation signal was

$$\rho_{x,x}(m) = \delta(m) \,,$$

which we can insert into our equation

$$\rho_{x,g}(k) = \int_{-\infty}^{+\infty} h(k-m)\delta(m)\mathrm{d}m \,. \tag{14.55}$$

We integrate over the delta function the same way we have done before to blend out $h(k)$ and obtain

$$\rho_{x,g}(k) = h(k) \,. \tag{14.56}$$

This result enables us to measure a room impulse response using a continuous signal and at the same time avoid potentially unstable methods such as deconvolution or division of two frequency functions. Figure 14.8e shows the practical result of this method for a 1 s Gaussian noise signal as excitation signal. The method allows us to extract the impulse response, but we have additional, residual noise between the reflections. This noise can be reduced by taking a longer excitation signal, and in practical measurements signals in the order of 1 min are often used.

Acknowledgments. I would like to thank my mentor Jens Blauert for teaching me the fundamentals of auditory signal processing. I learned the unique derivation of the convolution presented in this chapter when taking Jens' communication acoustics class in 1996. Rolf Walter at the University of Dortmund introduced me to the fundamentals of Fourier analysis and synthesis. Torben M. Pastore and Nikhil Deshpande were of great help proof-reading the manuscript.

References

14.1 J. d'Alembert: *Recherches sur Differens Points Importans du Systême du Monde (Researches on Different Important Points of the System of the World)*, Vol. 1. (Chez David l'aîné, Paris 1754)

14.2 P.-S. Laplace: *Mémoires de l'Académie Royale des Sciences de Paris in 1781* (Gautrier-Villars et Fils, Paris 1781)

14.3 A. Hald: *History of Mathematical Statistics from 1750 to 1930* (Wiley, New York 1998)

14.4 A. Dominguez: A history of the convolution operation (Retrospectroscope), IEEE Pulse **6**(1), 38–49 (2015)

14.5 G. Doetsch: Die Integrodifferentialgleichungen vom Faltungstypus, Math. Ann. **89**, 192–207 (1923)

14.6 P.S. Single, D.S. McGrath: Implementation of a 32768-Tap FIR filter using real-time fast convolution. In: *Proc. 87th Convent. Audio Eng. Soc.* (1989), Preprint 2830

14.7 J. Fourier: *Théorie Analytique de la Chaleur (The Analytic Theory of Heat)* (Firmin-Didot père et fils, Paris 1822)

14.8 A.V. Oppenheim, R.W. Schafer: *Digital Signal Processing* (Prentice Hall, Englewood Cliffs, New York 1975)

14.9 A. Bravais: Analyse mathématique sur les probabilités des erreurs de situation d'un point (Mathematical analysis of the probabilities of errors in a point's location), Mém. present. divers savants Acad. Sci. Inst. Fr., Sci. Math. Phys. **9**, 255–332 (1846)

14.10 F. Galton: Co-relations and their measurement, chiefly from anthropometric data, Proc. R. Soc. Lond. **45**, 135–145 (1888)

14.11 K. Pearson: Regression, heredity, and panmixia, Philos. Trans. R. Soc. Lond. **187**, 253–318 (1896)

14.12 J.M. Stanton: Galton, Pearson, and the Peas: A brief history of linear regression for statistics instructors, J. Stat. Educ. **9**(3) (2001), online, last accessed 02/10/2016

14.13 N. Wiener: Generalized harmonic analysis, Acta Math **55**(1), 117–258 (1930)

14.14 G.-B. Stan, J.-J. Embrechts, D. Archambeau: Comparison of different impulse response measurement techniques, J. Audio Eng. Soc. **50**(4), 249–262 (2002)

15. Audio Source Separation in a Musical Context

Bryan Pardo, Zafar Rafii, Zhiyao Duan

When musical instruments are recorded in isolation, modern editing and mixing tools allow correction of small errors without requiring a group to re-record an entire passage. Isolated recording also allows rebalancing of levels between musicians without re-recording and application of audio effects to individual instruments. Many of these techniques require (nearly) isolated instrumental recordings to work. Unfortunately, there are many recording situations (e.g., a stereo recording of a 10-piece ensemble) where there are many more instruments than there are microphones, making many editing or remixing tasks difficult or impossible.

Audio source separation is the process of extracting individual sound sources (e.g., a single flute) from a mixture of sounds (e.g., a recording of a concert band using a single microphone). Effective source separation would allow application of editing and remixing techniques to existing recordings with multiple instruments on a single track.

In this chapter we will focus on a pair of source separation approaches designed to work with music audio. The first seeks the repeated elements in the musical scene and separates the repeating from the nonrepeating. The second looks for melodic elements, pitch tracking and streaming the audio into separate elements. Finally, we consider informing source separation with information from the musical score.

One of the great advances of the 20th and 21st centuries has been the introduction and refinement of audio recording and the audio editing techniques recording allows. Remixing, editing, and remastering in the studio has enabled the creation of new genres of music (e.g., *musique concrete*, modern hip hop). When instruments are recorded in isolation, modern editing and mixing tools allow correction of small errors in music recordings without requiring a group to re-record an entire passage (e.g., a single missed note in the flute part). This also allows rebalancing of levels between musicians without re-recording (e.g., increasing the volume of the flute compared to the trumpet) and application of audio effects to individual instruments (e.g., adding reverberation to the vocals, but not the bass).

Many of these editing techniques require (nearly) isolated instrumental recordings. One cannot edit out a missed note in the flute, if that note is also recorded on the adjacent microphone for the violin. Unfortunately, there are many recording situations (e.g., a stereo recording of a 10-piece ensemble) where there are many more instruments than there are microphones. Thus, instruments are not recorded in isolation. Also, many legacy recordings are only available in the final stereo (two-channel) or mono (one-channel) mixture and the original premixed tracks are not available. All of these situations make many editing or remixing tasks difficult or impossible.

Audio source separation is the process of extracting individual sound sources (e.g., a single flute) from

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_15

a mixture of sounds (e.g., a recording of a concert band using a single microphone). Effective source separation would allow application of editing and remixing techniques to existing recordings with multiple instruments on a single track.

There have been many source separation approaches developed for general audio signals. Examples of general source separation algorithms include independent component analysis (ICA) [15.1], nonnegative matrix factorization (NMF) [15.2], nonnegative tensor factorization (NTF) [15.3], probabilistic latent component analysis (PLCA) [15.4], and robust principal component analysis (RPCA) [15.5]. While all of these have been applied to music, most do not leverage any particular features of music to aid in the separation process.

One element of a music scene which is highly salient is the repeating structure found in the music. *Schenker* asserted that repetition is what gives rise to the concept of the motive [15.6]. *Ruwet* used repetition as a criterion for dividing music into small parts, revealing the syntax of the musical piece [15.7]. *Ockelford* argued that repetition and imitation is what brings order to music [15.8].

Computational audio researchers have found repetitive elements in music audio useful for many purposes. Common applications include music summarization, e.g., see *Bartsch* [15.9], *Cooper* and *Foote* [15.10] and *Peeters* [15.11], audio segmentation [15.12], beat estimation [15.13], finding drum patterns in the audio [15.14], and structural analysis [15.15]. For a thorough review on music structure analysis, the reader is referred to [15.11, 16, 17].

The idea that repetition can be used for source separation is supported by recent findings in psychoacoustics. *McDermott* et al. established that the human auditory system is able to segregate individual sources by identifying them as repeating patterns embedded in the acoustic input, without requiring prior knowledge of the source properties [15.18]. Given this, could repetition in music be used as a basis for automatically separating the audio into a repeating background (e.g., a salsa montuno) and a varied foreground (the lead *salsero* sala singer)?

Another salient feature of a musical scene is melody. *Bregman* [15.19] shows that humans often link together distinct sound elements (e.g., notes) in time to produce perceptually salient entities called auditory streams. These streams often correlate strongly with melodies. Many trained musicians are able to segregate several melodies into independent elements they can attend to. Can a machine do the same? Often, a musical score can aid the trained musician to perform this task better. Can a score provide similar aid to an auditory streaming algorithm?

In this chapter we will focus on a pair of source separation approaches designed to work with music audio. The first seeks the repeated elements in the musical scene and separates the repeating from the nonrepeating. The second looks for melodic elements, pitch tracking and streaming the audio into separate elements. This second approach is then informed by a musical score to improve performance.

15.1 REPET

In this work, we begin with the observation that passages in many kinds of folk and pop music can be understood as a background component that is generally repeating in time, with a superimposed foreground component that is generally variable in time (e.g., a repeating accompaniment superimposed with varying vocals or a solo instrument). On this basis the *repeating pattern extraction technique* (REPET) was proposed. REPET is an intuitive approach for separating the repeating background from the nonrepeating foreground in an audio mixture. The basic idea is to identify repeating elements in the mixture by measuring self-similarity along time, derive repeating models by averaging the repeating elements over their repetition rates, and extract the repeating structure by comparing the repeating models to the mixture.

A number of experiments have shown that REPET can be effectively applied to separate pop songs into their music accompaniment and singing voice. Unlike other approaches for source separation, REPET does not depend on special parametrizations, does not rely on complex frameworks, and does not require external information. Because it is only based on repetition, it has the advantage of being simple, fast, blind, and therefore completely and easily automatable. More information about REPET, including source code, audio examples, and related publications can be found at [15.20].

15.1.1 Original REPET

The original REPET algorithm was designed to separate the repeating background from the nonrepeating

foreground in an audio mixture (e.g., the music accompaniment from the singing voice in a pop song) by identifying a period and modeling a segment for the periodically repeating patterns [15.21, 22].

The method can be summarized in three stages (Fig. 15.1):

- Identification of a repeating period
- Modeling of a repeating segment
- Extraction of the repeating structure.

Repeating Period Identification

In the first stage, the time-domain signal (top left of Fig. 15.1) is transformed into a time–frequency representation known as the spectrogram by using the short-time Fourier transform (STFT). A spectrogram represents sound as a two-dimensional structure, where the vertical axis shows frequency from low (at the bottom) to high. The horizontal axis shows time, from left to right. The spectrograms in Fig. 15.1 use red to indicate high energy and blue to show low energy.

The autocorrelation of each frequency channel is then computed. An autocorrelation measures the similarity of a signal with a delayed version of itself given different delays. The peak of the autocorrelation function indicates the period at which the sound repeats. To identify periodicity in the mixture, the beat spectrum [15.13] is derived from the spectrogram by computing the autocorrelation over time for every frequency channel and averaging the autocorrelations over the frequency channels. This is shown in the upper right of Fig. 15.1. Here, a peak is a candidate period at which the audio may repeat.

If periodically repeating patterns are present in the mixture, the beat spectrum forms peaks that are periodically repeating at different period rates, unveiling the underlying periodically repeating structure of the mixture, as shown in Fig. 15.1. A best-fit repeating period can then be identified from the beat spectrum, manually or by using an automatic period finder [15.21, 22].

Repeating Segment Modeling

In the second stage, the repeating period is used to find the period of the underlying repeating structure in the recording. This typically correlates to a pattern a few seconds long, such as a 4-chord repeating riff in a folk song. A model of what is repeating in the music is built by segmenting the audio at the points where the pattern repeats (middle panel of Fig. 15.1). The median value of the sound across all repetitions is then used to model the canonical repeating sound, as shown in Fig. 15.1.

This approach assumes the nonrepeating foreground (e.g., vocals or an instrumental soloist) has a sparse and varied time–frequency representation compared with the time–frequency representation of the

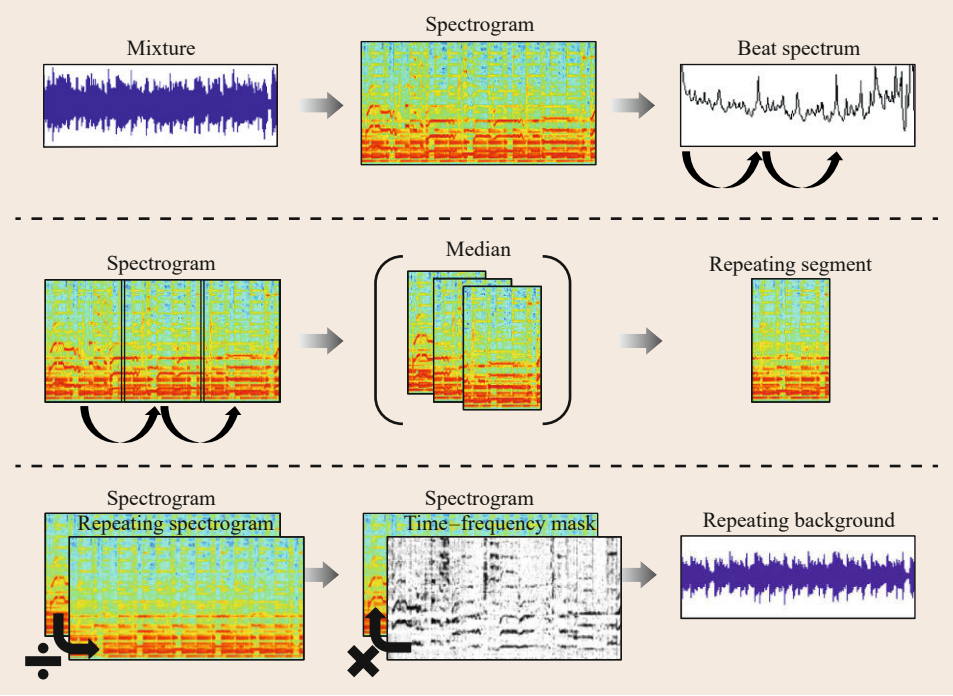


Fig. 15.1 Overview of the original REPET: (1) computation of the beat spectrum and identification of a repeating period (*top row*); (2) filtering of the spectrogram and modeling of a repeating segment (*middle row*); (3) derivation of the time–frequency mask and extraction of the repeating background (*bottom row*)

repeating background. Therefore, time–frequency bins with small deviations at their repetition rate would most likely represent repeating elements and would be captured by the median model. Conversely, time–frequency bins with large deviations at their repetition rate would most likely be nonrepeating elements (i. e., the nonrepeating musical foreground) and would be removed by the median model.

Repeating Structure Extraction

In the third stage, the repeating segment is used to derive a repeating spectrogram by taking, for every time–frequency bin, the minimum between the repeating model and the mixture spectrogram at period rate. This assumes the mixture spectrogram is the sum of a nonnegative repeating spectrogram and a nonnegative nonrepeating spectrogram. Thus, the mixture at any given time and frequency is assumed to always be as loud or louder than the individual mixture components (i. e., the repeating components are assumed not to cancel out the nonrepeating components).

The repeating spectrogram is then used to derive a time–frequency mask by dividing, for every time–frequency bin, the repeating spectrogram by the mixture spectrogram, as shown in Fig. 15.1. The rationale is that time–frequency bins that are likely to repeat at their repetition rate in the mixture spectrogram would have values near one in the time–frequency mask and would be weighted toward the repeating background. Conversely, time–frequency bins that are not likely to repeat at their repetition rate in the mixture spectrogram would have values near zero in the time–frequency mask and would be weighted toward the nonrepeating foreground.

The repeating background can then be obtained by multiplying, for every time–frequency bin, the time–frequency mask with the mixture. The nonrepeating foreground can be obtained by simply subtracting the repeating background from the mixture.

Experiments showed that REPET can be effectively applied to separate pop/rock song clips into their accompaniment music and singing voice [15.21, 22]. They also showed that REPET can be combined with other methods to improve background–foreground separation; for example, it can be used as a preprocessor to pitch detection algorithms to improve melody extraction [15.22], or as a postprocessor to a singing voice separation algorithm to improve music–voice separation [15.23]. Experiments further showed that REPET can be effectively applied to [separate the] accompaniment music and singing voice of separate full-track real-world songs, by simply applying the method along time via a sliding window [15.22]. There is, however, a trade-off for the window size in REPET: if the window is too long, the repetitions will not be sufficiently stable; if the window is too short, there will not be sufficient repetitions [15.22].

15.1.2 Adaptive REPET

Adaptive REPET is an extension of the original REPET and is designed to handle a varying periodic background, i. e., when the repeating period and/or the repeating patterns change over time (e.g., the succession of two homogenous sections in a song, such as a verse followed by the chorus), without the need for segmenting or windowing the audio [15.24].

As with the original REPET, the method can be summarized in three stages (Fig. 15.2):

- Identification of the repeating periods
- Modeling of a repeating spectrogram
- Extraction of the repeating structure.

Identification of Repeating Periods

In the first stage, the signal is transformed into a spectrogram. To identify local periodicities in the mixture, the beat spectrogram [15.13] is derived from the spectrogram by computing a beat spectrum for every time frame by sliding a window along time. In other words, each column in the beat spectrogram represents a beat spectrum at a given time.

If periodically repeating patterns are present in the mixture, the beat spectrogram forms horizontal lines that are periodically repeating vertically, corresponding to the succession of peaks in the concatenated beat spectra, unveiling the underlying periodically repeating structure of the mixture, as shown in Fig. 15.2. If variations of periodicity happen over time in the mixture, the horizontal lines in the beat spectrogram will show variations in their vertical periodicity. The repeating periods can then be identified from the beat spectrogram, for all the time frames in the mixture spectrogram, manually or by using an automatic period finder [15.24].

Repeating Spectrogram Modeling

In the second stage, the repeating periods are used to model an initial repeating spectrogram by taking, for every time frame in the mixture spectrogram, the median of the time–frequency bins at their period rate, as shown in Fig. 15.2.

The original REPET assumes that there is a single large period (e.g., the length of a measure) for all the repeating elements in the audio. The adaptive REPET considers each time frame individually (a single frame lasts roughly 20–40 ms) and tries to find similar frames separated from the current frame by some period (e.g., 2 s). If similar frames are found at this period (e.g.,

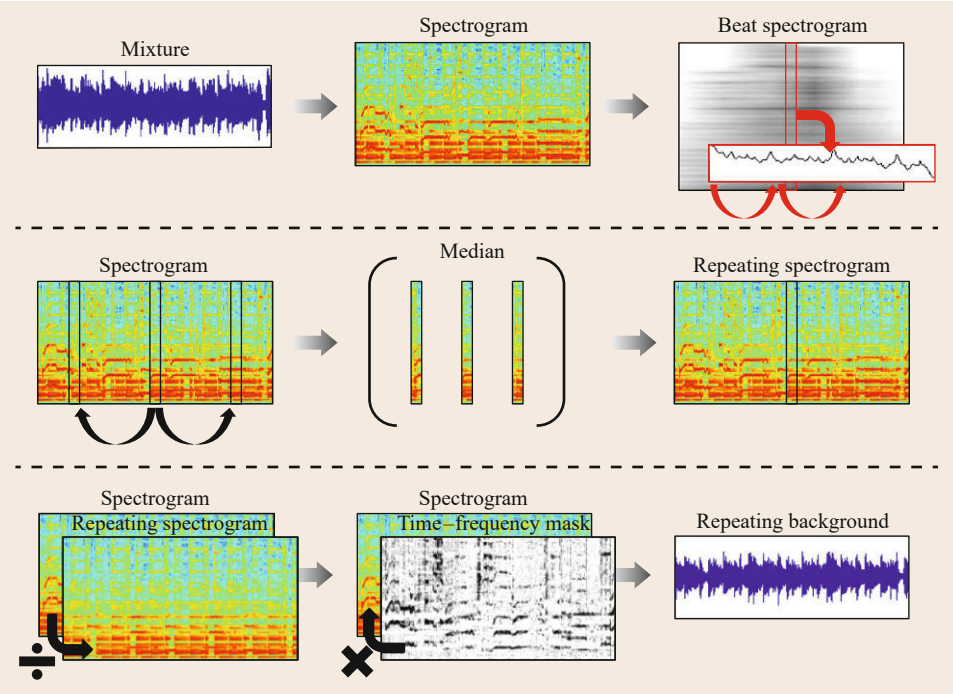


Fig. 15.2 Overview of adaptive REPET: (1) computation of the beat spectrogram and identification of the repeating periods (*top row*); (2) filtering of the spectrogram and modeling of a repeating spectrogram (*middle row*); (3) derivation of the time–frequency mask and extraction of the repeating background (*bottom row*)

similar frames at −4 s, −2 s, +2 s), the repeating spectrogram model for the current frame is built from those frames. Once this is done, it moves forward to the next time frame and tries to find a period at which the content of the new frame repeats. This period may or may not be the same as the previous time frame (e.g., a period of 1.9 s instead of 2 s). This lets it handle periodically repeating structures that vary slowly over time and also structures that are composed of interleaved repeating patterns of different periods (e.g., minimalist music).

Repeating Structure Extraction

In the third stage, the initial repeating spectrogram is used to derive a refined repeating spectrogram by taking, for every time–frequency bin, the minimum between the initial repeating spectrogram and the mixture spectrogram. The repeating spectrogram is then used to derive a time–frequency mask by dividing, for every time–frequency bin, the repeating spectrogram by the mixture spectrogram, as shown in Fig. 15.2.

The repeating background can then be obtained by multiplying, for every time–frequency bin, the time–frequency mask with the STFT of the mixture and transforming the result back to the time-domain. The nonrepeating foreground can be obtained by simply subtracting the repeating background from the mixture.

Experiments showed that adaptive REPET can be effectively applied to separate full-track real-world songs (e.g., a whole studio recording) into their accompaniment music and singing voice, unlike the original REPET which would only be meaningful for short excerpts [15.24].

15.1.3 REPET-SIM

REPET-SIM is a generalization of the REPET approach that was designed to handle nonperiodically repeating structures, i. e., when the repeating patterns happen intermittently or without a clear periodicity (e.g., repeated piano stabs in a jazz combo that use the same chord voicing, but whose rhythm varies). This is done by using a similarity matrix to identify the repeating elements [15.25].

As with the previous two methods, the method can be summarized in three stages (Fig. 15.3):

- Identification of the repeating elements
- Modeling of a repeating spectrogram
- Extraction of the repeating structure.

Repeating Element Identification

In the first stage, the signal is transformed into a spectrogram. To identify similarities in the mixture, the similarity matrix [15.26] is derived from the spectrogram by computing the cosine similarity between any two pairs of time frames.

If repeating elements are present in the mixture, the similarity matrix would form regions of high and low

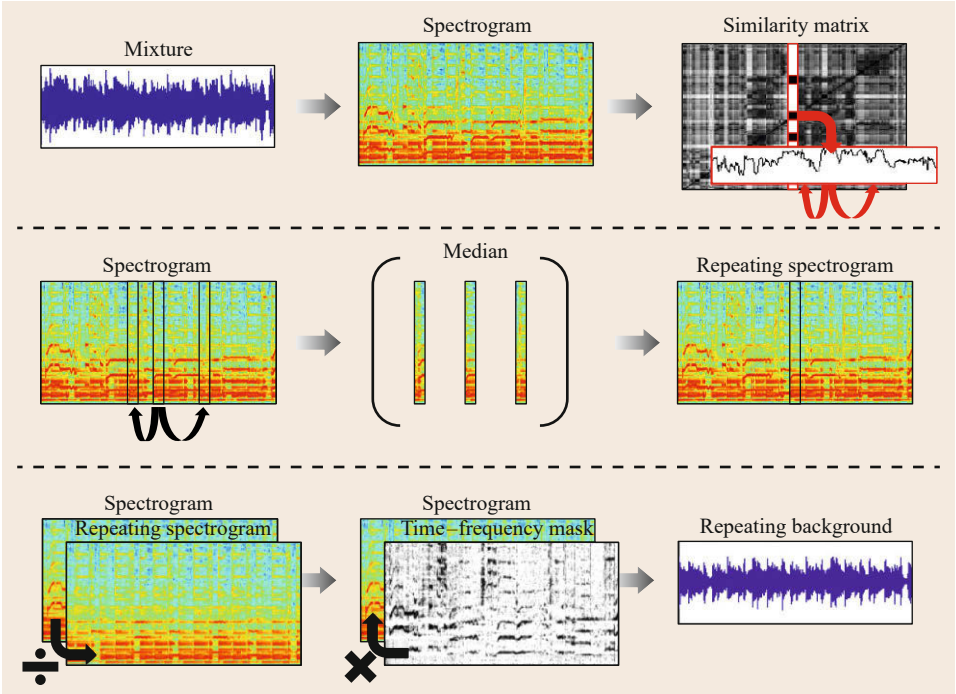


Fig. 15.3 Overview of REPET-SIM: (1) computation of the similarity matrix and identification of the repeating elements (*top row*); (2) filtering of the spectrogram and modeling of a repeating spectrogram (*middle row*); (3) derivation of the time–frequency mask and extraction of the repeating background (*bottom row*)

similarity at different time indices, unveiling the underlying repeating structure of the mixture, as shown in Fig. 15.3. The repeating elements can then be identified from the similarity matrix, for all the time frames in the mixture spectrogram, manually or by using an automatic peak finder [15.25].

Repeating Spectrogram Modeling

In the second stage, the repeating elements are used to model an initial repeating spectrogram by taking, for every time frame in the mixture spectrogram, the median of the time–frequency elements at their repetition rate, as shown in Fig. 15.3.

Compared with the other REPET methods (original and adaptive) that look for periodic similarity between events in the audio scene, REPET-SIM looks only for similarity, so that it can handle nonperiodically repeating structures, i. e., when the repeating patterns happen intermittently or without a clear periodicity.

Repeating Structure Extraction

In the third stage, the initial repeating spectrogram is used to derive a refined repeating spectrogram by taking, for every time–frequency bin, the minimum between the initial repeating spectrogram and the mixture spectrogram.

The repeating spectrogram is then used to derive a time–frequency mask by dividing, for every time–frequency bin, the repeating spectrogram by the mixture spectrogram, as shown in Fig. 15.3.

The repeating background can then be obtained by multiplying, for every time–frequency bin, the time–frequency mask with the STFT of the mixture and transforming the result back to the time-domain. The nonrepeating foreground can be obtained by simply subtracting the repeating background from the mixture.

Experiments showed that REPET-SIM can be effectively applied to separate full-track real-world songs into their accompaniment music and singing voice, also compared with the adaptive REPET [15.25]. Experiments also showed that REPET-SIM can be effectively applied to separate two-channel mixtures of one speech source and real-world background noise into their background noise and clean speech, by applying the method online for real-time computing [15.27].

Note that *FitzGerald* proposed a method very similar to REPET-SIM for music–voice separation [15.28].

15.2 Pitch-Based Source Separation

The previous sections focused on performing source separation using cues related to the repetitive, rhythmic structure of the music audio. Another fruitful approach is to use the melodic content as embodied by the pitches present in the audio. We now turn to this approach.

Pitched musical instruments (e.g., brass, woodwinds, strings, and keyboard instruments) are harmonic sound sources. Most of the energy in a harmonic sound is located at frequencies that are integer multiples of the fundamental frequency (F0). For example, if a piano plays a single note at A = 440 Hz, most of the energy in the sound will be concentrated at 440 Hz, 880 Hz, 1320 Hz, and so on. While F0 is a physical attribute and pitch is a perceptual attribute of a sound, for a harmonic sound, its pitch can be reliably matched to the F0. Therefore, we do not differentiate F0 from pitch in our discussions here.

Pitch information helps to separate harmonic sources from a mixture of sounds. In this section, we present our work on pitch-based source separation of audio mixtures composed of harmonic sound sources. Typically, this means separation of instruments from a music recording containing multiple concurrent instruments.

The first step is to estimate the pitches of these harmonic sources from the mixture [15.29]. This is called *multipitch estimation* (MPE). A frame is typically a 20–40 ms time window. MPE is typically performed in each individual time frame of the audio mixture. The second step is to connect the pitch estimates in different frames to form pitch trajectories, each of which corresponds to a source [15.30]. This is called *multipitch streaming*. The last step is to extract the harmonics from the pitch estimates of each source and reconstruct the source signal. In the following, we describe the three steps separately.

15.2.1 Multipitch Estimation

Multipitch estimation is the task of estimating pitches and the number of pitches in each frame of a harmonic sound mixture. In music information retrieval, the number of pitches is also called *polyphony*. Many methods have been proposed in the literature. Some do not employ any preprocessing of the signal and work with the full time domain or frequency domain signal [15.31–40]; others reduce the signal to a more compact representation such as employing an auditory filterbank as the front end [15.41–44] or representing the spectrum only with significant peaks [15.45–47]. Our method falls within the second category. More specifically, we represent the power spectrum of the audio mixture with both significant peaks and nonpeak regions and propose a maximum likelihood method to estimate the pitches and the polyphony from the peak–nonpeak-region representation [15.29].

Peak Detection

For harmonic sounds, significant spectral peaks ideally correspond to harmonics of the pitches. Therefore, detection of these peaks would help us infer the pitches. Taking one frame of the audio signal, we first perform a Fourier transform [15.48] to turn the audio into a spectral representation. The middle top panel of Fig. 15.1 shows a spectrogram. Here, each column of the image is the spectral representation of a single time step (typically a 20–40 ms window).

Spectral peaks at each time slice are detected by the peak detector described in [15.49]. Basically, there are two criteria that determine whether a local maximum in the spectrum should be labeled as a peak. The first criterion is global: the local maximum should not be less than some threshold (e.g., 50 dB) lower than the maximum of the spectrum across all frequencies in that time step. The second criterion is local: the local maximum should be locally higher than a smoothed version of the spectrum by at least some threshold (e.g., 4 dB). Finally, the peak amplitudes and frequencies are refined by quadratic interpolation [15.50].

We define the *nonpeak region* as those frequencies that are further than a quarter tone from any of the detected spectral peaks. Although the nonpeak region cannot tell us where the pitches should be, it can tell us where the pitches should not be. Pitches are not likely to be located at frequencies whose harmonics are in the nonpeak region.

Likelihood Function

We propose a maximum likelihood method to estimate the set of pitches from the peak–nonpeak representation of the mixture spectrum. The likelihood function is defined as the multiplication of the *peak-region likelihood* and the *nonpeak-region likelihood*, assuming conditional independence of peaks and the nonpeak region given the pitches. The peak region likelihood is defined as the probability of occurrence of the peaks, given an assumed set of pitches. The nonpeak region likelihood is defined as the probability of *not* observing peaks in the nonpeak region, given an assumed set of F0s. The peak region likelihood and the nonpeak region likelihood act as a complementary pair. The former helps find F0s that have harmonics that explain peaks, while the latter helps avoid F0s that would predict har-

monics where none are observed (i. e., in the nonpeak region).

To define the peak-region likelihood, we characterize each peak by its frequency and amplitude. We further consider the probability that each detected peak is a *normal* peak or a *spurious* peak. By *normal* we mean the peak is generated by some harmonic of the underlying pitches; and by *spurious* we mean the peak is due to other factors such as side-lobes, peak detection errors, etc.

We train the parameters of the likelihood functions using training data with ground-truth pitches and detected peaks and nonpeak regions. Two kinds of training data are used. The first kind is a set of isolated notes from the University of Iowa Musical Instrument Samples (MIS) dataset [15.51]. They are used to train the parameters of the nonpeak region likelihood model. The second kind is a set of randomly mixed chords using these isolated notes. They are used to train the parameters of the peak-region likelihood model. Ground-truth pitches are detected using the YIN [15.52] pitch tracking algorithm on isolated notes before mixing the isolated notes together. Peaks and the nonpeak region are detected using the proposed method.

Pitch and Polyphony Estimation

Given the set of detected peaks, frequencies within one semitone around each peak are considered as possible pitches (i. e., possible F0s). The underlying pitches in this frame are thus assumed to compose a subset of these frequencies. This helps to constrain the set of possible hypotheses to reasonable ones, given the data. The task of the maximum likelihood estimation is thus to find the subset of potential F0s that gives the highest likelihood to having generated the observed peaks in the spectrum.

It is, however, intractable to enumerate all these subsets. A typical scene may have 50–100 peaks. The set of all subsets of 100 peaks has 2^{100} elements and is not tractable to fully search. We therefore propose an iterative greedy search strategy [15.29]. We start from an empty subset to represent the estimated pitches. At each iteration, we add the one pitch estimate that most increases the likelihood of the subset. This strategy only enumerates a very small amount of subsets, as the choice of latter pitches depends on the choice of earlier pitches. However, the computational complexity is significantly reduced from exponential to linear with respect to the number of peaks.

An important question for this iterative greedy search process is *When should we stop*? In other words, how many pitches should we estimate in each time frame. Ideally, we hope that the likelihood function can take care of this problem and stop the process when the likelihood no longer increases on adding a new pitch. However, similar to many other maximum likelihood estimation problems, there is an overfitting problem. The likelihood typically increases with the number of pitches, although the increase becomes slower. We use a simple thresholding method to address this problem. We do not stop the iterative process until the number of pitches in the subset reaches a predefined maximally possible polyphony (e.g., 9). We then calculate the likelihood increase from 1 pitch to the maximally possible polyphony as the maximally possible likelihood increase. We then estimate the polyphony as the number of pitches that first surpasses 88% of the maximally possible increase. The threshold was tuned on a training set of musical chords with polyphony from 1–6. This simple polyphony estimation method is shown to work well on both musical chords and real music pieces.

Pitch Refinement

Pitches and polyphony estimated in individual frames may contain errors that make the pitch contours nonsmooth over time. By utilizing contextual information, it is possible to fix these errors and improve the pitch estimation accuracy.

We use a moving average to calculate the refined polyphony estimate with a triangular moving window with size of 19 frames. We also build a pitch histogram with a granularity of a semitone by counting the pitch estimates within the window and rank the pitches by their weighted counts according to the window. The top several pitches are returned as the refined pitches with equal weights, where the number is equal to the refined polyphony estimate. In experiments we show that this refinement step removes many insertion and deletion pitch estimation errors and significantly increases the estimation accuracy [15.29].

15.2.2 Multipitch Streaming

The pitches estimated in the previous section can help us separate harmonic sources in each individual frame. However, to separate the source signal over multiple frames, we need to connect these pitches into streams, each of which corresponds to a source. This is the multipitch streaming problem. Researchers have proposed using frequency–amplitude continuity to stream pitches [15.31, 36, 53, 54], but this approach only works within a note where the pitch does not change abruptly. For streaming pitches into noncontinuous pitch contours, we proposed the first method [15.30]. The basic idea is to use the timbre information. Notes performed by the same instrument have similar timbre compared to those performed by different instruments. Therefore,

we associate each pitch estimate with a timbre feature vector and perform clustering on the timbre feature vectors. Ideally, each cluster will correspond to one source and their pitches form the pitch stream of that source.

Timbre Features

We need to calculate a timbre feature vector for each pitch estimate in each frame and this feature should be calculated from the mixture signal directly. In our work we use two kinds of features. The first is called *harmonic structure* and is described in [15.49]. It is defined as a vector of relative logarithmic amplitudes of the harmonics of a pitch estimate. The harmonics are at integer multiples of the pitch. We use the first 50 harmonics to create a 50-dimensional timbre vector. We choose this dimensionality because most instruments have less than 50 prominent harmonics. For each harmonic, we use the peak-finder from [15.49] to see if there is a significant peak within a musical quarter-tone. If no peak is associated, the magnitude of the harmonic is set to 0 dB, otherwise it is set to the value of the nearest peak. Then, the representation is normalized. This feature has been shown to be similar for notes played by the same instrument within a narrow pitch range, while it is different for different instruments.

Another feature is called the *uniform discrete cepstrum* (UDC) [15.30]. It is calculated by taking the discrete cosine transform of a sparse log-amplitude magnitude spectrum where the nonzero elements are the harmonics of the pitch. This feature lies in the cepstral feature category and represents the spectral envelope of the harmonics of the target pitch. However, unlike other cepstral features such as the ordinary cepstrum or mel-frequency cepstral coefficients (MFCC), UDC can be calculated from the mixture spectrum directly for the target source, without resorting to source separation. UDC has been shown to outperform other cepstral representations and the harmonic structure feature in an instrument recognition task of musical chords [15.55].

Constrained Clustering

Given the feature vector of each pitch estimate, we perform K-means clustering on the pitches, to minimize the timbre inconsistency within each cluster. However, results show that there are two kinds of common errors. In Fig. 15.4b, a number of pitches are clustered into the wrong trajectory. For example, the pitches around Musical Instrument Digital Interface (MIDI) number 55 from 14.8 to 15.8 s form a continuous contour and are all played by the bassoon. However, in the clustering, some of them are assigned to the saxophone. In another example, from 16.8 to 17.6 s, the K-means

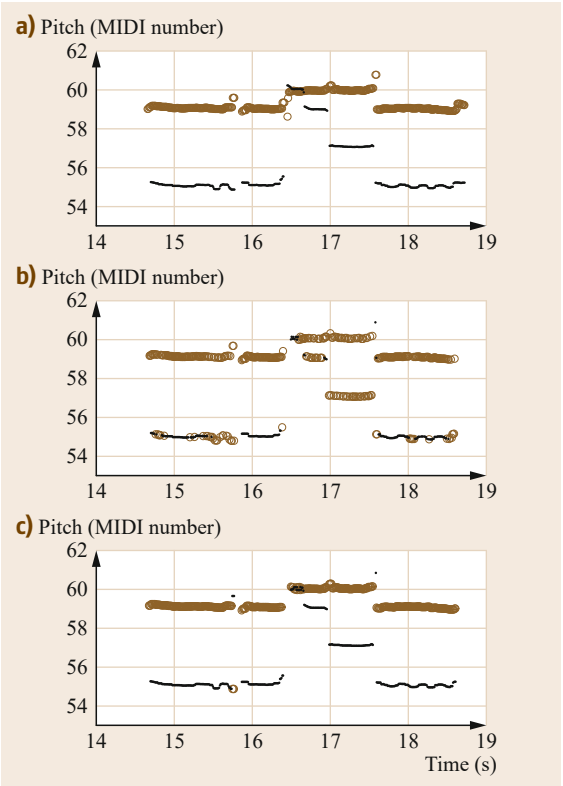


Fig. 15.4a–c Comparison of the ground-truth pitch streams **(a)**, K-means clustering ($K = 2$) results (i. e., only minimizing the objective function) **(b)** and the proposed method's results (i. e., considering both objective and constraints) **(c)**. Both the K-means and the proposed method take the ground-truth pitches as inputs, use 50 d harmonic structure as the timbre feature and randomly initialize their clusterings. Each point in these figures is a pitch. Different instruments are marked with different markers (*circles* for saxophone and *dots* for bassoon)

clustering places two simultaneous pitches into the saxophone stream. This is not reasonable as the saxophone is a monophonic instrument.

If we know that different sources do not often perform the same pitch at the same time and all sources are monophonic, we can impose two kinds of constraints on some pairs of the pitches to improve clustering: A *must-link* constraint is imposed between two pitches that differ less than Δ_t in time and Δ_f in frequency. It specifies that two pitches close in both time and frequency should be assigned to the same cluster. A *cannot-link* constraint is imposed between two pitches in the same frame. It specifies that two simultaneous pitches should be assigned to different clusters. These must-links and cannot-links form the set of all constraints. Figure 15.4c shows the result obtained from our pro-

posed algorithm, considering both the objective and constraints.

Iterative Algorithm

The formulated constrained clustering problem has two properties that mean existing constrained clustering algorithms [15.56–58] cannot be applied: 1) constraints are inconsistent with each other as they are imposed on pitch estimates that contain errors; 2) almost every pitch estimate is involved in some constraint. We therefore propose a new algorithm. It starts from an initial partition that satisfies a subset of all the constraints. Then it iteratively minimizes the objective function while incrementally satisfying more constraints. While the details of the algorithm can be found in [15.30], here we state its two important properties: 1) it always converges; 2) the timbre inconsistency objective function monotonically strictly decreases while the number of satisfied constraints monotonically increases.

15.2.3 Constructing Harmonic Masks

Given the estimated pitch stream for each harmonic source, we build a soft frequency mask around its harmonics to separate its magnitude spectrum from the mixture spectrum [15.59]. We then combine the separated magnitude spectrum with the phase spectrum of the mixture signal and perform an inverse Fourier transform to calculate the time-domain signal. Finally, the overlap-add technique [15.48] is applied to concatenate the current frame to previously separated frames.

To calculate the frequency masks for sources, we first identify their harmonics and overlapping situations from the estimated pitches. Each frequency bin is then classified into three kinds according to the number of harmonics that involve this frequency bin:

- Nonharmonic bin
- Nonoverlapping harmonic bin
- Overlapping harmonic bin.

For a nonharmonic bin, the masks are designed to evenly distribute the mixture energy to all active sources. For a nonoverlapping harmonic bin, the mixture energy is solely distributed to the source whose harmonic involves the bin.

The mask design for overlapping harmonic bins is the most difficult. One can calculate an average harmonic structure template for each source from the harmonic structures of its estimated pitches and then use the template to design the mask value at different harmonics [15.49]. This method considers the timbre model of the sources. A simpler method is to distribute the mixture energy to overlapping harmonics, in the inverse proportion to the square of the harmonic indices. This method does not model the timbre of sources, instead, it makes a general assumption that the harmonic amplitude decays at the same rate with respect to the harmonic index, regardless of pitch and instrument that produced the note. This is a very coarse assumption but it provides a simple and relatively effective way to resolve overlapping harmonics.

15.3 Leveraging the Musical Score

Previous sections described source separation algorithms that depend on repeating (rhythmic) content and pitch-based (melodic) content to perform separation. We now consider the case where one can improve a source separation algorithm by using information in addition to the audio recording. For many music pieces, music scores are widely available. In this scenario, score information can be leveraged to help analyze and separate the audio signal. More specifically, if the audio performance is faithful to the score and the audio and the score are aligned (i. e., synchronized), the score can tell us what pitches are likely being played at a time of the audio. This can greatly improve the accuracy of melodic, pitch-based source separation. We can use the score-provided pitch information to separate the harmonic sources in the audio. Such a system is called a score-informed source separation system.

In this chapter, we describe our work on an online score-informed source separation system called *Soundprism* [15.59]. Soundprism first aligns the audio with the score, then estimates the precise pitches based on the score-provided pitches in each frame and finally separates audio sources in the frame. All of these operations are performed in an online fashion, i. e., it processes the audio frame-by-frame in a serial fashion, without having the entire audio available. Figure 15.5 illustrates the system overview of Soundprism.

15.3.1 Audio–Score Alignment

The first stage of Soundprism is online audio-score alignment, also called *score following*. The score follower takes a piece of polyphonic music audio and its electronic score (e.g., MIDI) as input. It then outputs a score position for each time frame in a sequence. We

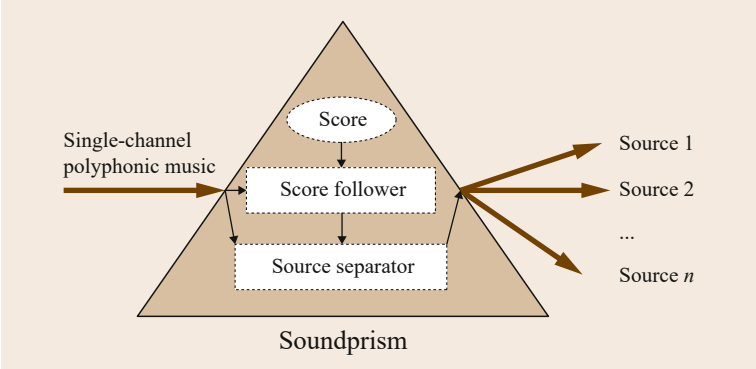


Fig. 15.5 System overview of Soundprism

propose a hidden Markov process model to do so, as illustrated in Fig. 15.6. The n-th time frame of the audio is associated with a 2-dimensional hidden state vector $\boldsymbol{s}_n = (x_n, v_n)^\top$, where x_n is its score position (in beats) and v_n is its tempo (in beats per minute [BPM]). Each audio frame is also associated with an observation variable $\boldsymbol{y}_n$, which represents the magnitude spectrum of the time frame. Our aim is to infer the current score position x_n from current and previous observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$. To do so, we need to define a process model to describe how the states transition, an observation model to evaluate hypothesized score positions for the current audio frame, and to find a way to do the inference in an online fashion.

Process Model

A process model defines the transition probability from the previous state to the current state, i. e., $p(\boldsymbol{s}_n|\boldsymbol{s}_{n-1})$. This tells us how the score position and tempo changes from one frame to another and the probability of the change. We use two dynamic equations to define this transition. While the detailed equations are ignored here, the basic idea is that the change of score position is determined by the tempo and the time interval between two adjacent frames. Also the tempo changes randomly around the previous tempo assuming a Gaussian distribution if the score position just passed a note onset or offset and does not change otherwise.

Note that we do not introduce randomness directly in score position. This is to avoid disruptive changes of score position estimates. In addition, randomness is only introduced when the score position has just passed a note onset or offset. This is because it is rather rare that the performer changes tempo within a note. Second, on the listener's side, it is impossible to detect the tempo change before hearing an onset or offset, even if the performer does make a change within a note. Therefore, changing the tempo state in the middle of a note is not in accordance with music performance, nor does it have evidence to estimate this change.

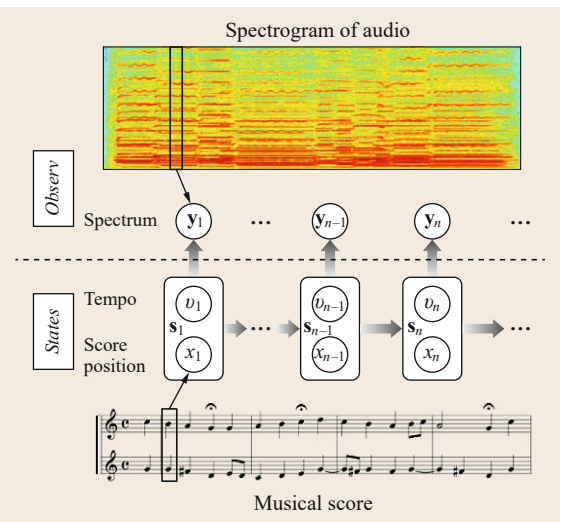


Fig. 15.6 Illustration of the state space model for online audio-score alignment

Observation Model

The observation model evaluates how well a hypothesized state (score position and tempo) can explain the observation, i. e., $p(\boldsymbol{y}_n|\boldsymbol{s}_n)$. In this work, we use the multipitch likelihood model as described in Sect. 15.2.1. The basic idea is that if the score pitches at the hypothesized score position fit well to the magnitude spectrum of the current frame $\boldsymbol{y}_n$, then the hypothesis is good. To calculate the multipitch likelihood, we just plug the score pitches into the likelihood function described in Sect. 15.2.1.

Clearly the observation model itself is not enough to estimate the hidden state (e.g., the score position), as the score may show the same pitches at different positions. In addition, the tempo dimension of the state does not play in the observation model. These problems, however, are addressed when the observation model and process model work together. The process model, con-

sidering the tempo dimension and the continuity of state changes, would only favor hypothesized states whose score position is close to the previous score position. This is the key idea of the hidden Markov process model.

Our observation model only considers information from the current frame and could be improved if considering information from multiple frames. *Ewert* et al. [15.60] incorporate interframe features to utilize note onset information and improve the alignment accuracy. *Joder* et al. [15.61] propose an observation model which uses observations from multiple frames for their conditional random field-based method. In the future we want to explore these directions to improve our score follower.

Inference

Given the process model and the observation model, we want to infer the state of the current frame from current and past observations. From a Bayesian point of view, this means we first estimate the posterior probability $p(\boldsymbol{s}_n|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_n)$, then decide its value using some criterion like maximum a posteriori (MAP) or minimum mean square error (MMSE). For hidden Markov processes, this posterior probability at the current frame can be updated from the previous frame. Therefore, we can estimate the posterior probability and the hidden states in an online fashion.

Here we use a *bootstrap filter*, one variant of particle filters [15.62, 63] to do the online update of the posterior probability. The process starts from an initialization of M particles. Their score positions are all set to the beginning of the score and their tempi are uniformly distributed between half and twice of the score-notated tempo. These particles represent an initialization of the posterior probability. To update the posterior probability when a new audio frame comes in, the particles are first moved using the process model, i. e., the score positions and tempi of the particles are changed. Then the observation likelihood of each particle is calculated using the observation model, which indicates the fitness of the particle to the current audio frame. The likelihood is set as the weight of the particle. These particles are then resampled with replacement according to their weights to generate a new set of M particles. Particles that do not fit to the current frame are less likely to remain in the new particle set, while particles that are a better fit to the current frame may retain multiple copies in the new particle set. A small random perturbation is imposed on these copies to prevent degeneracy of the particles. Now the new set of particles represent the new posterior probability. The average value of these particles is output as the estimate of the hidden state in the current frame.

The set of particles is not able to represent the distribution if there are too few and is time-consuming to update if there are too many. In our work we tried to use 100, 1000, and 10 000 particles. We find that with 100 particles, the score follower is often lost after a number of frames. But with 1000 particles, this rarely happens and the update is still fast enough. Therefore, 1000 particles are used in this chapter.

15.3.2 Pitch Refinement and Source Separation

Given the aligned score position for each audio frame, we know what instrument is playing what pitch in this frame from the score. This is important information for separating harmonic sources. However, the pitches provided by the score are integer MIDI pitch numbers. MIDI pitch numbers indicate keys on the piano keyboard. Typically, MIDI 69 indicates the A above Middle C. Assuming A440-based equal temperament allows translation from MIDI pitch to frequency in Hz. The resulting frequencies are rarely equal to the real pitches played in an audio performance. In order to extract the harmonics of each source in the audio mixture, we need to refine them to get accurate estimates of pitches played in the audio.

We refine the pitches using the multipitch estimation algorithm as described in Sect. 15.2.1, but restricting the search space within a semitone of the score-notated pitches. We also assume polyphony is given by the score.

Given the refined pitches in each audio frame, we use the same method as described in Sect. 15.2.3 to construct a harmonic mask for each pitch to separate the magnitude spectrum. Finally, we apply the inverse Fourier transform with the phase spectrum of the mixture signal and the overlap-add technique to reconstruct the separated source signals.

15.4 Conclusions

In this chapter we have outlined how to perform audio source separation on music using the cues of repeating musical structure and pitch content. We then showed how one can augment the pitch-based separation algorithm by leveraging the information in a musical score, where available. Moving forward,

we envision a combination of the score-informed pitch-based separation with separation based on rhythmic structure. Combining these should allow separation of musical instruments from an audio mixture in a large variety of contexts that are currently intractable.

References

15.1 P. Common, C. Jutten (Eds.): *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, 1st edn. (Academic, Oxford 2010)

15.2 T. Virtanen: Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria, IEEE Trans. Audio Speech Lang. Process. **15**(3), 1066–1074 (2007)

15.3 D. FitzGerald, M. Cranitch, E. Coyle: Non-negative tensor factorisation for sound source separation. In: *Irish Signals and Syst. Conf., Dublin* (2005)

15.4 P. Smaragdis, B. Raj, M.V.S. Shashanka: A probabilistic latent variable model for acoustic modeling. In: *NIPS Workshop Adv. Modeling Acoust. Process., Whistler* (2006)

15.5 P.-S. Huang, S.D. Chen, P. Smaragdis: Singing-voice separation from monaural recordings using robust principal component analysis. In: *37th Int. Conf. Acoustics, Speech and Signal Process., Kyoto* (2012)

15.6 H. Schenker: *Harmony*, Vol. 1 (Univ. Chicago Press, Chicago 1980)

15.7 N. Ruwet, M. Everist: Methods of analysis in musicology, Music Anal. **6**(1/2), 3–9 (1987)

15.8 A. Ockelford: *Repetition in Music: Theoretical and Metatheoretical Perspectives*, Royal Musical Association Monographs, Vol. 13, 2005)

15.9 M.A. Bartsch: To catch a chorus using chroma-based representations for audio thumbnailing. In: *IEEE Workshop Appl. Signal Process. Audio Acoust., New Paltz* (2001)

15.10 M. Cooper, J. Foote: Automatic music summarization via similarity analysis. In: *3rd Int. Conf. Music Inf. Retr., Paris* (2002)

15.11 G. Peeters: Deriving musical structures from signal analysis for music audio summary generation: Sequence and state approach, Comput. Music Modeling Retr. **2771**, 143–166 (2004)

15.12 J. Foote: Automatic audio segmentation using a measure of audio novelty. In: *IEEE Int. Conf. Multimedia and Expo, New York* (2000)

15.13 J. Foote, S. Uchihashi: The beat spectrum: A new approach to rhythm analysis. In: *IEEE Int. Conf. Multimedia and Expo, Tokyo* (2001)

15.14 K. Yoshii, M. Goto, H.G. Okuno: Drum sound identification for polyphonic music using template adaptation and matching methods. In: *ISCA Tutor. Res. Workshop on Stat. Percept. Audio Process., Jeju* (2004)

15.15 R.B. Dannenberg: Listening to Naima: An automated structural analysis of music from recorded audio. In: *Int. Comput. Music Conf., Gothenburg* (2002)

15.16 R.B. Dannenberg, M. Goto: Music structure analysis from acoustic signals, Handbook of Signal Process, Acoustics **1**, 305–331 (2009)

15.17 J. Paulus, M. Müller, A. Klapuri: Audio-based music structure analysis. In: *11th Int. Soc. Music Inf. Retr., Utrecht* (2010)

15.18 J.H. McDermott, D. Wrobleski, A.J. Oxenham: Recovering sound sources from embedded repetition, Proc. Nat. Acad. Sci. USA **108**(3), 1188–1193 (2011)

15.19 A. Bregman, C. Jutten: *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge 1994)

15.20 Interactive Audio Lab of Northwestern University: http://music.eecs.northwestern.edu/research.php?project=repet

15.21 Z. Rafii, B. Pardo: A simple music–voice separation system based on the extraction of the repeating musical structure. In: *36th Int. Conf. Acoust. Speech Signal Process., Prague* (2011)

15.22 Z. Rafii, B. Pardo: REpeating pattern extraction technique (REPET): A simple method for music–voice separation, IEEE Trans. Audio Speech Lang. Process. **21**(1), 71–82 (2013)

15.23 Z. Rafii, D.L. Sun, F.G. Germain, G.J. Mysore: Combining modeling of singing voice and background music for automatic separation of musical mixtures. In: *14th Int. Soc. Music Inf. Retr., Curitiba* (2013)

15.24 A. Liutkus, Z. Rafii, R. Badeau, B. Pardo, G. Richard: Adaptive filtering for music–voice separation exploiting the repeating musical structure. In: *37th Int. Conf. Acoustics, Speech and Signal Process., Kyoto* (2012)

15.25 Z. Rafii, B. Pardo: Music–voice separation using the similarity matrix. In: *13th Int. Soc. Music Inf. Retr., Porto* (2012)

15.26 J. Foote: Visualizing music and audio using self-similarity. In: *7th ACM Int. Conf. Multimedia, Orlando* (1999)

15.27 Z. Rafii, B. Pardo: Online REPET-SIM for real-time speech enhancement. In: *38th Int. Conf. Acoust. Speech and Signal Process., Vancouver* (2013)

15.28 D. FitzGerald: Vocal separation using nearest neighbours and median filtering. In: *23nd IET Irish Signals and Syst. Conf., Maynooth* (2012)

15.29 Z. Duan, B. Pardo, C. Zhang: Multiple fundamental frequency estimation by modeling spectral peaks and non-peak regions, IEEE Trans. Audio Speech Lang. Process. **18**(8), 2121–2133 (2010)

15.30 Z. Duan, J. Han, B. Pardo: Multi-pitch streaming of harmonic sound mixtures, IEEE Trans. Audio Speech Lang. Process. **22**(1), 1–13 (2014)

15.31 G.E. Poliner, D.P.W. Ellis: A discriminative model for polyphonic piano transcription, EURASIP J. Adv. Signal Process. **2007**, 48317-1–48317-9 (2007), https://doi.org/10.1155/2007/48317

15.32 M. Davy, S.J. Godsill, J. Idier: Bayesian analysis of polyphonic western tonal music, J. Acoustical Soc. Am. **119**, 2498–2517 (2006)

15.33 E. Vincent, M.D. Plumbley: Efficient Bayesian inference for harmonic models via adaptive posterior factorization, Neurocomputing **72**, 79–87 (2008)

15.34 K. Kashino, H. Murase: A sound source identification system for ensemble music based on template adaptation and music stream extraction, Speech Commun. **27**(3–4), 337–349 (1999)

15.35 M. Goto: A real-time music-scene-description system: Predominant-F0 estimation for detecting melody and bass lines in real-world audio signals, Speech Commun. **43**(4), 311–329 (2004)

15.36 H. Kameoka, T. Nishimoto, S. Sagayama: A multipitch analyzer based on harmonic temporal structured clustering, IEEE Trans. Audio Speech Lang. Process. **15**(3), 982–994 (2007)

15.37 S. Saito, H. Kameoka, K. Takahashi, T. Nishimoto, S. Sagayama: Specmurt analysis of polyphonic music signals, IEEE Trans. Speech Audio Process. **16**(3), 639–650 (2008)

15.38 J.-L. Durrieu, G. Richard, B. David: Singer melody extraction in polyphonic signals using source separation methods. In: *Proc. IEEE Int. Conf. Acoustics Speech Signal Process. (ICASSP)* (2008) pp. 169–172

15.39 V. Emiya, R. Badeau, B. David: Multipitch estimation of quasi-harmonic sounds in colored noise. In: *Proc. Int. Conf. Digital Audio Effects (DAFx)* (2007)

15.40 G. Reis, N. Fonseca, F. Ferndandez: Genetic algorithm approach to polyphonic music transcription. In: *Proc. IEEE Int. Symp. Intell. Signal Process* (2007)

15.41 T. Tolonen, M. Karjalainen: A computationally efficient multipitch analysis model, IEEE Trans. Speech Audio Process. **8**(6), 708–716 (2000)

15.42 A. de Cheveigné, H. Kawahara: Multiple period estimation and pitch perception model, Speech Commun. **27**, 175–185 (1999)

15.43 A. Klapuri: Multiple fundamental frequency estimation based on harmonicity and spectral smoothness, IEEE Trans. Speech Audio Process. **11**(6), 804–815 (2003)

15.44 A. Klapuri: Multiple fundamental frequency estimation by summing harmonic amplitudes. In: *Proc. ISMIR* (2006) pp. 216–221

15.45 R.J. Leistikow, H.D. Thornburg, J.S. Smith, J. Berger: Bayesian identification of closely-spaced chords from single-frame STFT peaks. In: *Proc. Int. Conf. Digital Audio Effects (DAFx'04), Naples* (2004) pp. 228–233

15.46 A. Pertusa, J.M. Inesta: Multiple fundamental frequency estimation using Gaussian smoothness. In: *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)* (2008) pp. 105–108

15.47 C. Yeh, A. Röbel, X. Rodet: Multiple fundamental frequency estimation of polyphonic music signals. In: *Proc. IEEE Int. Conf. Acoustics, Speech Signal Process. (ICASSP)* (2005) pp. 225–228

15.48 J.O. Smith: Spectral Audio Signal Processing, http://ccrma.stanford.edu/~jos/sasp/ (2014)

15.49 Z. Duan, Y. Zhang, C. Zhang, Z. Shi: Unsupervised single-channel music source separation by average harmonic structure modeling, IEEE Trans. Audio Speech Lang. Process. **16**(4), 766–778 (2008)

15.50 J.O. Smith, X. Serra: Parshl: An analysis-synthesis program for non-harmonic sounds based on a sinusoidal representation. In: *Proc. Int. Comput. Music Conf. (ICMC)* (1987)

15.51 L. Fritts, University of Iowa: http://theremin.music.uiowa.edu/MIS.html

15.52 A. de Cheveigné, H. Kawahara: YIN, a fundamental frequency estimator for speech and music, J. Acoustical Soc. Am. **111**, 1917–1930 (2002)

15.53 M. Ryynanen, A. Klapuri: Polyphonic music transcription using note event modeling. In: *Proc. IEEE Workshop on Appl. Signal Process. Audio Acoustics (WASPAA)* (2005) pp. 319–322

15.54 W.-C. Chang, A.W.Y. Su, C. Yeh, A. Robel, X. Rodet: Multiple-F0 tracking based on a high-order HMM model. In: *Proc. Int. Conf. Digital Audio Effects (DAFx)* (2008)

15.55 Z. Duan, B. Pardo, L. Daudet: A novel cepstral representation for timbre modeling of sound sources in polyphonic mixtures. In: *Proc. IEEE Int. Conf. Acoustics Speech Signal Process. (ICASSP)* (2014)

15.56 K. Wagstaff, C. Cardie: Clustering with instance-level constraints. In: *Proc. Int. Conf. Machine Learning (ICML)* (2000) pp. 1103–1110

15.57 K. Wagstaff, C. Cardie, S. Rogers, S. Schroedl: Constrained k-means clustering with background knowledge. In: *Proc. Int. Conf. Machine Learning (ICML)* (2001) pp. 577–584

15.58 I. Davidson, S.S. Ravi, M. Ester: Efficient incremental constrained clustering. In: *Proc. ACM Conf. Knowl. Discovery and Data Mining (KDD)* (2007) pp. 240–249

15.59 Z. Duan, B. Pardo: Soundprism: An online system for score-informed source separation of music audio, IEEE J. Selected Topics Signal Process. **5**(6), 1205–1215 (2011)

15.60 S. Ewert, M. Müller, P. Grosche: High resolution audio synchronization using chroma onset features. In: *Proc. IEEE Int. Conf. Acoustics Speech Signal Process. (ICASSP)* (2009) pp. 1869–1872

15.61 C. Joder, S. Essid, G. Richard: A conditional random field framework for robust and scalable audio-to-score matching, IEEE Trans. Audio Speech Lang. Process. **19**(8), 2385–2397 (2011)

15.62 A. Doucet, N. de Freitas, N.J. Gordon (Eds.): *Sequential Monte Carlo Methods in Practice* (Springer, New York 2001)

15.63 M.S. Arulampalam, S. Maskell, N. Gordon, T. Clapp: A tutorial on particle filters for online nonlinear-non-Gaussian Bayesian tracking, IEEE Trans. Signal Process. **50**(2), 174–188 (2002)

16. Automatic Score Extraction with Optical Music Recognition (OMR)

Ichiro Fujinaga, Andrew Hankinson, Laurent Pugin

Optical music recognition (OMR) describes the process of automatically transcribing music notation from a digital image. Although similar to optical character recognition (OCR), the process and procedures of OMR diverge due to the fundamental differences between text and music notation, such as the two-dimensional nature of the notation system and the overlay of music symbols on top of staff lines. The OMR process can be described as a sequence of steps, with techniques adapted from disciplines including image processing, machine learning, grammars, and notation encoding. The sequence and specific techniques used can differ depending on the condition of the image, the type of notation, and the desired output.

Several commercial and open-source OMR software systems have been available since the mid-1990s. Most of them are designed to be used by individuals and recognize common (post-18th-century) Western music notation, though there have been some efforts to recognize other types of music notation such as for the lute and for earlier Western music.

Even though traditional applications of OMR have focused on small-scale recognition tasks, typically as an automated method of musical entry for score editing, new applications of large-scale OMR are under development, where automated recognition is the central technology for building *full-music* search systems, similar to the large-scale full-text recognition efforts.

16.1 History

Computer-based optical recognition technologies, including both OMR and OCR, have been under development since the early days of computing technology. Computerized OCR was first developed in 1951 [16.1] and sold to large corporations, such as Reader's Digest and AT&T, to help process subscription and billing information [16.2].

Pruslin [16.3] demonstrated the first optical music recognition system. This system operated on a single measure of common Western music notation (CWMN), and was capable of recognizing a limited set of musical symbols. Several years later, *Prerau* [16.4] introduced the *DO-RE-MI* OMR system, capable of recognizing three measures of printed CWMN consisting of a single voice on two staves in a single font.

The goal of most early OMR development was a universal recognition system, capable of recognizing the entirety of music notation output, in much the same way that early OCR systems were envisioned as a tool capable of complete and accurate transcription of all textual documents [16.5, 6]. Despite understanding the limitations of the early OMR systems, Prerau concludes that the technique *should be able to be expanded to the recognition of all printed music* [16.7]. This early optimism was driven by a näive (in hindsight) understanding of the diversity and complexity of document contents. For text, page features such as

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_16

columns, figures, tables, footnotes, and even headings posed a challenge for accurate textual recognition. For music recognition, the troubles were even more acute since music notation styles and practices would vary by composer, publisher, repertoire and historical practices. By the 1990s, the goal of creating a universal recognition system had been largely abandoned in favor of repertoire and application-specific OMR systems capable of transcribing a well-defined subset of music documents [16.8]:

> [...] *in practice, composers and publishers often feel free to adapt old notation to new uses, and invent new notation, as they see fit. There are in fact national dialects of music notation, and musical works use many different levels of notational complexity. Thus it may not be possible to devise a single recognition system capable of recognizing all music notation.* [Pruslin 1966] *states that a complete solution to the music recognition problem is the specification of: which notes are present, what order they are played in, their time values or durations, and volume, tempo, and interpretation. This level of recognition suffices for only some of the applications listed* [later in this paper].

As research and development continued through the 1990s and 2000s, OMR systems were developed to specialize in transcribing specific repertoires or styles of notation. In addition to CWMN recognition systems, repertoire-specific recognition systems exist for many different music notation styles, including lute tablature [16.9–11], Byzantine chant [16.12, 13], mensural notation, both in print [16.14] and manuscript sources [16.15], and others.

Several OMR toolkits have also been developed to help assemble bespoke OMR systems. These toolkits have been used to build several of the aforementioned systems, and provide a generalized structure and toolset from which these customized OMR systems may be built. Examples of these frameworks include the CANTOR system [16.16] and the Gamera system [16.17]. While these systems present a more flexible approach to OMR, they require significantly more development expertise to create and run OMR than *turnkey* systems and, as such, are generally only used in research contexts.

16.2 Overview

The recognition process can generally be defined as identifying and contextualizing from a signal the information contained in it. It is a discriminative response to a specific stimulus that makes it possible to assign each object to a particular class. Recognizing shapes and reading are typical recognition processes performed by humans, and cognitive research has studied the complexity of this task. *Marr* [16.18], a pioneer in cognitive science, has proposed a theory of vision describing it as a bottom-up process, in which an image is first deconstructed into bidimensional primitives and then reconstructed as a spatial object. Since then, other studies [16.19] have shown that such a theory, however rational, is too restrictive and that a top-down activity is part of the human recognition processes, in particular when the subject does not easily understand the scene or the context. For example, when a human is reading a highly degraded document, the type of document and the context will be instrumental clues to deciphering its contents. For example, knowing that the document is a string quartet and reading the beginning of the bottom staff provides the reader with the information that the content to be read at that place is quite likely to be an F clef. Studies on human text reading have highlighted different reading techniques [16.20]. In most cases, the brain does not operate by letter decoding but rather by adopting a more global approach based on the linguistic context. This explains why reading known words within which letters are missing or are inverted can usually be achieved without problems. Conversely, when one has to read an unknown word, reading is performed via letter decoding and assembly.

The goal pursued in optical recognition of music scores and of documents is similar to the recognition process as described above. A major problem, however, is that it is extremely difficult to formalize this process algorithmically, especially if the aim is to design a system that can operate simultaneously with the bottom-up and top-down analysis mechanisms.

Among the various applications developed in the field of document recognition, the most advanced ones are certainly the optical character recognition (OCR) applications for which many functional and business solutions already exist. In fact, very early on, the results obtained in this field allowed complete solutions to be implemented for the recognition of typewritten texts or fonts, at least for fairly good documents having a relatively simple structure. In most systems, the recognition is performed by a sequential process of preprocessing, segmentation, classification and validation. This approach provides excellent results for two main reasons: firstly because character segmentation is

relatively easy and can be done simply by analyzing verticals projections, and secondly because classifiers such as neural networks (NN) or *k*-nearest-neighbor (*k*-NN) that are often used in these systems perform particularly well for this type of operation and on this type of data.

The results obtained in the recognition of cursive handwriting, or related tasks, are significantly less successful and it is unanimously agreed that this task is much more difficult [16.21]. In addition, it is striking to see how difficult it is, if not impossible, to adapt techniques traditionally used for recognition of printed texts to the recognition of handwritten text. In fact, the problem faced with handwriting recognition is a problem that can be summed up in a well-known paradox stated by *Sayre* [16.22]: *to recognize, we need to segment the input signal, but in order to segment appropriately we need to recognize*. The best results in handwriting recognition were obtained using other techniques than NN or *k*-NN that enable validating the recognition at different linguistic levels. The approach that has been the most widely used in recent years is hidden Markov models (HMM), especially because of their excellent noise absorption capacity and, above all, because of their linguistic context integration capacity [16.23].

16.3 OMR Challenges

OMR is a special case of document recognition. From a technical point of view, it has many similarities with OCR. Many of the challenges encountered in character recognition also arise with music. A recurring problem is the quality of documents considered. Some techniques can show promising results and yield high recognition rates when the documents considered are very clean, are straight, and show perfectly printed symbols. The difficulty of the task, however, increases dramatically with document degradation, if the document is skewed or curved, or if some symbols are poorly printed or partially deleted. In real-life cases, however, it is very rare that the documents to be processed are clean, straight, and perfectly printed. In most cases, we can expect imperfections of various types. They can be grouped into two distinct categories: the imperfections in the document itself and degradation introduced by the document acquisition phase.

Imperfections in the document can vary from one type of document to another, from one document to another, and sometimes even from one page to another. The following problems are commonly encountered:

- Printing imperfections (such as uneven absorption of ink by the paper)
- Partial erasure of symbols
- Ink bleed-through from the opposite side of the paper
- Stains
- Paper degradation (yellowing, foxing, mold and mildew etc.)
- Holes or tears.

An essential step in the recognition process is digitization, where the analog signal (an original document, a photo etc.) is converted into a digital image. This document acquisition phase is not always limited to a single capture event. In practice it is composed of several successive steps, and in the recognition stage the user may not have a clean and clear image of the original physical document but only a version that previously went through different imaging processes. In some cases for instance, the document is photographed before being digitized. A source may have been photographed and then transferred to a microfilm or microfiche, and ultimately only the microfilm is available to be digitized. Each step of the of document acquisition can introduce various types of unwanted artifacts:

- Document image skewing
- Noise
- Bulged or curved appearance of the document image
- Nonuniform lighting
- Borders around the document image
- Partial content of other pages around the document image.

An important consideration during this phase is the digitization resolution. The resolution has a direct effect on the size and definition of details that can be captured from the physical object. Musical notation contains specific features that depend on relatively small symbols, for example staccato, dotted notes and accidentals. *Ng* and *Boyle* [16.24] showed that with a document scanned at a resolution of 300 dpi, the distinction between a sharp and a natural could be ambiguous. Several studies [16.25, 26] have shown that as a result the appropriate resolution for capturing details from a music document is higher than that for text.

While OCR and OMR are similar tasks in many aspects, they also have fundamental differences that make OMR a more difficult task than OCR. The first difference concerns the placement of symbols on a page

and how they are arranged. Text is, for the most part, unidimensional. Characters and words are typically composed in a horizontal line, and each line has no vertical relationship with the parallel lines above or below it. This approach is not appropriate for music, because it is necessary in music to consider both the vertical and the horizontal dimensions simultaneously. The addition of the vertical dimension poses challenges in accurately determining the pitch of a note, for example, or the component pitches of a chord, while simultaneously information embedded in the horizontal dimension determines co-occurrences of both sounding and nonsounding events (i. e., notes that sound at the same time, or ties that control the sounding duration of a note).

Another difference is the musical characteristic that symbols are superimposed on staff lines. Superimposition of the elements is known to be a difficult problem of recognition of forms [16.27]. Many segmentation algorithms widely used in textual recognition are ineffective with music scores because they operate by edge detection. Thus, in OMR, the superposition of the symbols makes segmentation a particularly critical phase. Most musical recognition systems include the removal of the staff lines, which is particularly difficult when a musical symbol merges with a staff line, for example on top of an F clef. In practice, several factors can make staff line detection and removal difficult:

- The lines are usually not perfectly straight
- The line thickness is often variable
- The lines may be interrupted at certain points.

Many musical symbols have similar characteristics to each other (e.g., a half note and a quarter note are both made of a stem and a note head of the same size), but must be interpreted in vastly different ways. The visual difference between a note with a dotted duration and a note with a staccato is simply a slight shift in the position of the dot, but one means to lengthen the sounded note, and the other means to shorten it! This often leads to classification problems. Many musical symbols are not made of a single graphical element, but are *compound* symbols composed of two or more distinct elements. This is the case, for example, of an F clef, where the entire symbol is composed of three components: a curved line and two dots. Elements of compound symbols can belong to different types of symbols. A small dot can be the point of an F clef, a point of a dotted note, or a staccato point. Similarly, a sharp can be part of a key signature, the accidental of a note, or belong to a figured bass.

16.4 Technical Background

Several studies cover the early OMR techniques exhaustively, including *Blostein* and *Baird* [16.8], *Selfridge-Field* [16.28], *Bainbridge* and *Bell* [16.27], and more recently *Rebelo* et al. [16.21].

The nature of musical symbols has meant that from very early on OMR research had to look at structural recognition methods. The earliest research quickly showed that a functional recognition approach would work only for excessively simple cases. In general, symbols are grouped into two distinct categories: on one side, the symbolic, which can be treated as characters (keys, alterations, rests etc.), and on the other side, those that *Martin* and *Bellissant* [16.29] name the *iconic*, or assembled (e.g., note heads, stems, flags and beams), which are made of different primitives that may undergo various transformations and are assembled following certain rules.

Most OMR systems developed to date have more or less the same pipeline architecture that can be broken down into distinct phases:

1. Preprocessing of the image
2. Detection and removal of staff lines
3. Segmentation of the objects
4. Reconstruction of musical symbols
5. Classification and interpretation.

In many systems, musical rules are applied to increase accuracy. This can happen during several phases of the process, but most of the time the rules are applied during the reconstruction of the musical symbols from the primitives and during the classification of the symbols and the interpretation of the music content.

Using grammar to describe the musical notation and musical knowledge is a common technique [16.30]. Various methods have been proposed to integrate grammatical methods developed originally for languages into the bidimensional space of OMR. The transposition from one to two dimensions increases the complexity of the grammar, in both the design of the grammars and their use. This results in grammars that are either fairly simple but incomplete, or very complex but challenging to manage. One approach to simplify the problem is to have two distinct grammatical levels: low-level grammars, for describing the structure of a musical symbol constructed from primitives, and high-level grammars

for the description of the organization of musical notation itself [16.31].

16.4.1 Preprocessing

In many OMR research projects no particular attention is given to the preprocessing. The reason is that many of the problems to be solved are common to that faced by OCR, or even more generally in document recognition. In most researches a reference is made to the solutions proposed in these two research areas.

However, some more specific studies focusing on old documents looked more precisely at the thresholding problem during the binarization phase. This is the case with *MacMillan* et al. for the Levy project [16.32]. It shows that in this practical case, it is necessary to use more advanced methods of binarization. As does Ng for manuscript sources, MacMillan et al. use some locally adaptive thresholding algorithms. This technique allows the binarization to be optimized for document images with nonuniform lighting [16.33]. Furthermore, *Burgoyne* et al. [16.34] found that binarization methods that worked well for text documents did not work well for music documents and vice versa.

16.4.2 Staff-Line Detection and Removal

The detection and the removal of the staff lines is a key phase in OMR systems. This phase is often part of the preprocessing phase since usually the detection of the staff lines is used also to correct the skew of the document. Ideally, the staff lines are straight, parallel to each other, have constant thickness, and are horizontal. In reality, for various reasons, the lines may be curved, not parallel to each other, have varying thickness, and be skewed. A wide range of solutions has been proposed of tackle these challenges [16.21, 35]. These include projections, line tracing, two-dimensional vector fields, skeletonization, and graph-based approaches.

Once detected, staff lines are usually removed without removing the musical symbols. This step is critical, in particular with symbols touching the staff lines. In such cases, the removal of the staff lines may breakup the musical symbols into smaller pieces, which will then make their classification difficult without reconstruction. The difficulty of the task is therefore to find a solution that does not erase the staff line when it intersects with a symbol.

16.4.3 Recognition Architectures

The techniques used in OMR for locating and classifying symbols varies significantly from system to system and have evolved considerably since the first systems. This has been made possible by the considerable and steady increase of computational resources offered by new machines. Despite this diversity, some techniques are recurrent and are to be found similar in many approaches.

The simplest approach to locating and classifying the symbols is to look for the connected components and to classify them. This technology, already used by Prerau, generally uses simple measures such as the size of the bounding box of the symbol or the surface of the symbol. As simple as it is, this approach is still relatively effective because of the morphological diversity of some musical symbols (including in size) and can still be used today for a first classification phase.

Another approach widely used for the localization of the umsic symbols is to look for some easily detectable primitive and then to look around it for other primitives belonging to the expected symbol. For example, *Martin* and *Bellissant* [16.29] locate the note stems and then search for ellipses of note heads in the adjacent area. *Miyao* [16.36] uses a similar technique tailored for recognizing polyphonic passages. First, the note stems are located together with all the note heads and with stem flag candidates around each stem. Then, two neural networks are used to identify from the candidates those that are attached to a specific stem. With *Rossant* [16.37], vertical segments are not only used to locate the stems but also a whole series of symbols in which a vertical segment of a certain size is included: these symbols belong to various categories, such as notes, accidentals, bar lines, or certain types of rests. For the remaining symbols, Rossant performs a template correlation on limited zones: for locating whole note or half note rests, this area can be limited to a staff range, whereas the location of the whole notes, the correlation must be performed on a wider area.

Fujinaga [16.38] uses mainly vertical and horizontal projections. The symbols are located by vertical projection of the area between the upper line and the lower line. Different local projections are used to calculate the height, width, area, and the number of peaks to the vertical projection of the symbol. The data is used for classification in conjunction with some syntactic rules. For example, the first symbol on a staff is expected to be a clef; or a beam of eighth notes can contain only notes, rests or accidental; or a duration dot can be placed only after a note or rest. Classification is performed using a k-NN classifier together with a genetic algorithm.

Kato and *Inokuchi* [16.39] use a blackboard expert system where knowledge of musical notation is used in the process to remove ambiguities. The recognition is performed measure by measure once the clef, the

key signatures and the time signatures have been recognized. To allow optimal recognition of the different musical symbols that differ significantly in their size, their position, and their possible appearance frequency or importance, a wide variety of recognition methods are used. The architecture consists of four separate modules, which communicate with each other through a shared common data structure representing a measure. The role of the various modules is as follows:

1. Extraction of the primitives
2. Reconstruction of the symbols
3. Recognition of the symbols
4. Semantic analysis.

The memory shared by the four modules consists of five distinct layers representing five different levels of abstraction for a measure:

1. Pixels of the image
2. Primitives (note heads, stems, accidentals etc.)
3. Musical symbols (reconstituted notes, rests etc.)
4. Musical meaning of symbols (pitch, duration etc.)
5. Possible interpretation of the content of the measurement.

The recognition process is guided by a variable threshold that controls the recognition. Each layer makes assumptions that are verified by the upper layer from a tight line. If an upper layer considers a hypothesis unacceptable, the threshold is released and the treatment is reperformed at the lower level with the new threshold. For example, an unrecognized primitive will be put back in the pixels of the image. A document of good quality will be recognized with tighter thresholds and faster than poor-quality paper.

Ng and *Boyle* [16.24] use an iterative process of segmentation and recognition. Recognition is performed at different phases of the segmentation process, rather than requiring all the symbols to be segmented into primitive or broken signs prior to recognition. The purpose of this method is to avoid over-segmentation of the symbols. A first pass of recognition is performed directly after the removal of staff lines based on simple rules and using a *k*-NN classifier. For nonrecognized symbols, different types of subsegmentation and heuristics are performed.

Coüasnon [16.40] proposes a solution that uses the syntactic structure of musical notation. The idea is to use the a priori knowledge given by the music writing rules in order to guide the process of segmentation and labeling of objects. Coüasnon distinguishes between two types of information: the *physical* information corresponding to the arrangement of notes and attributes on the partition, and the *logical* information corresponding to the transformation of the notes in music writing. He defines a grammar that models this separation into two levels. The terminals of this grammar are the basic entities of the document description and are recognizable without contextual information in order to have information on which the recognition process can be based. They consist of terminal segments and symbols. The recognition process is performed in two phases corresponding to the two levels of information: graphic recognition and syntactic recognition. The first phase deals with the notes and recognizes the relative positioning of the attributes, while the second phase recognizes the symbols connected with a voice (slurs, dynamics etc.) and assigns the notes to the different voices depending on the vertical alignment and the number of beats per measure. The method specifically targets orchestral scores, where a distinction must be made between voices and staves: there may be up to three voices per staff and, conversely, a voice may be written on more than one staff.

Bainbridge and *Bell* [16.31] provide an extensible solution also based on the use of a grammar. The solution is meant to allow the recognition of different types of music notations and not be limited to CWMN. The system, named CANTOR, consists of distinct modules to be applied after the removal of stave lines:

1. Primitive recognition
2. Primitive assembly
3. Musical semantics.

Recognition of primitives is defined through a language specifically designed for the task. The idea is to accommodate a wide range of music writing by easily assigning a primitive recognition for each shape (e.g., projections, Hough transform etc.). Assembly of the primitive, the second module, is based on a slightly modified definite clause grammar, which has the limited scope of describing the taxonomy of musical notation. The purpose of keeping the scope of the grammar limited is to prevent it from becoming too complex and difficult to manage. The purpose of the last phase is to combine the symbols recognized by looking at their position in order to produce a structure representing the musical content of the image.

16.4.4 OMR Aggregation

Several researches have tried to improve OMR results by aggregating multiple commercial OMR tools. This approach relies on the fact that the commercial OMR tools all have their strengths and weaknesses and thus it should be possible to improve the overall recognition results by comparing the output of several tools. The first attempts were made by *Knopke* and *Byrd* [16.41]. Their approach involves two steps. First, the strengths

and the weaknesses of each tool that will be used are evaluated. Then the results of this evaluation are used for weighting the output of each tool when comparing their output and when the tools do not agree. *Bugge* et al. [16.42] propose a similar approach together with the definition of a dedicated format (MusicXiMpLe), a subset of the MusicXML file format used for the alignment of the tool's output and that is meant to facilitate the alignment of the data. More recently, *Church* and *Cuthbert* [16.43] have proposed a different approach that integrates rhythmic analysis in the alignment process.

16.5 Adaptive OMR

A wide range of music printing techniques have been used throughout history [16.44, 45], each having their own graphic particularities. This has led to the creation of highly varied music document symbols and notation practices. For each printing technique, publisher, editor, composer, or musical repertoire, the appearance of the notation can vary significantly. Printers often had their own distinctive font. The font shapes also vary depending on the size of the book, ranging from small in-octavo formats to larger in-folio or even in-plano formats. Font designs also have trends and changes occur over time. For example, the shape of the note heads in music fonts of the 16th and 17th centuries was generally diamond shaped. It then gradually changed and by the 18th century round note heads had become the trend.

In many OMR systems, the recognition of symbols is performed using supervised machine-learning algorithms. Supervised learning algorithms require ground-truth datasets of symbols in order to be trained. For each ground-truth symbol (e.g., a note head), features are extracted and fed to the algorithm for training. Once trained, the system will be able to identify similar symbols of the same category. When building an OMR system, gathering the ground-truth data for training is a highly time-consuming task since the amount of data required can be quite high, the data should ideally come from a wide range of sources, and each symbol needs to be correctly labeled. The high variability in the data makes it unrealistic to build an OMR system optimized for recognizing any document, even if targeting documents from a reduced historical period or one of restricted type.

A solution for tackling this issue is adaptive OMR [16.38, 46]. The assumption with adaptive OMR is that the system is not likely to consistently reach 100% accuracy and that most of the time, OMR workflows require human verification and correction to achieve usable results. In this context, adaptive OMR uses these correction results as further training data, feeding them back to the system and retraining the recognition system. As a result, systems will *learn* from their previous mistakes, correctly identifying previously misrecognized symbols through the expansion of its training data.

Pugin et al. [16.47] describe how this can be achieved using maximum a posteriori (MAP) adaptation, a technique widely used in handwriting and speech recognition. In a preliminary phase, a book-independent (BI) model is trained using ground-truth data taken from a selection of different books drawn from different printers and featuring variations in font shape and size, and is used as a seed for the recognition system. Thus the BI model gives acceptable results in general but is not specifically optimized for a particular source. In real-world usage, as each page of a book is recognized and corrected, the BI model is amended and optimized with MAP adaptation for the symbols in that book. As soon as the user has corrected the recognition errors on a newly processed page, that page is used as ground-truth to adapt the BI model. Eventually, after a user has corrected a number of pages, a book-dependent (BD) model emerges that is optimized for a particular set of sources (e.g., from a specific printer). The BD model can be saved and used for recognizing similar documents, creating a more optimal *bootstrap* than the general BI model.

16.6 Symbolic Music Encoding

The output of an OMR process is a machine-readable encoded music score. Whereas for text recognition, unformatted text files can serve as a basic encoded output of OCR processes, there is no equivalent for music encoding. There is no uniformly recognized basic music encoding scheme equivalent to unicode or to ASCII (American Standard Code for Information Interchange). For OMR applications, this means that

the encoded output has to be structured according to a defined music code, be it designed specifically for the OMR application, such as the Liszt format for the SharpEye commercial application, for example, or a more generic musical code.

Over the years, hundreds of musical codes have been proposed, illustrating both the complexity of music notation and the wide range of applications codes can serve [16.48]. Selfridge-Field groups them into three categories:

1. The codes targeting sound applications
2. The codes targeting notational applications
3. Those targeting analytical or more abstract applications.

Looking at how note pitches are handled is a simple way to help understand the difference between these code categories. In the first category, the most widely used code is undoubtedly MIDI (Musical Instrument Digital Interface), whose primary goal was to allow sound information to be exchanged between instruments. One particularity of MIDI regarding pitches is that it does not make the difference between, say, an F-sharp and a G-flat even if they were noted differently in the score. The second category of codes is meant to capture visual characteristics of music notation. Musical codes that belong to this category include DARMS (Digital Alternative Representation of Music Scores), one of the oldest computer codes; SCORE, the code used by the Score music notation software application developed by Leland Smith; or the notation interchange file format (NIFF). One particularity of the codes for notational applications is that they often do not have a concept of pitch. They do not store the notes by referring to the pitch name and the octave but instead by referring to the staff line on which they appear. The C_4 (C of the fourth octave) with the treble G-clef will be coded with the same staff line parameters as an E_2 with the bass F clef. Codes from the later category that target analytical applications, such as the plain and easy code, the kern representation, or the Essen associative code (EsAC), to mention only a few, store the note information through its pitch name and its octave.

Codes for notational applications are well suited to OMR applications. In fact, NIFF was developed with OMR in mind and is still used by some commercial software applications. NIFF was designed in the mid-1990s and subsequently came to be supported by a fair range of commercial OMR systems, including SharpEye, SmartScore, and PhotoScore. The use of NIFF remained limited, however, and it never really took off beyond its use by commercial OMR applications, which eventually abandoned it. Nor did another attempt in music code definition that was to extend MIDI for OMR with the expressive MIDI [16.48].

One reason NIFF failed to establish itself as a standard file format may be that it is a binary format. (Binary codes had the advantage of being more compact than ASCII codes, but disk space is no longer an issue for this type of data). But another reason might be that it is a notational code, with the limitation that only the graphical component of music notation is taken into account. In OMR, the recognition task acts as an encoding process that moves from a graphical domain to a logical domain. With a notational code such as NIFF, the processing of the data remains limited. For example, the pitch name and the octave of a note can remain unknown since only the position on the staff line is stored. In many uses of OMR data, however, further processing will be desirable in order to move to a code of the third category, the codes for an analytical or more advanced application. Typically, this will mean further processing the data in order to determine the pitch names and the octaves by taking into consideration the clef of the staff (or possible intermediate clef changes) and the key signature (or possible accidentals appearing previously in the measure). When the goal of the OMR process is the reediting of the original image, then a notational code can suffice, assuming that the editing tool can read the notational code produced by the OMR process. (It should to be noted that such an approach can have advantages regarding the impact of some recognition errors. For example, if a clef of a staff was wrongly recognized, it will have no impact on the accuracy of the content of the staff itself). This is, however, only one possible use case of OMR. Typical use cases of OMR require analytical data, for example to make an arrangement or transposition of the musical content. The analytical data can be derived from notational codes, but this is not necessarily trivial to do. It is one reason why encoding output formats other than NIFF were desirable and why it eventually became obsolete.

One analytical format that is widely used as an output of OMR processes is MusicXML, which is a code that started as an XML (extensible markup language) representation of MuseData [16.49]. It was developed by Michael Good and is now owned by MakeMusic. MusicXML is primarily an interchange format for exchanging music data between computer applications. This makes it well suited for exporting data from an OMR system to other types of applications, but it is not designed to represent and store the information of an OMR process.

Some XML alternatives to MusicXML were proposed for encoding music notation. One targets a wide range of applications, including OMR: the IEEE 1599–2008 standard designed by the IEEE 1599 Working

Group for XML Musical Application [16.50]. To our knowledge this standard is not currently used by any OMR systems.

16.6.1 The Music Encoding Initiative (MEI)

Over the last few years, a community-based project began to occupy an increasingly important role: the Music Encoding Initiative (MEI) [16.51]. The project was started around 2000 by Perry Roland from the University of Virginia. It was directly inspired by the Text Encoding Initiative (TEI), a leading project in text studies that became over the years the commonly accepted standard for representing and encoding texts. MEI pursues similar goals for musical documents. It is expressed in the form of an XML schema that defines the structure of the corresponding XML musical data. The first version of the MEI schema was released in 2010 in ODD (one document does it all), which is a schema definition solution developed by the TEI community that regroups in one document the schema definition and all the documentation related to it. The MEI schema is regularly updated to incorporate the latest improvements.

One feature of MEI in contrast to many attempts to define a musical code is that it is driven by an open community of contributors representing a wide range of backgrounds and interests. They include technologists, musicologists with various repertoires of expertise, and librarians. Beside the fact that MEI is community driven, it also has the particularity of accommodating a wide range of music notation and not being limited to CWMN – even though this is its first target. This flexibility is greatly facilitated by its modular approach and its nonmonolithic design. Each music notation type can be defined as a separate module, for example, MEI already includes in its core set of modules specialized modules for mensural and neume notations. The modularity of MEI does not serve only the separation of music notation types. MEI includes distinct modules for the metadata, for pointers and references, for linking with facsimile or for the definition of graphics, shapes, and symbols. Furthermore, each module can be modified or new modules added, should the default settings of MEI not be appropriate. This can happen when encoding a different notation type or when another type of application is targeted by the encoding. MEI includes a so-called customization service that allows part of the schema to be redefined or new encoding concepts to be introduced [16.52].

Even though MEI was not designed specifically for OMR applications, its richness and flexibility makes it perfectly appropriate for them [16.53]. The aforementioned module for linking with facsimile images is of particular interest. It makes it possible to easily and precisely refer from the encoding to zones in an image. This is similar to what is achieved with the hOCR format developed by Google for their open-source OCR project OCRopus. Since the output is text with OCR, they can use HTML (hypertext markup language) to mark up the text output [16.54]. However, the hOCR format enriches the HTML with additional information on layout, referring to an image, and data such as character recognition confidences. This is done in a way that does not affect the structure of the HTML content, which remains standard HTML. In a similar (though not identical) way, the MEI can include additional information, including references to image zones without the logical structure of the MEI to be modified. With MEI, references to images work by defining a facsimile subtree in the MEI file that regroups a set of surfaces (typically each one representing a page) that will contain a reference to an image and a list of zones in it. Each zone is identified with a unique identity to which any element in the encoding (a clef, for example) can refer. This not only enables robust linking between the encoding and the image to be generated, but it also has the advantage of keeping the information concerning the references to the images separate from the logical musical data.

16.7 Tools

16.7.1 Commercial OMR Software

Although several commercial OMR software packages have been marketed, currently only a handful products have stood the test of time. One of the survivors is also the first commercial OMR software ever published. Initially under the name MIDISCAN, Musitek released its product in August 1993 [16.55]. It was renamed later as SmartScore and its Lite version was bundled with a music editing software Finale 2003 in the spring of 2002. Another popular music editing software Sibelius (now owned by Avid Technology) uses Photoscore [16.56], which was originally released by Neutron for the Acorn system in 1997 (for Windows in 1999 and for the Mac in 2000). The ScoreMaker by Kawai has been available on the Windows system (in Japanese only) since 1995 and the capella-scan [16.57] from Germany has been around since about 2000 [16.58]. According to the developer, Graham Jones, the development of SharpEye started in 1996 [16.59], but it has not been updated

since 2006 (version 2.86) when he *transferred rights in SharpEye to another company* [16.60]. All of the commercial OMR software mentioned above are designed to work with CWMN.

16.7.2 Open-Source Tools and Toolkits

A few OMR tools are available as open-source. We can make the distinction between end-user ready-to-use desktop applications and toolkits available for creating custom OMR applications.

Desktop Applications

For CWMN, the only open-source tool available is Audiveris [16.61]. It is written in Java and is available under the General Public License (GPL) v2 open-source license. For text recognition, Audiveris uses the Google OCR Tesseract engine. The recognition engine is based on neural networks that can be retrained for specific sources. Audiveris includes a user interface for data correction and exports to MusicXML. Version 5 is currently in preparation.

One open-source OMR project specifically targets Renaissance music prints: the Aruspix project [16.62]. The project focuses on the development of techniques and tools for processing early typographic music prints from the 16th and 17th centuries. The Aruspix software application is a desktop application written in C++ (cross-platform) and is available under the GPL v3 license. Aruspix uses HMM (hidden Markov models) for the recognition with an original approach without staff removal [16.14]. It includes a user editor for correcting the results and an adaptive feature based on MAP adaptation [16.47]. Aruspix uses MEI as internal and output format.

Toolkits

The most widely used toolkit for building OMR systems is Gamera [16.17, 63]. Gamera is written in C++ and Python and is available under the GPL v2 license. It is designed as a toolkit for building pattern recognition systems with a strong focus on document recognition and OMR in particular. It includes a whole range of image processing algorithms together with a *k*-NN classifier. With this tool, the users can build their own recognition system by putting together scripts that will perform selected operations. These can include various preprocessing operations, such as noise removal, blurring, deskewing, contrast adjustment, sharpening, binarization, and morphology, for example. At the core of the Gamera system is the segmentation and the classification. Several algorithms are provided for these tasks together with a user interface for labeling the data. Both the image processing and the classifier parts are easily extendable, if necessary, with either C++ or Python extensions.

Gamera is distributed with add-on toolkits specifically designed for building OMR applications. The MusicStaves toolkit implements various algorithms for removing the staff lines [16.35]. It can be used in an interactive mode through the Gamera graphical user interface or in a noninteractive mode from the command line or scripts. The OTR (optical tablature recognition) toolkit is a package for the recognition of lute tablature [16.9, 11]. It includes scripts for the recognition of French, German, and Italian tablatures, all being slightly different tablature notations. Gamera also includes a Psaltiki recognition toolkit for the recognition of Byzantine chant notation used in chant notation of the Eastern churches [16.13]. It also distributes an OCR toolkit with custom page segmentation algorithms and heuristics rules for dealing with diacritics.

Gamera has been used for projects on other music notation, including neumes [16.64]. Gamera is also often used for evaluating research techniques, such as with Aruspix [16.65] where two OMR approaches are compared, or on more specific OMR steps such as staff line removal [16.66–68].

16.8 Future

Recent directions taken in OMR developments together with changes in technology make it possible to envisage completely new OMR paradigms. For decades, the goal of OMR has focused almost exclusively on providing image transcriptions for further processing with external software applications. Commercial OMR applications are usually desktop software applications that take an input image and output an encoded score without making further use of the link established between the image and its content.

Recent advances that can change this paradigm are manifold. First of all, open-source developments, such as Gamera, open new perspectives. The OMR technology is no longer embedded in a black box but instead remains open in modules that can be modified and assembled differently according to the needs and type

of documents to be processed. Many music documents raise similar though not identical challenges, and being able to adjust the tools is essential. The aforementioned Gamera MusicStaves toolkit is a perfect example in that regard.

Adaptive-OMR design is another direction that opens new perspectives and that also differs from the design of most commercial OMR desktop applications. Having systems that improve themselves over time is essential for bringing OMR to a next level. Correcting OMR errors is a highly time-consuming task, and the advantages gained by being able to feed this knowledge back into the system appear self-evident. Nonetheless, for years users have been using OMR systems and correcting their output without exploiting this data goldmine. This wasteful practice can be changed, but only if the output of the OMR process preserves the link between the output data and the image. It appears that the developments of MEI will play a key role in this endeavor. Having a standard format for preserving OMR data will allow the creation of large datasets. They will be usable for training or improving any OMR systems, which in turn will greatly facilitate the development and improvement of OMR technology.

One significant technological change that has occurred over the last few years is the emergence of online applications that can run in web browsers. It is now possible to develop software applications that run online without any application or plugin to be downloaded and installed locally. This radically changes the way software applications can be made available to the users. Over the next few years we can expect to see online OMR tools appearing that will be a significant breakthrough from the desktop applications currently available. In this context, the development of MEI engraving tools, such Verovio [16.69, 70] will be essential for making the OMR output editable online. For OMR, having online tools will also make it possible to develop adaptive systems where the corrections of one user can be immediately incorporated into the system and benefit all users, not only the one who made the corrections.

Online technology also transforms the way data can be made available. We now have access to thousands of images of music sources that are being digitized and made available by music libraries and archives all around the world. These images are an unparalleled resource for musicians, musicologists, and other scholars alike. However, only OMR technology can fully revolutionize the way they are made available to the user. The recent developments of Diva.js [16.71, 72] offer a glimpse of the future, where the output of the OMR process in MEI is displayed on top of high-resolution images directly in the web browser. This is a setup that is widely known for books but is still lacking for music, and large-scale online OMR technology is the key to fill this gap.

As large amounts of score data become available in symbolic formats, the next challenge is how to effectively use these large corpora of musical data. There are two basic issues: searching and analysis. Searching music is complex: there are pitches, rhythms, text, multiple voices sounding simultaneously, chords, different instruments etc. Linking metadata for sources and works (e.g., date of composition, location, or genre) and musical content is also essential. Analyzing music is also challenging given the large amounts of symbolic data that were not available previously. Thus, there will be questions such as what are the best ways to search through these corpora? What should the queries look like? What kinds of user interfaces are needed for queries and displays? What types of analysis of music are possible given these large datasets? The answers to these and other questions will create new research avenues paved with the aid of OMR technology.

References

16.1 D.H. Shepard: Apparatus for reading, Patent Application 2664758 (1951)

16.2 D. Martin: David H. Shepard, 84, Dies; Optical Reader Inventor, New York Times, 11 December 2007

16.3 D. Pruslin: *Automatic Recognition of Sheet Music*, Sc. D. Diss. (Massachusetts Institute of Technology, Cambridge 1966)

16.4 D. Prerau: *Computer Pattern Recognition of Standard Engraved Music Notation*, PhD Diss. (Massachusetts Institute of Technology, Cambridge 1970)

16.5 A. Samuel: The banishment of paper-work, New Sci. **21**(380), 529–530 (1964)

16.6 S. Mori, C. Suen, K. Yamamoto: Historical review of OCR research and development, Proc. IEEE **80**(7), 1029–1058 (1992)

16.7 D.S. Prerau: Computer pattern recognition of printed music. In: *Fall Joint Computer Conference 1971*, AFIP Conf. Proc., Vol. 39 (1971) pp. 153–162

16.8 D. Blostein, H.S. Baird: A critical survey of music image analysis. In: *Structured Document Image Analysis*, ed. by H.S. Baird, H. Bunke, K. Yamamoto (Springer, Berlin 1992) pp. 405–434

16.9 C. Dalitz, T. Karsten: Using the Gamera framework for building a lute tablature recognition system. In: *6th Int. Soc. Music Inf. Retr. Conf. (ISMIR)* (2005)

pp. 478–481
16.10 L.L. Wei, Q.A. Salih, H.S. Hock: Optical tablature recognition (OTR) system: Using Fourier descriptors as a recognition tool. In: *2008 International Conference on Audio, Language and Image Processing, Shanghai* (2008) pp. 1532–1539, https://doi.org/10.1109/ICALIP.2008.4590235
16.11 C. Dalitz, C. Pranzas: German lute tablature recognition. In: *Int. Conf. Document Anal. Recognit. (ICDAR)* (2009) pp. 371–375
16.12 V.G. Gezerlis, S. Theodoridis: Optical character recognition of the orthodox hellenic byzantine music notation, Pattern Recognit. **35**(4), 895–914 (2002)
16.13 C. Dalitz, G.K. Michalakis, C. Pranzas: Optical recognition of psaltic Byzantine chant notation, Int. J. Doc. Anal. Recognit. (IJDAR) **11**(3), 143–158 (2008)
16.14 L. Pugin: Optical music recognition of early typographic prints using hidden Markov models. In: *7th Int. Conf. Music Inf. Retr. (ISMIR)* (2006) pp. 53–56
16.15 L. Tardón, S. Sammartino, I. Barbancho, V. Gómez, A. Oliver: Optical music recognition for scores written in white mensural notation, EURASIP J. Image Video Process. **2009**, 843401 (2009), https://doi.org/10.1155/2009/843401
16.16 D. Bainbridge: *Extensible Optical Music Recognition*, PhD Diss. (University of Canterbury, Canterbury 1997)
16.17 K. MacMillan, M. Droettboom, I. Fujinaga: Gamera: Optical music recognition in a new shell. In: *Proc. Int. Comput. Music Conf.* (2002) pp. 482–485
16.18 D. Marr: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (Freeman, New York 1982)
16.19 T. Pun: C. De. Garrini: Cybernétique et vision par ordinateur. In: *Le déficit visuel, de la neurophysiologie à la pratique de la réadaptation*, ed. by A.B. Safran, A. Assimacopoulos (Masson, Paris 2014) pp. 213–224
16.20 R. Bruyer: *Le Cerveau Qui Voit* (Editions Odile Jacob, Paris 2000)
16.21 A. Rebelo, I. Fujinaga, F. Paszkiewicz, A.R.S. Marcal, C. Guedes, J.S. Cardoso: Optical music recognition: State-of-the-art and open issues, Int. J. Multimed. Inf. Retr. **1**(3), 173–190 (2012)
16.22 K.M. Sayre: Machine recognition of handwritten words: A project report, Pattern Recognit. **5**, 213–228 (1973)
16.23 T. Plötz, G. Fink: Markov models for offline handwriting recognition: A survey, Int. J. Document Anal. Recognit. **12**, 269 (2009)
16.24 K.C. Ng, R.D. Boyle: Recognition and reconstruction of primitives in music scores, Image Vis. Comput. **14**(1), 39–46 (1996)
16.25 I. Fujinaga, J. Riley: Recommended best practices for digital image capture of musical scores. In: *3rd Int. Conf. Music Inf. Retr. (ISMIR)* (2002) pp. 261–263
16.26 W. Koseluk: Digitalization of musical sources: An overview. In: *The Virtual Score: Representation, Retrieval, Restoration, Computing in Musicology*, Vol. 12, ed. by W.B. Hewlett, E. Selfridge-Field (MIT Press, Cambridge 2001) pp. 219–226
16.27 D. Bainbridge, T. Bell: The challenge of optical music recognition, Comput. Humanit. **35**, 95–121 (2001)
16.28 E. Selfridge-Field: Optical recognition of musical notation: A survey of current work. In: *Computational Musicology: An International Directory of Applications*, Vol. 9, ed. by W.B. Hewlett, E. Selfridge-Field (1993) pp. 109–146
16.29 P. Martin, C. Bellissant: Low-level analysis of music drawings. In: 1st Int, Conf. Doc. Anal. Recognit., ICDAR **pp**, 417–425 (1991)
16.30 H. Fahmy, D. Blostein: A graph grammar programming style for recognition of music notation, Mach. Vis. Appl. **6**, 83–99 (1993)
16.31 D. Bainbridge, T. Bell: A music notation construction engine for optical music recognition, Softw. Pract. Exp. **33**(2), 173–200 (2003)
16.32 K. MacMillan, M. Droettboom, I. Fujinaga: Gamera: A structured document recognition application development environment. In: *2nd Int. Symp. Music Inf. Retr. ISMIR* (2001) pp. 173–178
16.33 K.C. Ng: Music manuscript tracing. In: *4th Int. Workshop, Graphics Recognit.: Algorithms and Applications (GREC)* (2001) pp. 322–334
16.34 J. Burgoyne, L. Pugin, G. Eustace, I. Fujinaga: A comparative survey of image binarisation algorithms for optical recognition on degraded musical sources. In: *8th Int. Conf. Music Inf. Retr. (ISMIR)* (2007) pp. 509–512
16.35 C. Dalitz, M. Droettboom, B. Pranzas, I. Fujinaga: A comparative study of staff removal algorithms, IEEE Trans. Pattern Anal. Mach. Intell. **30**(5), 753–766 (2008)
16.36 H. Miyao: Stave extraction for printed music scores. In: *3rd Int. Conf. Intell. Data Eng. Automated Learning (IDEAL)* (2002) pp. 562–568
16.37 F. Rossant: A global method for music symbol recognition in typeset music sheets, Pattern Recognit. Lett. **23**(10), 1129–1141 (2002)
16.38 I. Fujinaga: Exemplar-based learning in adaptive optical music recognition system. In: *Int. Comput. Music Conf* (1996) pp. 55–60
16.39 H. Kato, S. Inokuchi: A recognition system for printed piano music using musical knowledge and constraints. In: *Int. Assoc. Pattern Recognit. Workshop on Syntactic and Struct. Pattern Recognit* (1990) pp. 231–248
16.40 B. Coüasnon: Formalisation grammaticale de la connaissance a priori pour l'analyse de documents: Application aux partitions d'orchestre. In: *Actes du dixième congrès Reconnaissance des Formes et Intelligence Artificielle, Rennes* (1996) pp. 465–474
16.41 I. Knopke, D. Byrd: Towards musicdiff: A foundation for improved optical music recognition using multiple recognizers. In: *8th Int. Conf. Music Inf. Retr. (ISMIR)* (2007) pp. 123–126
16.42 E.P. Bugge, K.L. Juncher, B.S. Mathiesen, J.G. Simonsen: Using sequence alignment and voting to improve optical music recognition from multiple recognizers. In: *12th Int. Soc. Music Inf. Retr. Conf. (ISMIR)* (2011) pp. 405–410
16.43 M. Church, M.S. Cuthbert: Improving rhythmic transcriptions via probability models applied post-

OMR. In: *15th Int. Soc. Music Inf. Retr. Conf. (ISMIR)* (2014) pp. 643–648
16.44 H.E. Poole: Music printing. In: *Music Printing and Publishing*, ed. by D.W. Krummel, S. Sadie (Norton, New York 1990) pp. 3–78
16.45 R. Rasch (Ed.): *Music Publishing in Europe 1600–1900 Concepts and Issues, Bibliography* (Berliner Wissenschafts, Berlin 2005)
16.46 F. Rossant, I. Bloch: Robust and adaptive OMR system including Fuzzy modeling, Fusion of musical rules, and possible error detection, EURASIP J. Adv. Signal Process. **2007**, 81541 (2007)
16.47 L. Pugin, J.A. Burgoyne, I. Fujinaga: MAP adaptation to improve optical music recognition of early music documents using hidden Markov models. In: *8th Int. Conf. Music Inf. Retr. (ISMIR)* (2007) pp. 513–516
16.48 E. Selfridge-Field: *Beyond MIDI: The Handbook of Musical Codes* (MIT Press, Cambridge 1997)
16.49 Makemusic Inc.: musicXML, http://www.musicxml.com (2017)
16.50 WG_1599 – Working Group for XML Musical Application: 1599-2008 – IEEE Recommended Practice for Defining a Commonly Acceptable Musical Application Using XML, http://standards.ieee.org/findstds/standard/1599-2008.html (2017)
16.51 Music Encoding Initiative: http://www.music-encoding.org
16.52 A. Hankinson, P. Roland, I. Fujinaga: The music encoding initiative as a document-encoding framework. In: *12th Int. Soc. Music Inf. Retr. Conf. (ISMIR)* (2011) pp. 293–298
16.53 A. Hankinson, L. Pugin, I. Fujinaga: An interchange format for optical music recognition applications. In: *11th Conf. Int. Soc. Music Inf. Retr. (ISMIR)* (2010) pp. 51–56
16.54 T.M. Breuel, U. Kaiserslautern: The hOCR microformat for OCR workflow and results. In: *Int. Conf. Document Anal. Recognit. (ICDAR)* (2007) pp. 1063–1067
16.55 S. George: Evaluation in the visual perception of music. In: *Visual Perception of Music Notation: Online and Offline Recognition*, ed. by S. George (IRM, Hershey 2004) p. 308
16.56 M. Dawe: About Neuratron, http://www.neuratron.com (2015)
16.57 capella-software AG: Products, http://www.capella.de/us/index.cfm/products (2017)
16.58 Visiv Ltd: User comments, reviews, etc., http://www.visiv.co.uk/quote.htm (2006)
16.59 Visiv Ltd: Version History, http://www.visiv.co.uk/vershv2.htm (2006)
16.60 Graham Jones: http://www.indriid.com/grahamjones.html
16.61 Wikipedia: Audiveris, https://en.wikipedia.org/wiki/Audiveris (2017)
16.62 Laurent Pugin: Aruspix, http://www.aruspix.net
16.63 Christoph Dalitz: GAMERA Project, http://gamera.informatik.hsnr.de
16.64 G. Vigliensoni, J.A. Burgoyne, A. Hankinson, I. Fujinaga: Automatic pitch detection in printed square notation. In: *Proc. Int. Soc. Music Inf. Retr. Conf., Miami* (2011) pp. 423–428
16.65 L. Pugin, J. Hockman, J.A. Burgoyne, I. Fujinaga: Gamera versus Aruspix: Two optical music recognition approaches. In: *9th Int. Conf. Music Inf. Retr. (ISMIR)* (2008) pp. 419–424
16.66 J. Cardoso, A. Capela, A. Rebelo, C. Guedes: A connected path approach for staff detection on a music score. In: *Proc. 15th IEEE Int. Conf. Image Process* (2008) pp. 1005–1008
16.67 A. Dutta, U. Pal, A. Fornés, J. Lladós: An Efficient Staff Removal Approach from Printed Musical Documents. In: *Proc. 2010 20th Int. Conf. Pattern Recognit* (2010) pp. 1965–1968
16.68 A. Fornés, V.C. Kieu, M. Visani, N. Journet, A. Dutta: The ICDAR/GREC 2013 Music Scores Competition: Staff removal, Lect. Notes Comput. Sci. **8746**, 207–220 (2014)
16.69 Laurent Pugin: Verovio, http://www.verovio.org
16.70 L. Pugin, R. Zitellini, P. Roland: Verovio: A library for engraving MEI music notation into SVG. In: *15th Int. Conf. Music Inf. Retr. (ISMIR)* (2014) pp. 107–112
16.71 McGill University: http://ddmal.github.io/diva.js (2016)
16.72 A. Hankinson, W. Liu, L. Pugin, I. Fujinaga: Diva: A web-based document image viewer. In: *Conf. Theory Prac. Digital Libraries* (2011)

17. Adaptive Musical Control of Time-Frequency Representations

Doug Van Nort, Phillippe Depalle

In this chapter we consider control structures and mapping in the process of deciding upon the underlying sonic algorithm for a digital musical instrument. We focus on control of timbral and textural phenomena that arise from the interaction and modulation of stationary spectral components, as well as from stochastic elements of sound. Given this observation and general design criteria, we focus on a family of sound models that parameterize the stationary and stochastic components using a spectral representation that is commonly based on an underlying short-time Fourier transform (STFT) analysis. Using this as a fundamental approach we build a dynamic model of sound analysis and synthesis, focusing on a design that will simultaneously lead to musically interesting transformations of textural and noise-based sound features while allowing for control structures to be integrated into the sound dynamics. Building upon well-established adaptive algorithms such as the Kalman Filter, we present a recursive-exponential implementation, and exploit a fast algorithm derivation in order to process both additive data and the full underlying phase vocoder. The model is further augmented to allow for nonlinear adaptive control, pointing towards new directions for adaptive musical control of time-frequency models.

At this point in the field of digital music synthesis and control, there are many high-quality algorithms available for sound synthesis and transformation. Similarly, the issue of mapping and control structures to expressively manipulate these algorithms is also a well established field [17.1, 2]. The dominant control paradigm thus far has been to impose structures onto synthesis algorithms in a top-down fashion, following established paradigms such as continuous timbre space navigation [17.3]. However, there exists a largely unexplored territory that can advance the design of systems for digital, expressive musical control: the concurrent design of sound synthesis and control algorithms from a very low level. Such an approach can allow for nuanced manipulation that deeply integrates sonic and expressive goals, by expressing both the temporal dynamics of input gestures as well as how this control affects the sound over time. This requires one to consider control in a lower-level part of the instrumental design hierarchy, and to *embed this in the description of the underlying sound process*. Such an approach can become particularly powerful in the case of analysis/transformation/synthesis systems based on spectral models. In this chapter we present a particular class of adaptive analysis/synthesis systems whose representation can function as a hybrid between signal

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_17

and physical modeling approaches, as well as between source-filter and additive models. The models we develop for sound transformation are novel and interesting in their own right, allowing for various textural and spectral effects. However we further reconsider classic techniques in a new framework not simply for new transformation possibilities, but extend this by building the normally *higher-level* control layer into the model itself, showing how mapping and control can be considered at the most atomic level, so that the design of musical control is deeply tied into the very definition of the system.

When one considers the type of sound they want to control they are, at least to some extent, imagining the space of possible transformations of this sound. Therefore, it is beneficial to consider the transformations afforded by a given sound model, in conjunction with the myriad ways that this model may be controlled. This chapter will focus on a family of sound models that parameterize the stationary and stochastic components using a spectral representation that is commonly based on an underlying short-time Fourier transform (STFT) analysis [17.4]. Using this as a fundamental approach the chapter presents a dynamic model of sound analysis and synthesis, focusing on a design that will simultaneously lead to interesting transformations of spectral, textural and noise-based sound features while allowing for control structures to be integrated into the sound dynamics. This is achieved through a novel approach based on a recursive-exponential implementation, augmented to allow for nonlinear adaptive control.

17.1 State-Space Analysis/Synthesis

The dominant paradigm for understanding a linear system (audio filters, the Fourier transform, etc.) typically lies in the frequency domain through a transfer function representation, which describes the input/output relationship that the system will produce. Any such system can also be represented as a recursive difference equation that expresses not just the out-of-time effect of the system, but the dynamics of this system over time [17.5]. This is achieved through a state-space representation (SSR): rather than model a system in terms of its transfer-function representation one can express its actions on a stochastic process $\boldsymbol{y}[n]$ by way of the two state-space equations

$$\begin{aligned} \boldsymbol{s}[n+1] &= \mathbf{A}\boldsymbol{s}[n] + \boldsymbol{w}[n] \quad \text{and} \\ \boldsymbol{y}[n] &= \mathbf{B}\boldsymbol{s}[n] + \boldsymbol{v}[n]\,, \end{aligned} \tag{17.1}$$

where the sequence $\boldsymbol{s}[n]$ is the state of the process at time n, and the upper equation represents the internal dynamics of the process as governed by dynamics matrix $\mathbf{A}$. The lower equation projects the state vector, which may be hidden, into a vector of observable output variables using observation matrix $\mathbf{B}$. Both $\boldsymbol{w}$ and $\boldsymbol{v}$ are assumed to be zero-mean, Gaussian white-noise processes. The first affects the progression of the state while the second is additive noise present in the output process $\boldsymbol{y}$.

Using the SSR, one can express and modify the dynamics of a sound transformation system, building a control structure around this representation. As we will see, methods for statistical estimation can be applied to this framework, and can be extended to modeling in nonstationary and nonlinear environments. A further advantage is that a physically inspired model of a control/sound system can be used to constrain the interaction, so that physical and more abstract signal models can be used in a hybrid fashion. The approach presented here takes advantage of this potential for parameter estimation, accounting for nonlinearity and model hybridity.

17.1.1 The State-Space Phase Vocoder

In general, the phase vocoder is a widely used tool for the analysis, transformation and synthesis of audio signals. It began as an attempt to efficiently code and transmit voice signals using filter banks [17.6], was later represented by the STFT [17.7] and then began to find use in musical applications [17.8, 9]. The most common effects generated by the use of the phase vocoder are pitch shifting and time scaling, which are achieved through altering the time/frequency block increment size between the analysis and synthesis step and then interpolating. If the step increment for both analysis and synthesis is subjected to certain constraints based on the type of windowing function used in the STFT, then the input signal is perfectly reconstructed upon resynthesis. However the phase vocoder becomes musically interesting when the signal is distorted by transformations such as pitch/time scaling, cross-synthesis and others in which the amplitude and phase of each frequency bin are modified over time. As the representation itself is purely deterministic and able to capture the signal entirely, these distortions are externally applied to the spectral data in an intermediate (i. e., between analysis and synthesis) step.

Now, the creation of an SSR-based phase vocoder is possible by exploiting a recursive description of the discrete Fourier transform (DFT), as was presented in [17.10]. The approach exploits the fact that the complex exponentials of the DFT can be expressed as

$$e^{jn\theta} = e^{j\theta} e^{j(n-1)\theta} , \tag{17.2}$$

for time n and frequency θ. Thus the DFT matrix and its inverse can be expressed as a first-order recursion, which from the above equations we can see is a necessity in order to work within this SSR framework. Therefore the DFT matrix may be represented by the state matrix $\mathbf{A}$, which becomes a block diagonal matrix consisting of 2×2 rotation submatrices, and can be thought of as expressing the *process of the Fourier transform* in the way one might imagine a heterodyning filter to act over time in an analog filter-bank DFT implementation [17.11]. The state $\boldsymbol{s}$ in this case is a vector representing the spectral frames acted on by the DFT matrix at each time step, while $\boldsymbol{v}$ can be thought of as an additive output noise similar to the residual found in the spectral modeling synthesis (SMS) approach [17.4]. This basic representation forms the foundation for the control/sound models that are derived throughout this chapter. In [17.12] we presented musically inspired effects created by exploiting its explicit representation of spectral dynamics and a stochastic component, known as the stochastic state-space phase vocoder (SSSPV). We extended the generic structure of (17.1) by introducing noise into the state matrix $\mathbf{A}$. This resulted in modifying the dynamic of the process, which led to more dramatic effects. Indeed, adding such noise in the dynamic of the synthesis with different time-varying behaviors allows for effects such as concurrent random modulations between partials that can induce jitter and lead to an influence of a given sound's texture [17.13].

17.1.2 Related Work

A state-space approach to analysis/synthesis was presented in [17.14] in which the real and imaginary components of p sinusoidal partials, tracked over time, were represented in the state vector. The observation matrix summed across the real components of the partials, and the addition of observation noise generated a sinusoid+noise resynthesis. The same underlying model was also used in [17.15], though the spectral components that were tracked did not necessarily represent partials. Similarly, a recursive state-space formulation is presented in [17.16] wherein the state is comprised of the real and imaginary components for N evenly spaced frequency bins. Thus, this implementation maintains all of the data from the phase vocoder while the aforementioned work directly tracks partials and so is a sinusoidal model. The motivation differs in [17.16] as well, with the goal being the interpolation of missing audio samples whereas the former two projects were concerned with building an analysis/synthesis scheme for audio transformations. The work presented in this chapter is situated between these two as the motivation is towards musical transformations, yet preserves the lower-level representation given by the complete Fourier spectral frames.

Each of these approaches utilizes adaptive filtering and estimation to jointly estimate and predict spectral models of a musical signal that spans from the low level of spectral envelope and partials up to musical structure. The unifying characteristics across each is that they rely on a state-space modeling framework. This approach to signal representation allows for hybrids between a signal model and a physical model, as was shown by [17.17] and [17.18].

17.1.3 Beyond SSSPV: From Effect to Transformation

The SSSPV can be considered as a phase vocoder in which a stochastic element has been built into the representation via a state-space framework. The focus with this implementation is on the fact that through the embedding of noise within the system representation itself (rather than as input to the state or observation equations) a sound can be resynthesized with an added textural or roughness quality. While this approach can provide a family of interesting effects, there are some limitations with this implementation. Primarily, this state-space representation diverges from a classic phase vocoder in that it is actually a recursive implementation of the DFT. In other words, it is only defined for a single analysis window, and the input signal must be reintroduced at the boundary between windows. This is not a problem for this set of noise-based effects, but it inhibits the implementation of deeper transformations based on time stretching and pitch shifting across large time scales.

The use of the state-space representation is interesting with the SSSPV in that it gives access to the internal dynamics of the analysis/synthesis process for control and processing. However a deeper reason for using the state-space approach is that it allows for the use of tools such as the Kalman filter for parameter estimation. To this end, we build upon the state equations from (17.1), and using the same underlying spectral model as in SSSPV, design an STFT-based recursive implementation that functions as a *true* phase vocoder, capable of deeper transformations than noise-based effects.

17.2 Recursive, Infinite-Length Windows

Creating a state-space recursion based solely on (17.2) defines a DFT that implicitly assumes a periodic input signal whose fundamental period is defined by the N input samples. It is further an implicit rectangular windowing, and by reintroducing the input each N samples (as in SSSPV) there is no overlap in the analysis. In order to extend this, we will illustrate how the STFT can also be expressed with a one-step recursion. In the process, this will introduce an infinite-length window into the definition, which provides certain beneficial time-frequency properties; in particular, through a single-sided exponential window [17.19] $h[m]$ defined as

$$h[m] = \mathrm{e}^{\lambda m} u_+[-m] \,, \tag{17.3}$$

where $u_+[m]$ is the unit step function. The use of this window allows one to accurately detect the beginning of signals (due to the discontinuous front edge of the window) as well as to tune the influence of the past samples through changing the damping value $\lambda > 0$. Following this, the STFT

$$X_{n,k} = \sum_{m=-\infty}^{\infty} x[m]h[m-n]\mathrm{e}^{-\mathrm{j}\omega_k m} \tag{17.4}$$

for real input signal $x[m]$ becomes

$$X_{n,k} = \sum_{m=-\infty}^{n} x[m]\mathrm{e}^{\lambda(m-n)}\mathrm{e}^{-\mathrm{j}\omega_k m} \,. \tag{17.5}$$

The one-step recursion can be derived by looking at the value

$$X_{n+1,k} = \sum_{m=-\infty}^{n+1} x[m]\mathrm{e}^{\lambda(m-(n+1))}\mathrm{e}^{-\mathrm{j}\omega_k m} \,, \tag{17.6}$$

which expands to

$$X_{n+1,k} = \sum_{m=-\infty}^{n} x[m][\mathrm{e}^{\lambda(m-n)}\mathrm{e}^{-\lambda}]\mathrm{e}^{-\mathrm{j}\omega_k m} + x[n+1]\mathrm{e}^{-\mathrm{j}w_k(n+1)} \tag{17.7}$$

and thus

$$X_{n+1,k} = \mathrm{e}^{-\lambda} X_{n,k} + x[n+1]\mathrm{e}^{-\mathrm{j}\omega_k(n+1)} \,. \tag{17.8}$$

Therefore the STFT at any given time – when using this window function – is a product of the exponentially damped previous time step and the Fourier transform component from the current time-series value. In light of the state-space representation developed in the previous section, we can now rewrite the state equation as

$$\boldsymbol{s}[n+1] = \hat{\mathbf{A}}\boldsymbol{s}[n] + \mathbf{D}_n \boldsymbol{u}[n+1] + \boldsymbol{w}[n] \,, \tag{17.9}$$

where

$$\hat{\mathbf{A}} = \mathrm{e}^{-\lambda}\mathbf{A} \quad \text{and} \quad \mathbf{D}_n = (\bar{\mathbf{A}})^n \,, \tag{17.10}$$

with $\bar{\mathbf{A}}$ defined as the block diagonal from $\mathbf{A}$ extracted and collapsed in order to form a $2N\times 2$ matrix. Finally $\boldsymbol{u}[n] = [x[n+1]\ 0]^{\mathrm{T}}$ is the real-valued time series at time $n+1$ and functions as the *input vector* in terms of the state-space formalism.

Now, this implementation is an adaptive version of the STFT in which the infinite-length exponential window acts as a forgetting factor. This smoothing extends the previous recursive DFT into an STFT through a consideration of all past sample values. From this point onward, let us refer to this implementation as the recursive exponential STFT, or RESTFT (RESTFT). Note that this differs from a standard STFT wherein window length N is directly related to time-frequency resolution in that this gives rise to N frequency bins. Instead, in this case frequency resolution is not directly tied to window duration [17.20], but rather is a factor both of the window overlap as well as the damping value λ. These can therefore be adapted based on the transformation and synthesis requirements.

Now, we may build an adaptive framework on top of RESTFT using the Kalman filter, in order to parameterize the relative sine/noise quality of both attack and sustain such that one may reestimate these trajectories under time stretching and other time-varying transformations. Several algorithms were created in order to reach this goal, beginning with the creation of the initial Kalman framework.

17.3 Kalman Filter-Based Phase Vocoder

Taking full advantage of deeper sound transformations and control parameter estimation with the RESTFT representation requires the implementation of a Kalman filter (KF). The KF is an algorithm developed in order to optimally estimate the state of a linear dynamical system perturbed by Gaussian noise [17.21, 22]. Many variants have been established in order to extend the tracking, estimation and prediction properties of this filter to nonlinear and non-Gaussian systems [17.23]. As the STFT is a linear transform, the standard form of the KF suffices in order to model the sound process. The result is a model that allows one to estimate the magnitude and phase of each bin for every time step, and to extract *both the state and observation noise* for individual control and processing. As is noted in [17.24], in which the author creates a similar Kalman-based additive model specialized to damped percussive sounds, the state and observation noise sources relate to the transient noise and sustain noise (respectively) of the underlying sound signal. In this chapter we focus on these noise values, in order to define novel transformations of the noise component that adapt to varying time-stretching, such as one finds in scrubbing-based gestural control.

Now, the Kalman filter is a recursive process that consists of a time update and a measurement/observation update. The former first predicts future values of the state, and the latter then modifies this in order to provide an adjusted estimate for the current time step. More precisely, the a priori state estimate $\hat{\boldsymbol{s}}_n^-$ and the state covariance estimate $\mathbf{P}_n^-$ are conditioned on all prior values as

$$\begin{aligned}\hat{\boldsymbol{s}}_n^- &= \hat{\mathbf{A}}\hat{\boldsymbol{s}}_{n-1} + \mathbf{D}_{n-1}\boldsymbol{u}_{n-1} \quad \text{and} \\ \mathbf{P}_n^- &= \hat{\mathbf{A}}\mathbf{P}_{n-1}\hat{\mathbf{A}}^{\mathrm{T}} + \mathbf{Q}\,. \end{aligned} \tag{17.11}$$

These are the *time update* equations. Recall that the underlying system is perturbed by process and observation noise $\boldsymbol{w}_n$ and $\boldsymbol{v}_n$, which are zero-mean Gaussian white noise with variance $\mathbf{Q}$ and $\mathbf{R}$ respectively. The values $\hat{\boldsymbol{s}}_n$ and $\mathbf{P}_n$ are the a posteriori state and state covariance values, acquired from the following *measurement update* equations

$$\begin{aligned}\mathbf{K}_n &= \mathbf{P}_n^-\mathbf{B}^{\mathrm{T}}(\mathbf{B}\mathbf{P}_n^-\mathbf{B}^{\mathrm{T}} + \mathbf{R})^{-1}\,, \\ \hat{\boldsymbol{s}}_n &= \hat{\boldsymbol{s}}_n^- + \mathbf{K}_n(\boldsymbol{y}_n - \mathbf{B}\hat{\boldsymbol{s}}_n^-)\,, \end{aligned} \tag{17.12}$$

and

$$\mathbf{P}_n = (\mathbf{I} - \mathbf{K}_n\mathbf{B})\mathbf{P}_n^-\,. \tag{17.13}$$

A heuristic understanding of these equations would be to first consider the a priori state and covariance estimates as predictions based on all past values, with no noise disturbance; the updated estimates are defined by adjusting this initial estimate based on the influence of the *innovations sequence* $\boldsymbol{\epsilon}_n = (\boldsymbol{y}_n - \mathbf{B}\hat{\boldsymbol{s}}_n^-)$, which represents the difference between the observed value and the ideal noise-free observation. The degree of influence from the innovation is *tuned* by the so-called Kalman gain $\mathbf{K}_n$.

Following [17.24] we extract the residual from the state-space model in order to drive the resynthesis, though the extraction method differs as does the underlying representation (exponential STFT versus damped sinusoidal model). Thus the estimated process and observation noise sources are defined as

$$\hat{\boldsymbol{w}}_n = \hat{\boldsymbol{s}}_n - \hat{\mathbf{A}}\hat{\boldsymbol{s}}_{n-1} \quad \text{and} \quad \hat{\boldsymbol{v}}_n = \boldsymbol{y}_n - \mathbf{B}\hat{\boldsymbol{s}}_n\,. \tag{17.14}$$

We therefore have an estimate of the state $\hat{\boldsymbol{s}}_n$ – providing instantaneous estimates of magnitude and phase values – which gives rise to estimates of the excitation noise $\hat{\boldsymbol{w}}_n$ and the additive output noise $\hat{\boldsymbol{v}}_n$.

Finally, in order to provide the first a priori estimates, the model must be given initial conditions for the state and state covariance matrix. In the absence of any specific knowledge about the signal, we may safely define these such that

$$\hat{\boldsymbol{s}}_0 = \mathbf{0}_{2N\times 1} \quad \text{and} \quad \mathbf{P}_0 = \sigma^2\mathbf{I}_{2N\times 2N}\,. \tag{17.15}$$

The value $\sigma^2 \ll 1$ is the state covariance – it reflects the level of uncertainty in our initial state and governs the rate of convergence of the filter. If we were absolutely certain that the initial state was identically zero, we could let $\sigma^2 = 0$ as well. When the state of the input at time zero is unknown, one may include σ^2 as a model parameter that must be tuned. The other parameters are the noise covariance for the state and process noise values, the damping value λ for matrix $\hat{\mathbf{A}}$ and the state size N. The three noise covariance values $(\sigma^2, \mathbf{Q}, \mathbf{R})$ influence the ability of the tracking of spectral data, changing the relative amount of energy that will be present in the residual signals $\hat{\boldsymbol{w}}_n$ and $\hat{\boldsymbol{v}}_n$ as compared to the state estimate $\hat{\boldsymbol{s}}_n$. Meanwhile, the values λ and N together define the time/frequency resolution of the analysis as well as the computational power required (in the case of N). Therefore, while the Kalman-based RESTFT will produce a *perfect reconstruction* of an input signal if there are no modifications to the spectral or residual data, this is a parametric approach in which the relative contribution of each time-series $\hat{\boldsymbol{s}}_n$, $\hat{\boldsymbol{w}}_n$ and $\hat{\boldsymbol{v}}_n$ can be tuned by altering these model parameters.

17.3.1 Discussion

This KF-based RESTFT framework becomes musically interesting when used to reestimate state and noise values after applying time/frequency transformations such as pitch shifting or time stretching, or when one controls the relative level of each as well as the variance of the noise processes over time.

By adding a layer of control for such adaptive signal behavior, one is providing a bottom-up definition of mapping in which the general dynamic rules are described, rather than a static parametric mapping function. In this way, control and mapping are embedded in the signal definition, and happen *at the level of the sound transformation*.

Again, the RESTFT state size is defined by the number of phase vocoder spectral bins. This state size is a strong limiting factor for computation of the analysis, and so we have developed an algorithm adapted from the control theory literature, the so-called Chandrasekhar-type recursions [17.25], in order to make high-quality transformations more feasible by allowing for more reasonable window (i. e., state matrix) sizes, and therefore higher frequency resolution. The details of that algorithm may be found in the appendix.

17.4 Additive Layer and Higher-Level Architecture

The advantages gained by having an adaptive framework based on the Kalman filter can be extended from the phase vocoder to a higher-level, additive framework as well. In this implementation, rather than tracking the state of N evenly spaced frequency bins, we can model L sinusoidal partials ($L \ll N$ in general) so that

$$x_n = \sum_{k=1}^{L} \left[a_{k,n} \cos\left(\sum_{r=0}^{n} \omega_{k,r} + \phi_k \right) \right] . \qquad (17.16)$$

In terms of the underlying state-space representation, rather than the time-invariant state matrix defined by (17.10), the system here is nonstationary with time-varying state matrix $\mathbf{A}_n$ where the frequency values to each block diagonal component are given from the L frequencies $\{\omega_{1,n}, \ldots, \omega_{L,n}\}$. Therefore the matrix is of size $2L \times 2L$, where L is generally below 100. Note that the time resolution is therefore not bound implicitly to N samples in the same way as with the phase vocoder. Rather, the time-varying matrix $\mathbf{A}_n$ can be updated as much (i. e., each sample) or as little as desired, and commonly every H samples (i. e., hope size).

Now, note that the infinite-length exponential windows are still used in this analysis, preserving the time-frequency properties that exist with the RESTFT. This analysis stage does not do any partial tracking itself. Rather, peaks of interest are given to the analysis system at any time stamp deemed appropriate. Therefore, another method may be used in order to provide high-quality partial tracking (e.g., [17.26]) and this information can be fed to the Kalman-based system in order to provide a complete set of data on the suggested location of each partial. The additive Kalman layer, then, computes a *reanalysis* that extracts specific magnitude and phase values at each partial while producing the residual in the process. As with the KF-RESTFT, the sound's transient noise is captured by the process $\hat{\boldsymbol{w}}_n$ while the additive noise is represented by $\hat{\boldsymbol{v}}_n$. Therefore, rather than providing an improved partial-tracking algorithm, this additive implementation is geared towards providing a more refined analysis of the sine/noise energy decomposition. This serves the ultimate goal of improved noise-based transformations, in this case relative to a parametric additive framework. The overall sound processing architecture is depicted in Fig. 17.1, which shows the two analysis/synthesis schemes – KF-RESTFT and the Kalman-additive model – in parallel. It is not necessarily suggested that both schemes must be used in tandem, but rather this illustrates the manner in which they relate to one another in terms of model, input and control parameters. In each case, the input value x_n is sent into the analysis. In terms of the additive model, it is assumed that some other method is used for determining the list of peaks $F_\tau = \{\omega_{0,\tau}, \ldots, \omega_{L-1,\tau}\}$ in the *peak tracking* step. For both the additive and phase vocoder implementations, a set of model parameters must be provided. With the phase vocoder, these parameters are

$$\theta = \{\lambda, N, r, \sigma^2, q\} , \qquad (17.17)$$

where again the first two influence the time-frequency resolution and the latter three are the noise covariance values for the state, process and observation. For the additive model, the parameter set $\bar{\theta}$ varies slightly

$$\bar{\theta} = \{\lambda, L, H, r, \sigma^2, q\} , \qquad (17.18)$$

where L is the number of partials, and H is the hop size of the peak-tracking analysis. In general, these model parameters are not controlled online (and so have no inputs in this diagram), but they may potentially be changed in an adaptive fashion [17.27]. Once the Kalman-based additive or phase vocoder analyses are complete, they produce state estimates $\hat{\boldsymbol{s}}_n$ as

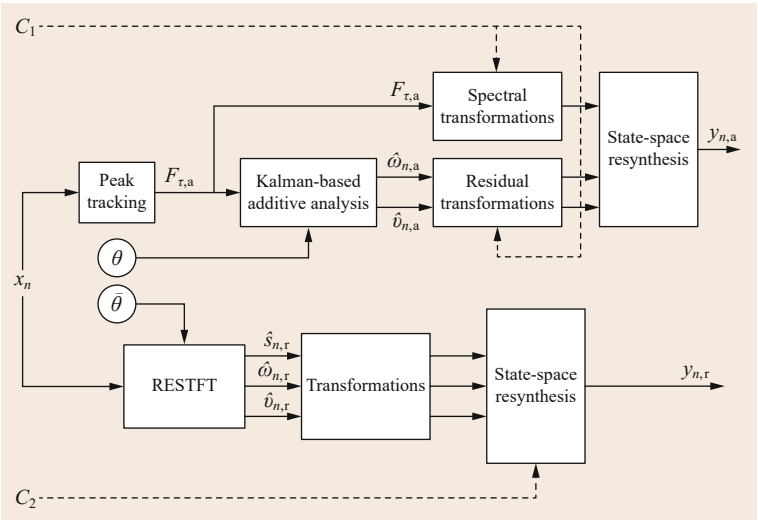


Fig. 17.1 Model architecture for Kalman-based additive and recursive exponential STFT models, shown here in parallel. Amplitude or phase values may be extracted for processing during the respective transformation stages

well as residuals $\hat{\boldsymbol{w}}_n$ and $\hat{\boldsymbol{v}}_n$. These are the values that are more likely controlled, or otherwise transformed before resynthesis. Thus, any possible control input would map directly into these values. In the case of the additive model, the frequency values F_τ may be transformed separately, so that partials may be independently processed. After time/frequency transformations are applied to the spectral frames of the state or to the residual values, the entire process is resynthesized using the state-space model, by (17.14).

17.5 Sound Transformations

Depending on the type of input signal, these two models are fairly robust to changes made simply to the noise covariances. However these values do exhibit an influence when the signal in question is time stretched, and so become indirect tuning parameters in adding textural qualities to the signal. It is shown in [17.24] that in fact it is only the *ratio* of the state/observation noise covariances that affect the resultant sound. This ratio changes the relative influence of the state or observation noise process over the output, and so this must be tuned in conjunction with the transformations – whether applied primarily to the input (state) or output (observation) residual. At the same time, time-varying modifications of the partials (in the case of the additive model) can be achieved without time stretching by focusing on the frequency trajectories and even the window damping value λ. Therefore, the transformations made possible with this approach can be broken down into three broad categories defined by:

- Textural effects achieved through a combination of processing the two noise time series and time stretching.
- Spectral effects achieved through modifying the state matrix.
- Cross-synthesis by combining the noise time series and state matrix values from different input signals.

As noted previously, the underlying sound model introduced in this chapter is a combination of a source-filter and an additive/phase vocoder approach. This fact is embedded in the above classification of transformations: the first can be considered a source-filter type of processing where the noise excitation is transformed and conditioned by the state matrix filter, while the second is primarily focused on additive transformations such as transposing all or certain partials over time by altering the damping parameter for the state matrix. This second class can also exhibit source-filter-type processing if blocks of the state matrix itself are processed (rather than its frequency/damping parameters), thereby changing the filter response of the model. The third class combines both of these approaches: the excitation of one sound may be altered and crossed with another sound whose spectral content has likewise been processed by altering its state matrix parameters. A full treatment of these classes is beyond the scope of this chapter, but the reader can see [17.28] for an expanded discussion.

17.6 Adaptive Control of Sound Transformations

The model developed here for sound transformation is novel and interesting in its own right, allowing for various textural and spectral effects. However another reason for reconsidering classic techniques in a new framework is not simply for new transformation possibilities, but also to build the normally *higher-level* control layer into the model itself, showing how mapping and control can be considered at the most atomic level. The use of the estimation framework afforded by the Kalman filter was chosen not only for residual estimation, but because it can allow for *consideration of control dynamics* during the analysis stage in parallel with sound dynamics. In this way, mapping becomes an expression of the way control dynamics covary with and influence sound dynamics, and vice versa. Therefore we will illustrate an augmentation of the sound model to include control dynamics. We will first look at an example based on a simple source-filter model to make the process clear and to illustrate the modeling of a time-domain autoregressive moving average (ARMA) process; a basic control system that can be used modularly to build more complex designs. After this we will return to a more musically motivated example based on the RESTFT and physically inspired control dynamics.

Recall that the RESTFT/additive model is based on a linear state-space equation, which expresses the temporal evolution of the STFT. While it is sufficient to use a linear model in order to describe the underlying analysis/synthesis system and certain sound transformations, as soon as one wishes to express musically interesting control of the spectral state-evolution, the system becomes nonlinear very quickly. In this case, one must expand to a nonlinear state space, and extend the modeling approach applied to RESTFT accordingly. This augmentation process begins with the modeling assumption that the state evolution and observation equations are governed by some nonlinear functions f and g such that

$$\begin{aligned} \boldsymbol{s}[n+1] &= f(\boldsymbol{s}[n], \boldsymbol{u}[n], \boldsymbol{w}[n]) \quad \text{and} \\ \boldsymbol{y}[n] &= g(\boldsymbol{s}[n], \boldsymbol{u}[n], \boldsymbol{v}[n]) \,, \end{aligned} \tag{17.19}$$

where all input/output parameters maintain their meaning from the previous state-space equations. A direct application of the previous model is not possible, as the Kalman filter is defined only for a linear system. However, several nonlinear extensions to this technique have been developed, with one of the more popular variants being the extended Kalman filter (EKF) [17.22]. The EKF has been used in many applications for nonlinear state estimation and control, making it a standard for certain areas such as in design of navigation systems.

The essential idea of this technique is to create a nominal linear trajectory $\bar{\boldsymbol{s}}_k$ in state space around the true state trajectory. This is calculated from the a posteriori estimate $\hat{\boldsymbol{s}}_{k-1}$ at a previous time step and without input noise via

$$\begin{aligned} \bar{\boldsymbol{s}}_k &= f(\hat{\boldsymbol{s}}_{k-1}, \boldsymbol{u}_{k-1}, \boldsymbol{0}_{N\times 1}) \quad \text{and} \\ \bar{\boldsymbol{y}}_k &= h(\bar{\boldsymbol{s}}_k, \boldsymbol{u}_k, 0) \,. \end{aligned} \tag{17.20}$$

Once this trajectory is calculated, a local linear estimate of the nonlinear state evolution is computed at each sample and the standard Kalman equations are applied to this linearized form. To achieve this, a Taylor series approximation of the nonlinear functions f and g are calculated as

$$\boldsymbol{s}_k \approx \bar{\boldsymbol{s}}_k + \bar{\mathbf{A}}_k(\boldsymbol{s}_{k-1} - \hat{\boldsymbol{s}}_{k-1}) + \bar{\mathbf{W}}_k w_{k-1} \,, \tag{17.21}$$

$$\boldsymbol{y}_k \approx \bar{\boldsymbol{y}}_k + \bar{\mathbf{H}}_k(\boldsymbol{s}_k - \hat{\boldsymbol{s}}_k) + \bar{\mathbf{V}}_k \boldsymbol{v}_{k-1} \,, \tag{17.22}$$

where $\bar{\mathbf{A}}$ and $\bar{\mathbf{H}}$ are the Jacobian matrices for functions f and g respectively, taken with respect to the state input. $\bar{\mathbf{W}}$ and $\bar{\mathbf{V}}$ are similarly computed, with respect to the input noise processes. Finally, with this local linearization at each instance – using the nominal state trajectory and Jacobian matrices – we may use the classic Kalman algorithm to produce state, residual and output estimates with appropriately modified versions of (17.11) through (17.13).

17.6.1 Example 1: Control of Modulated Source–Filter Model

In order to understand the paradigmatically different nature of this approach to control in comparison to standard views on parameter mapping, consider the following example system, which begins from the desire to design a Wacom tablet-based control of a modulated source-filter model. The first step is to express a second-order infinite impulse response (IIR) filter in a first-order recursion as a time-domain autoregressive (AR) process. The initial second-order expression begins with output $y[n]$ and filter coefficients $b_1[n], b_2[n]$ that relate to each other as

$$y[n] = b_1[n]y[n-1] + b_2[n]y[n-2] + a_0 x[n] \,, \tag{17.23}$$

where $x[n]$ is the input source signal and a_0 is an input gain. We may then express the state vector as

$$\boldsymbol{s}[n] = \begin{pmatrix} x_1[n] \\ x_2[n] \\ x_3[n] \\ x_4[n] \end{pmatrix} = \begin{pmatrix} y[n] \\ y[n-1] \\ b_1[n] \\ b_2[n] \,, \end{pmatrix} \tag{17.24}$$

which gives rise to the following state equation

$$\boldsymbol{s}[n+1] = \begin{pmatrix} x_3[n+1]x_1[n] + x_4[n+1]x_2[n] \\ x_1[n] \\ x_3[n] \\ x_4[n] \end{pmatrix} + \begin{pmatrix} a_0 x[n] \\ 0 \\ 0 \\ 0 \end{pmatrix} + \boldsymbol{w}[n]\,, \tag{17.25}$$

where again $\boldsymbol{w}[n]$ is a white noise process. Note that computing the state at time n requires knowledge of the state itself, and so a priori values $\hat{x}_3$ and $\hat{x}_4$ are used in practice.

While this equation already introduces nonlinearities due to the interaction between state values, it is more intuitive from a musical point of view to control the center frequency (f_c) and bandwidth (B_w) of the filter. Therefore, using knowledge of the relationship between these parameters for a second-order IIR filter, we can rewrite the state equation in a more intuitive manner – though at the cost of introducing more nonlinearities. Further, a two-dimensional control input $u[n] = \{u_1[n], u_2[n]\}$ is introduced to influence these two parameters, so that bandwidth changes linearly and the center frequency is sinusoidally modulated. To achieve this mapping we may include the control inside the state dynamics function. The state vector then becomes

$$\boldsymbol{s}[n] = \begin{pmatrix} y[n] \\ y[n-1] \\ f_c[n] \\ B_w[n] \end{pmatrix}, \tag{17.26}$$

while the new state equation is

$$\boldsymbol{s}[n+1] = \begin{pmatrix} -\phi(\hat{x}_3[n+1], \hat{x}_4[n+1])x_1[n] + \psi(\hat{x}_3[n+1])x_2[n] \\ x_1[n] \\ x_3[n] + c_1[n] \\ x_4[n]\cos(2\pi u[n]n) \end{pmatrix}, \tag{17.27}$$

where

$$-\phi(a,b) = -2\mathrm{e}^{\frac{-a\pi}{f_s}}\cos\left(\frac{2\pi b}{f_s}\right), \tag{17.28}$$

and

$$\psi(a) = \mathrm{e}^{\frac{-2a\pi}{f_s}}, \tag{17.29}$$

for any real value a, b where f_s is the audio sampling rate. This new state space has several aspects that a traditional Kalman filter cannot handle: nonlinear state dynamics, use of the a priori state values in the state vector and the inclusion of control input in the state dynamics function itself. However, with the use of an EKF variant, we can estimate the state and control, driving the system as desired.

Now, with all of the complexity embedded in the state dynamics, the observation on this state is a simple projection of $x_1[n]$ onto output $z[n]$

$$z[n] = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \boldsymbol{s}[n]\,. \tag{17.30}$$

For this example, the input *source* $x[n]$ is white noise, and so to align with our previous notation let $x[n] = w[n]$, while again the control input $u[n]$ comes from an external control source: the two-dimensional position data from a graphics tablet (though any continuous control could be used). Given this representation, the EKF derivation begins with the Jacobian matrix for the state equation f, which becomes

$$\bar{\mathbf{A}}_n = \begin{pmatrix} -\phi(\hat{x}_3[n], \hat{x}_4[n]) & \psi(\hat{x}_3[n]) & \frac{\partial f}{\partial x_3} & \frac{\partial f}{\partial x_4} \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{17.31}$$

where

$$\frac{\partial f}{\partial x_3} = -2\hat{x}_1\mathrm{e}^{\frac{-\pi\hat{x}_3}{f_s}}\cos\left(\frac{2\pi\hat{x}_4}{f_s}\right) - \frac{2\pi\hat{x}_2}{f_s}\mathrm{e}^{\frac{-2\pi\hat{x}_3}{f_s}}, \tag{17.32}$$

and

$$\frac{\partial f}{\partial x_4} = \frac{4\pi}{f_s}\mathrm{e}^{\frac{-\pi\hat{x}_3}{f_s}}\sin\left(\frac{2\pi\hat{x}_4}{f_s}\right). \tag{17.33}$$

After estimating the linearized state matrix, one can further estimate the state by the linear approximation as in (17.21), centered about the nominal state trajectory defined by (17.27). The matrices $\bar{\mathbf{W}}$ and $\bar{\mathbf{V}}$ are simply the identity multiplied by their initial noise covariances, and the matrix $\mathbf{H}$ remains a simple projection. At this point, we can use the Kalman filter loop in order to estimate the state and output, given the control input values c_1 and c_2. The a priori prediction at the next step is computed, in order to be used in the future calculation of the Jacobian matrix for the nonlinear state dynamics. Therefore, in practice we use a one-step prediction in order to drive the nonlinear control dynamics, rather than directly affect the B_w and f_c parameters. In this sense the system implements predictive control.

17.6.2 Example 2: Dynamic Control of Partials/Residual

We now return to the time-frequency models developed in this chapter in order to embed control dynamics within them, towards a more musically inspired example. The classic view of analysis/synthesis methods would consider control input to act on an intermediate transformation stage, between analysis and synthesis, as depicted in Fig. 17.1. However, if one were to consider control dynamics in the model during the analysis stage, then the process would be modeled as the outcome that occurs from the particular control input that is enacted at the time of analysis, such as in the previous example in which the process was considered as a frequency-modulated source-filter model. This difference must be considered when planning for the transformation stage in the context of analysis-synthesis modeling. In the original Kalman implementation, the analysis stage was simply modeled as a linear STFT, and the parameters were modified or substituted afterwards. Instead, we can *augment* the initial model to include potentially nonlinear control dynamics, which are then estimated during the analysis stage.

Control of State-Space Additive Model

In order to explore an implementation of this instrument design methodology, we draw upon the first author's performance practice of using a Wacom graphics tablet as an input control device for *scrubbing* type of interactions [17.1]. This example combines the ARMA modeling approach and an EKF estimator for control dynamics (as in Example 1, Sect. 17.6.1) with the Kalman-based additive sound model. That is, the latter model is used for the sound analysis/modeling stage, and is augmented with temporal dynamics for the control analysis/transformation stage. Therefore this approach can properly be considered as *analysis-augmentation-transformation-synthesis*.

While the full example is beyond the scope of this chapter (see [17.28] for more), the basic intent is to build a system that controls timbral/textural characteristics through gestural dynamics that are built into the control structure. We can work with the low-level spectrotemporal model as expressed by (17.9) in order to add a control for the separate levels of the sine versus residual level such as

$$\boldsymbol{s}[n+1] = \hat{\mathbf{A}}\boldsymbol{\alpha}\boldsymbol{s}[n] + \mathbf{D}_n\boldsymbol{u}[n+1] + \boldsymbol{\beta}\boldsymbol{w}[n]\,. \quad (17.34)$$

where $\boldsymbol{\alpha}$ is a $1\times N$ weighting for the partial values and $\boldsymbol{\beta}$ is similarly a $1\times N$ weighting for the input residual value. The overall perceptual effect depends heavily on the choice of the distribution of the individual weight values. For example, the individual $\boldsymbol{\alpha}$ weights can target certain partials, be drawn from another spectral envelope to create cross-synthesis, or control the odd/even balance, which can have a strong effect on the perceived timbre. Similarly, the relative balance of the partial/residual magnitude spectrum over time can strongly affect the dynamic textural properties of the sound.

Implementing Control Gestural Dynamics

Consider the use of the velocity from the Wacom tablet x-y values, conditioned by a leaky integrator, to guide certain aspects of a sound's resynthesis. This sort of gesture requires a sustained motion at a fairly high speed (which is a function of the integrator's response time) in order to maintain the influence. Therefore, oscillatory motion at regular speeds will result in a sustained value of the control output. A perceptually coherent mapping arises when this controls the stable part of the spectra, represented by the partial values of the state vector. For musical reasons, we will also map this motion to the playback time of the sample being scrubbed. For this example, assume we condition these dynamics on the x-position input, and call it c_1. In classic instrument design cases, the differentiator and integrator are thought of as *black boxes*, only considered in terms of their input/output effects. However, we can represent their dynamics as a recursive difference equation as well, with the velocity found simply by

$$\delta[n] = k_\mathrm{r}(c_1[n] - c_1[n-1])\,, \quad (17.35)$$

where k_r is the rate of the control signal. At this point, we can apply the leaky integrator to the velocity, which can be expressed as

$$L[n] = \frac{1}{k_\mathrm{r}}\delta[n] + L[n-1]\,2^{\frac{-1}{k_\mathrm{r}\lambda}}\,, \quad (17.36)$$

where λ is the response time of the integrator. Therefore, from a signal processing point of view, this system is a first-order filter acting on the velocity of control input c_1. In order to design another expressive control gesture for y-position input c_2 that has considerably different dynamics, we can extend to a second-order filter and add resonance to the system. Generally speaking we can condition this input by the equation

$$z[n] = b_1 z[n-1] + b_2 z[n-2] + b_0 c_2[n]\,, \quad (17.37)$$

where z is the system output and b_1 and b_2 are general filter coefficients while b_0 is the input gain. The temporal behavior of this system depends heavily on these latter three coefficients. Further, to leverage the dynamics of this system we need to express it in a one-step

recursion, which we can do by defining an intermediate equation as

$$z[n] = z[n-1] + \frac{1}{k_\mathrm{r}} d[n] \,, \tag{17.38}$$

with new intermediate variable d. Taking this further, define the general filter coefficients as

$$b_0 = \frac{a_1}{k_\mathrm{r}^2} \,, \quad b_1 = 2 - \frac{a_1 - a_2}{k_\mathrm{r}^2} \,,$$
$$b_2 = \frac{a_2}{k_\mathrm{r}^2} - 1 \,, \tag{17.39}$$

where again a_0, a_1 and a_2 are tunable filter coefficients. Inserting (17.38) and (17.39) into the general second-order equation results in a first-order recursion of two variables defined as

$$d[n] = d[n-1] + \frac{1}{k_\mathrm{r}} (a_1 (c_2[n] - z[n-1]) - a_2 d[n-1]) \,. \tag{17.40}$$

Taken together, (17.38) and (17.40) constitute a state-space form of the general second-order system with state variables d and z. Further, this particular choice of b coefficients was made due the dynamics that it allows for: rather than building up energy, the conditioned response here is such that fast input control changes result in an oscillation that converges to the given value (provided a coefficient values are subject to certain boundary constraints). From a perceptual control point of view, this behavior is an interesting addition to the instrument design when used for control of the input residual gain. With such a mapping, sharp control actions result in a high excitation noise level, having amplitude modulation. Gradual movements have little or no perceptible modulations, depending on the filter coefficients. In the context of control dynamics, *Menzies* [17.29] has referred to this as a resonant-follower.

While these two dynamical control systems can be interpreted as signal processing tools, they also possess a more intuitive physical interpretation. The leaky integrator is in fact a spring-damper system, with the response time parameter acting as a damping coefficient on the system. Similarly, the second-order resonant filter expresses a mass-spring-damper (MSD) system, which can be checked by discretizing the physical equation and setting the mass to 1. Thus we can provide a physically grounded expression by setting the coefficients to $a_1 = K_\mathrm{s}$ and $a_2 = K_\mathrm{d}$, which are spring and damping constants respectively. Further, the state variables z and d represent position and velocity, where this relationship can be seen by comparing (17.35) and (17.38). Therefore the perceptual effect of building up energy in the former case or of modulating response in the later is in fact a product of designing the physics of a real system that has a resistive (spring-damper) and resonant (MSD) feel to it. When designing this physically inspired interaction *in tandem with* the signal-focused STFT/additive sound model the result is a hybrid approach that falls between signal and physical modeling.

In order to augment the overall system to include these elements, however, we need to modify the expression a bit. If we kept (17.40) and (17.38) in their current form, the output would depend on an intermediate parameter at the same time step and an a priori estimate would need to be used. This can be avoided by combining these equations so that

$$z[n] = z[n-1] + \left(1 - \frac{K_\mathrm{d}}{k_\mathrm{r}}\right) d[n-1] + \frac{K_\mathrm{s}}{k_\mathrm{r}} (c_2[n] - z[n-1]). \tag{17.41}$$

This results in a single expression of the second-order control system in a first-order recursion.

Augmented State Space: Time–Domain Control, Time/Frequency Sound Model

With this, the mapping and control structures are fully defined, as seen in Fig. 17.2. However, we must integrate these new equations with the sound model's state space. While the essence of what we want to achieve may be expressed by

$$\boldsymbol{s}[n+1] = \hat{\mathbf{A}}\mathbf{L}[n]\boldsymbol{s}[n] + \mathbf{D}_n \boldsymbol{u}[n+1] + z[n]\boldsymbol{w}[n] \,, \tag{17.42}$$

this does not represent a full model of the dynamics, as the control structure must be fully represented in the state equation. We can achieve this by augmenting the state vector to include the three control process variables and the mapping dynamics. At this point, the system is thus modeled as an *observation of input/output values, leading to an estimate of sound/control dynamics which in turn are synthesized to produce the final audio output*.

The control-augmented dynamics model is now nonlinear due to the interaction between state variables. Because of this, we must use the EKF structure in order to take advantage of the underlying Kalman framework for estimating the dynamics. For ease of reading, the full derivation is presented in the appendix. In short, after taking into consideration the nonlinear interaction between control and sound dynamics, we have arrived

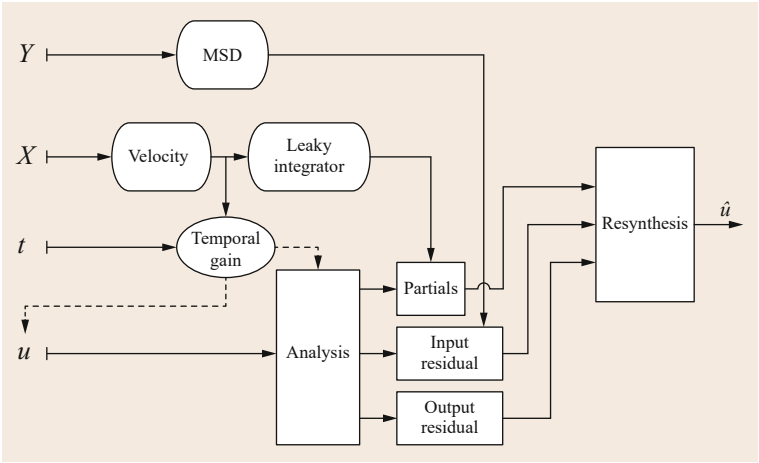


Fig. 17.2 The mapping structure for the given instrument design. *X*-*Y* tablet values feed the physically inspired dynamics of the leaky integrator and mass-spring-damper systems respectively. Time (of contact) is an implicit input that controls the *scrubbing*. This control is outside the model and so is not discussed in this chapter, but note that it may control either sound input u or the model parameters by way of some time-scaling algorithm such as a subtractive optimally localized averages technique. Finally, the analyzed sound u is an input to the analysis process, which is then affected by the control dynamics before being resynthesized into output $\hat{u}$

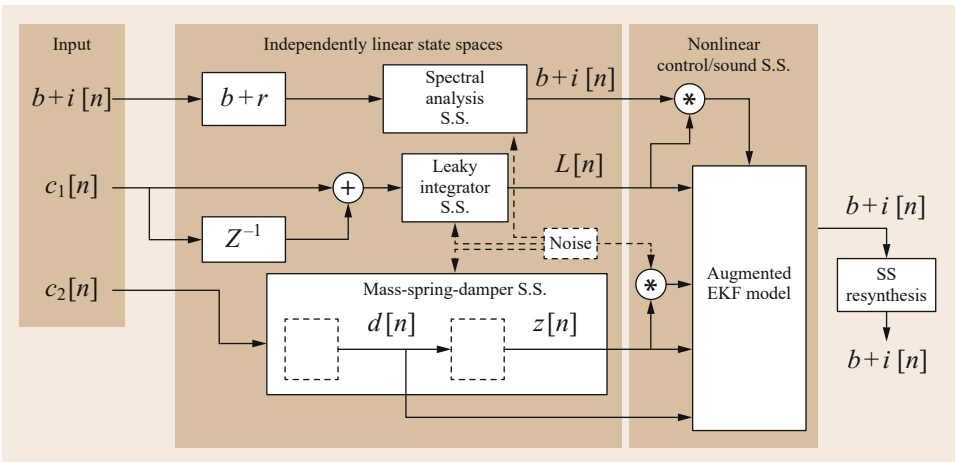


Fig. 17.3 Nonlinear state-space system for controlling time/frequency model. Analyzed input sound and control values are observed, while linear dynamical systems for control/sound are combined into a nonlinear control-sound dynamics model, augmented by an EKF. This dynamics model predicts the control/sound state value, and this is resynthesized to produce new sound output

at a complete instrument design that takes into account gestural dynamics. The layout for the entire instrumental system is depicted in Fig. 17.3. This example serves to show the process of designing an instrument that builds control and sound dynamics together in an estimation framework for control/sound analysis/synthesis. After designing the dynamics, state-space and extended Kalman filtering, the system's runtime behavior is as follows:

- Sample the control and sound input simultaneously.
- Drive the respective control/sound state equations with these values.
- Estimate the intermediate control/sound state variables using the EKF based on a nonlinear control/sound interaction model.
- Use the estimated state values to drive this nonlinear model, thereby resynthesizing new sound output $\boldsymbol{y}[n]$.

17.7 Chapter Summary

This chapter has presented a canonical way to build sound transformation instruments using a combination of ARMA and time/frequency modeling approaches, the state-space representation for dynamics, the Kalman filter for estimation and the EKF for nonlinear control. The combined use of these tools has led to both a set of novel algorithms as well as a working methodology for embedding control-sound dynamics *at the sound level* of an instrumental system, so that control structures can be built from the bottom up in a temporally focused way. The issues of sound representation and transformation were thus intersected with that of mapping and control structure design. We thus examined mapping from the point of view of a constraint on a dynamical system, expressing the co-evolution of parameters. This work shifts the focus from modeling parameter space to modeling the temporal response of action-to-sound coupling in a way that combines physically modeled dynamics with abstract time/frequency signal models. Going beyond the analysis-transformation-synthesis paradigm, the chapter introduced augmentation with nonlinear control dynamics, thus allowing for modeling of the control/sound interaction in a way that may be customized to contemporary digital musical performance contexts.

From a design point of view, in using the framework developed in this chapter the actual output may adhere closely or loosely to the control input, depending on the noise statistics among other tuning parameters. One may tune the degree of adherence to the dynamical systems model or add additional control values into the state vector. While it is beyond the scope of this chapter, some future directions include a more robust tuning by adaptively estimating the noise statistics at the same time as the state dynamics [17.27] or using the control dynamics residual itself as an interesting signal to be sonified or mapped to additional sound parameters.

17.A Appendix 1: Chandrasekhar Implementation

The computation of the state estimate in (17.12) is the essential step in which the Kalman gain – adapted by the updated state covariance – determines the amount of influence that the innovation sequence (i. e., the *novel* information) will have on the newly predicted state. This, as well as the Kalman gain of (17.12) may be factored differently, so that

$$\hat{\mathbf{K}}_n = \hat{\mathbf{A}}\mathbf{P}_n^{-}\mathbf{B}^{\mathrm{T}} , \tag{17.43}$$

$$\hat{\mathbf{R}}_n = (\mathbf{B}\mathbf{P}_n^{-}\mathbf{B}^{\mathrm{T}} + \mathbf{R}) , \tag{17.44}$$

$$\hat{s}_n = \hat{s}_n^{-} + \hat{\mathbf{K}}_n(\hat{\mathbf{R}}_n)^{-1}\boldsymbol{\epsilon}_n . \tag{17.45}$$

The quantity $\hat{\mathbf{R}}_n$ is the variance matrix of the innovations sequence $\boldsymbol{\epsilon}_n$. Using this factorization, it was shown in [17.25] that the matrices $\hat{\mathbf{K}}_n$ and $\hat{\mathbf{R}}_n$ may be computed by utilizing intermediate operations that act on matrices having a substantially smaller rank – and so possibly much less computation time. These added operations are referred to as Chandrasekhar-type recursions. The RESTFT fits this case and so was rewritten in this form. In particular, intermediate matrices $\mathbf{Y}_n$ and $\mathbf{M}_n$ arise, which are defined as

$$\mathbf{Y}_n = [\hat{\mathbf{A}} - \hat{\mathbf{K}}_n(\hat{\mathbf{R}}_n)^{-1}\mathbf{B}]\mathbf{Y}_{n-1} , \tag{17.46}$$

$$\mathbf{M}_{n+1} = \mathbf{M}_n + \mathbf{M}_n\mathbf{Y}_n^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}(\hat{\mathbf{R}}_n)^{-1}\mathbf{B}\mathbf{Y}_n\mathbf{M}_n , \tag{17.47}$$

which then are used to redefine the Kalman equations as a recursion with

$$\hat{\mathbf{K}}_{n+1} = \hat{\mathbf{K}}_n + \hat{\mathbf{A}}\mathbf{Y}_n\mathbf{M}_n\mathbf{Y}_n^{\mathrm{T}}\mathbf{B}^{\mathrm{T}} , \tag{17.48}$$

$$\hat{\mathbf{R}}_{n+1} = \hat{\mathbf{R}}_n + \mathbf{B}\mathbf{Y}_n\mathbf{M}_n\mathbf{Y}_n^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}t . \tag{17.49}$$

These new equations for $\mathbf{Y}_n, \mathbf{M}_n, \hat{\mathbf{K}}_n$, and $\hat{\mathbf{R}}_n$ are of size $N\times\alpha$, $\alpha\times\alpha$, $N\times p$ and $p\times p$ respectively, where p is the output dimension and α is the rank of a matrix equation defined by the initial state covariance and Kalman gain matrices [17.25]. As the system observation is of the scalar audio value in our case, the matrix $\hat{\mathbf{K}}_n$ reduces to a column vector while $\hat{\mathbf{R}}_n$ is simply a scalar. Thus the inversion of this latter value in (17.47) can be replaced by a simple multiplicative inverse, further speeding up the calculations, which is particularly time saving in the Matlab environment. Finally it can be shown – in this particular case of scalar observation – that the value $\alpha = 1$ as well, and so the overall speed for the calculation of the state estimate and related values is improved dramatically. Further, note that certain quantities such as $\mathbf{M}_n\mathbf{Y}_n^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}$ are used several times, so that they can be stored in memory and only need be computed once.

17.A.1 Initial Conditions

The initial values for the matrices $\hat{\mathbf{K}}_n$ and $\hat{\mathbf{R}}_n$ are simply derived as

$$\mathbf{R}_0 = \mathbf{R} + \mathbf{B}\mathbf{P}_0\mathbf{B}^{\mathrm{T}} , \tag{17.50}$$

$$\mathbf{K}_0 = \hat{\mathbf{A}}\mathbf{P}_0\mathbf{B}^{\mathrm{T}} . \tag{17.51}$$

However those for $\mathbf{Y}_n$ and $\mathbf{M}_n$ require further derivation. It is noted in [17.25] that in particular cases the initial values $\mathbf{Y}_0$ and $\mathbf{M}_0$ assume a simple form. When the value $\hat{\mathbf{s}}_0$ is known with a high degree of certainty – and one may safely assume that $\sigma^2 = 0$ – then the values reduce to

$$\mathbf{Y}_0 = \mathbf{I}_{2N\times 2N} , \tag{17.52}$$

$$\mathbf{M}_0 = \mathbf{Q} . \tag{17.53}$$

At the same time, it is proven that if the state matrix $\hat{\mathbf{A}}$ is a stability matrix – meaning that its eigenvalues lie within the unit circle and the state sequence converges to a stationary process – then the initial conditions take another form. Our sound model is a limit case of this, and so we include this as a possible scenario in the algorithm. In this case, the initial values become

$$\mathbf{Y}_0 = \hat{\mathbf{A}}\mathbf{P}_n\mathbf{B} \tag{17.54}$$

$$\mathbf{M}_0 = (r + \mathbf{B}\mathbf{P}_n\mathbf{B}^{\mathrm{T}})^{-1} . \tag{17.55}$$

As these two scenarios are likely but not guaranteed outcomes for any given audio signal, we consider them both and leave this as an option when conducting a given analysis. That said, the latter case is generally a sufficient solution.

Now, this entire derivation has been based on the assumption that we have a stationary process, with $\hat{\mathbf{A}}$ constant over time. This is the case with our implementation based on the STFT. However, if we wish to extend this into an additive implementation then the Chandrasekhar recursions may not be used. Fortunately this is not a concern, as the need for this fast algorithm arises precisely in the case of having large state sizes brought on by the STFT. Therefore, fast implementations are used when they are needed, while the standard form still works in the additive case where state sizes are more manageable (e.g., N less than 100).

17.B Appendix 2: Example 2 EKF Derivation

The derivation of the full control structure from the example of Sect. 17.6 is calculated as follows here. It begins by augmenting the state vector to include the three control process variables. Let

$$\boldsymbol{s}_c[n] = \begin{pmatrix} \boldsymbol{s}[n] \\ L[n] \\ d[n] \\ z[n] \end{pmatrix} , \tag{17.56}$$

which results in the new state equation defined by

$$\boldsymbol{s}_c[n+1] = \mathbf{A}_{c,n} f(\boldsymbol{s}_c[n]) + \mathbf{D}_{c,n}\boldsymbol{u}_c[n] + z[n]\boldsymbol{w}[n] , \tag{17.57}$$

where

$$f(\boldsymbol{s}_c[n]) = \begin{pmatrix} L[n-1]\boldsymbol{s}[n-1] \\ L[n-1]\, 2^{\frac{-1}{k_r\lambda}} \\ d[n-1] - \dfrac{1}{k_r}(K_s z[n-1] + K_d d[n-1]) \\ z[n-1] + \left(1 - \dfrac{K_d}{k_r}\right) d[n-1] - \dfrac{K_s}{k_r} z[n-1] \end{pmatrix} . \tag{17.58}$$

Note that the scalar $L[n-1]$ is multiplied across all values of the state $\boldsymbol{s}[n-1]$, so that the augmented state vector that results from this function is of size $N+3$ for state size N. Further $\mathbf{A}_{c,n}$ is the block-diagonal matrix $\hat{\mathbf{A}}_n$ with a 3×3 identity matrix added to the diagonal, with proper zero padding of the first columns and rows of this new state matrix in order to make it well defined.

Meanwhile, input vector $\boldsymbol{u}_c[n]$ accounts for the current sound input value $\boldsymbol{u}[n]$ as well as the control inputs, so that

$$\boldsymbol{u}_c[n] = \begin{pmatrix} \boldsymbol{u}[n] \\ c_1[n] \\ c_1[n-1] \\ c_2[n] \end{pmatrix} . \tag{17.59}$$

Likewise $\mathbf{D}_{c,n}$ extends the matrix $\mathbf{D}_n$ in order to account for the input value's contribution to the control dynamics. Thus the control matrix $\mathbf{C}$ is added to the block diagonal of $\mathbf{D}_{c,n}$, where

$$\mathbf{C} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & K_s/k_r \\ 0 & 0 & K_d/k_r \end{pmatrix} , \tag{17.60}$$

with zero padding of the first columns and rows of $\mathbf{D}_{c,n}$ in the same manner as applied to $\mathbf{A}_{c,n}$. Similarly $\boldsymbol{u}_c[n]$ is the input value $\boldsymbol{u}[n]$ padded with zeros, thereby finalizing this new form of the process equation. The observation $\mathbf{B}_c$ is essentially the same as in (17.1) except that it acts on the augmented state vector $\boldsymbol{s}_c$, and thus has necessary zero padding and identity elements prepended to the added rows. This new matrix is needed to account for the two control values, which are added to the observation that still includes audio output $\boldsymbol{y}[n]$.

17.B.1 EKF for Control/Sound Integration

We must linearize about a nominal state trajectory

$$\bar{\boldsymbol{s}}_c[n] = \mathbf{A}_{c,n-1} f(\hat{\boldsymbol{s}}_c[n-1]) + \mathbf{D}_{c,n-1}\boldsymbol{u}_c[n-1] \quad (17.61)$$

and calculate the nominal state based on the a posteriori estimate of the last time step; we also have control dynamics that are taken into account in making the state prediction. Given this estimate we can linearize the state trajectory in a similar fashion to the expression of (17.21). The primary difference is that we must take the control into account for the state noise process, and thus we arrive at

$$\boldsymbol{s}_c[n] \approx \bar{\boldsymbol{s}}_c[n] + \bar{\mathbf{A}}_n(\boldsymbol{s}_{n-1} - \hat{\boldsymbol{s}}_{n-1}) + \bar{\mathbf{W}}_n \boldsymbol{w}[n]\,, \quad (17.62)$$

where the two Jacobians are found to be

$$\bar{\mathbf{A}}_n = \begin{pmatrix} \hat{\mathbf{L}}[n-1] & \hat{\boldsymbol{s}}[n-1] & 0 & 0 \\ 0 & 2^{\frac{-1}{\lambda k_r}} & 0 & 0 \\ 0 & 0 & 1+K_\mathrm{d} & 0 \\ 0 & 0 & 1-\dfrac{K_\mathrm{d}}{k_\mathrm{r}} & 1-\dfrac{K_\mathrm{s}}{k_\mathrm{r}} \end{pmatrix} \quad (17.63)$$

for the process linearization and

$$\bar{\mathbf{W}}_n = \begin{pmatrix} \hat{z}[n-1] & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (17.64)$$

for the linearization of the input residual control matrix. (In practice, the value $\hat{\boldsymbol{L}}[n]$ from $\bar{\mathbf{A}}_n$ is actually an $N\times N$ matrix comprised of copies of $\hat{\boldsymbol{L}}[n]$ where N is the number of partials, and the same for the $\hat{z}$ entry of the latter matrix. We express it in this simplified form for clarity of notation, without any loss of generality.) Using the newly linearized approximation from (17.62) and the already-linear observation equation, we can use the standard Kalman loop – in this case as it was defined for analysis of the additive sound model.

References

17.1 D. Van Nort, M. Wanderley, P. Depalle: Mapping control structures for sound synthesis: Functional and topological perspectives, Comput. Music J. **38**(3), 6–22 (2014)

17.2 M. Wanderley (ed.): Mapping strategies in real-time computer music, Organised Sound **7**(2) (2002)

17.3 D. Wessel: Timbre space as a musical control structure, Comput. Music J. **3**(2), 45–52 (1979)

17.4 X. Serra, J.O. Smith: Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition, Comput. Music J. **14**(4), 14–24 (1990)

17.5 T. Kailath: *Linear Systems* (Prentice-Hall, Englewood Cliffs 1980)

17.6 J.L. Flanagan, D.I.S. Meinhart, R.M. Golden, M.M. Sondhi: Phase vocoder, J. Acoust. Soc. Am. **38**(5), 939–940 (1965)

17.7 J.B. Allen, L.R. Rabiner: A unified approach to short-time Fourier analysis and synthesis, Proc. IEEE **65**, 1558–1564 (1977)

17.8 J.A. Moorer: The use of the phase vocoder in computer music applications, J. Audio Eng. Soc. **26**(1/2), 42–45 (1978)

17.9 M. Dolson: The phase vocoder: A tutorial, Comput. Music J. **10**(4), 14–27 (1986)

17.10 G.H. Hostetter: Recursive discrete fourier transformation, IEEE Trans. Audio Speech Signal Process. **28**(2), 184–190 (1980)

17.11 D. Arfib, F. Keiler, U. Zoelzer, V. Verfaille, J. Bonada: Time-frequency processing. In: *DAFX: Digital Audio Effects*, 2nd edn., ed. by U. Zoelzer (Wiley, Chichester 2011)

17.12 D. Van Nort, P. Depalle: A stochastic state-space phase vocoder for synhesis of roughness. In: *Proc. Int. Conf. Digit. Audio Efects (DAFx 06)* (2006) pp. 177–180

17.13 S. Dubnov, N. Tishby, D. Cohen: Influence of frequency modulating jitter on higher order moments of sound residual with applications to synthesis and classification. In: *Proc. Int. Comput. Music Conf. (ICMC 96)* (1996) pp. 378–385

17.14 H.D. Thornburg, R.J. Leistikow: Analysis and resynthesis of quasi-harmonic sounds: An iterative filterbank approach. In: *Proc. Int. Conf. Digit. Audio Efects (DAFx 03)* (2003)

17.15 Y. Qi, T.P. Minka, R.W. Picard: Bayesian spectrum estimation of unevenly sampled nonstationary data. In: *Proc. Int. Conf. Acoust. Speech Sig. Process. (ICASSP 02)* (2002)

17.16 A.T. Cemgil, S.J. Godsill: Probabilistic phase vocoder and its application to interpolation of missing values in audio signals. In: *Proc. 13th Eur. Sig. Process. Conf., Antalya* (2005)
17.17 P. Depalle, S. Tassart: State space sound synthesis and a state space synthesiser builder. In: *Proc. Int. Comput. Music Conf. (ICMC 95)* (1995)
17.18 X. Rodet, D. Matignon, P. Depalle: State space models for wind-instrument synthesis. In: *Proc. Int. Comput. Music Conf. (ICMC 92)*, Vol. 277–280 (1992)
17.19 S. Tomazic: On short-time fourier transform with single-sided exponential window, Signal Process. **55**(2), 141–148 (1996)
17.20 S. Tassart: Infinite length windows for short-time fourier transform. In: *Proc. Int. Comput. Music Conf. (ICMC 98)* (1998)
17.21 R.E. Kalman: A new approach to linear filtering and prediction problems, Trans. ASME J. Basic Eng. **82**(1), 35–45 (1960)
17.22 G. Welch, G. Bishop: An Introduction to the Kalman Filter, Tech. Rep. TR 95-041 (University of North Carolina, Department of Computer Science, Chapel Hill 1995)
17.23 M.S. Grewal, A.P. Andrews: *Kalman Filtering: Theory and Practice* (Prentice Hall, Englewood Cliffs 1993)
17.24 H. Thornburg: *Detection and Modeling of Transient Audio Signals with Prior Information*, Ph.D. Thesis (Stanford University, Palo Alto 2005)
17.25 M. Morf, G.S. Sidhu, T. Kailath: Some new algorithms for recursive estimation in constant, linear, discrete-time systems, IEEE Trans. Automat. Contr. **19**, 315–323 (1974)
17.26 P. Depalle, G. Garcia, X. Rodet: Tracking of partials for additive sound synthesis using hidden markov models. In: *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP 93)*, Vol. 1 (1993) pp. 225–228
17.27 A. Moghaddamjoo, R.L. Kirlin: Robust adaptive kalman filtering with unknown inputs, IEEE Trans. Acoust. Speech Signal Process. **37**(8), 1166–1175 (1989)
17.28 D. Van Nort: *Modular and Adaptive Control of Sound Processing*, Ph.D. Thesis (McGill University, Montreal 2010)
17.29 D. Menzies: Composing instrument control dynamics, Organised Sound **7**(3), 255–266 (2002)

18. Wave Field Synthesis

Tim Ziemer

Wave field synthesis enables acoustic control in a listening area by systematic regulation of loudspeaker signals on its boundary. This chapter starts with an overview including the history of wave field synthesis and some exemplary installations. Next, the theoretic fundamentals of wave field synthesis are detailed. Technical implementations demand drastic simplifications of the theoretical core, which come along with restrictions of the acoustic control as well as with synthesis errors. Simplifications, resulting synthesis errors as well as the working principles of compensation methods and their effects on the wave field are extensively discussed. Finally, the current state of research and development is addressed.

18.1 Overview

Wave field synthesis is the concept of creating a desired sound field within an extended listening area. The objective is to control the sound field in space rather than discrete audio channels. This way of thinking has the objective to overcome the shortcoming of conventional audio systems, which give minor control over source width, distance and lateral sources as well as have the limitation of one single listening position, the *sweet spot*. Early wave field synthesis attempts arose in the 1930s [18.1], the time in which also the first stereo concepts came up. The first explicit formulations of a *wave front synthesis* concept come from the Delft University of Technology i. e., *Berkhout* et al. entitled *acoustic control* [18.2–4]. The idea of full acoustic control has nearly unlimited possibilities: a virtual singing bird could be placed on the shoulder of one listener. And this effect is not only audible for this very listener; everybody would localize the birdsong as coming from her shoulder. A virtual helicopter could circle above their heads and a virtual car could dash towards a listener. But as the sound field of the actual physical happening is recreated, the sound is *virtual* in terms of artificial creation, but absolutely real in terms of physical acoustics. Thus, a wave field synthesis system could make an electric piano sound as broad as a concert grand piano, a car like the Boston Symphony Hall, the cinema like a solutional cave or a living room like a soccer stadium.

Of course, wave field synthesis can also be used as the acoustic counterpart to stereoscopic video in movies and virtual reality applications. Acoustic control also delivers pragmatic solutions: speech intelligibility of public announcements in stadiums could be improved by creating one plane wave instead of using delay lines. The room acoustics of multipurpose halls could be artificially adjusted to the specific performance. Noise disturbance and unwanted echoes could be eliminated by either creating sound waves directed only towards the target receiver or by creating cancellation signals for quiet zones. These objectives sound almost too good to be true and of course comprehensive acoustic

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_18

control is a challenging goal. At the moment concessions have to be made. Several wave field synthesis approaches enable acoustic control to a certain degree. The terms *wave field synthesis* (WFS) and *sound field synthesis* (SFS) comprise concepts of wave front synthesis, *higher-order ambisonics* (HOA) [18.5–11], sound radiation synthesis [18.12–15] and other methods [18.16–20].

The research project CARROUSO, funded by the European Community from 2001 to 2003, brought universities and industry partners together to achieve huge steps in the development and applicability of wave front synthesis [18.21–23] paving the way for market-ready audio systems. Today, application areas of wave field synthesis systems are especially music playback and concert hall auralization [18.19, 23–25], network music performance and virtual conference rooms [18.26, 27], listening room compensation [18.28] and research [18.8, 9, 29]. Although wave field synthesis is still at a stage of research and development, several systems are already in use e.g., in cinemas, theaters, clubs, themed environments and universities. In the auditorium of the Technical University in Berlin a wave field synthesis system amplifies the lecturer without being too loud in the frontal region and without creating the annoying echo, typical for delay lines in PA systems. Additionally, a number of music compositions has been created for this system. It includes almost 1000 loudspeakers, individually controlled by a cluster of computers, just to synthesize the sound field in the horizontal plane [18.25]. This is a typical order of magnitude for large rooms. In cooperation with Fraunhofer IDMT the Audi Q7 was equipped with a 62-channel wave field synthesis system to auralize the room acoustics of concert halls or churches for driver and passengers [18.30]. Systems for entertainment are installed in TV sound boards, clubs, cinemas and at festivals to create both natural and surreal immersive, spatial sound experiences for every single listener [18.25, 31].

Implementations come together with problems that need to be compensated, resulting in limitations concerning the synthesis precision, extent of the listening area, number and potential positions of virtual sources, complexity of sound radiation patterns, auralization of virtual room acoustics or alike. The name already implies that the theoretical fundamentals of wave field synthesis are the wave equation and its solutions with appropriate boundary conditions. These are discussed next, followed by an extensive examination of wave front synthesis, the most common sound field synthesis approach. Finally, alternative approaches as well as the current state of research and development are addressed.

18.2 Wave Equation and Solutions

18.2.1 Homogeneous Wave Equation

The first base equation of the wave field is Euler's equation of motion

$$\rho_0 \frac{\partial \boldsymbol{v}(\boldsymbol{x},t)}{\partial t} = -\nabla p(\boldsymbol{x},t)\,, \tag{18.1}$$

describing the flow of frictionless fluids by means of time t, position vector $\boldsymbol{x}$, particle velocity vector $\boldsymbol{v}$, pressure p, ambient density ρ_0 and the nabla operator ∇. In Cartesian coordinates the following is valid

$$\boldsymbol{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

$$\boldsymbol{v} = \begin{bmatrix} u(x) \\ v(y) \\ w(z) \end{bmatrix}$$

$$\nabla \equiv \frac{\partial}{\partial \boldsymbol{x}} = \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z}\,. \tag{18.2}$$

The second base equation of the wave field is a form of the continuity equation

$$c^2 \rho_0 \nabla \boldsymbol{v}(\boldsymbol{x},t) + \frac{\partial p(\boldsymbol{x},t)}{\partial t} = 0 \tag{18.3}$$

with the propagation velocity c. Differentiating (18.3) with respect to time and replacing the velocity term by the right side of the equation of motion, (18.1), yields the homogeneous wave equation for pressure

$$\nabla^2 p(\boldsymbol{x},t) - \frac{1}{c^2}\frac{\partial^2 p(\boldsymbol{x},t)}{\partial t^2} = 0\,, \tag{18.4}$$

where ∇^2 is the Laplace operator

$$\nabla^2 \equiv \frac{\partial^2}{\partial \boldsymbol{x}^2} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\,. \tag{18.5}$$

By differentiating the continuity equation with respect to $\boldsymbol{x}$ and the equation of motion with respect to t yields the homogeneous wave equation for velocity

$$\nabla^2 \boldsymbol{v}(\boldsymbol{x},t) - \frac{1}{c^2}\frac{\partial^2 \boldsymbol{v}(\boldsymbol{x},t)}{\partial t^2} = 0\,. \tag{18.6}$$

Equations (18.4) and (18.6) look very similar and can have the same solution. But sound pressure and particle velocity are not the same. Their relationship is described in (18.1). The sound pressure gradient is proportional to the temporal derivative of the particle velocity i. e., the particle acceleration. Sound pressure p and sound velocity $\boldsymbol{v}$ are perturbations of the state of equilibrium that propagate as sound waves. Solutions of the wave equation are called *sound field* or *wave field*. The equations assume the conditions listed in Table 18.1 [18.32–34].

Table 18.1 Preconditions for the wave equations

1.	The propagation medium is homogeneous
2.	The medium is quiescent and vortex free
3.	State changes are adiabatic i. e., no heat interchange between areas of low pressure and areas of high pressure due to the rapid movement of the particles
4.	Pressure and density perturbations are small compared to static pressure and density
5.	Relationships in the medium are subject to linear differential equations
6.	The medium exhibits no viscosity
7.	The medium is source-free

18.2.2 Homogeneous Helmholtz Equation

By means of a Fourier transform

$$P(\boldsymbol{x},\omega) = \int_{t=-\infty}^{\infty} p(\boldsymbol{x},t)\mathrm{e}^{\mathrm{i}\omega t}\mathrm{d}t \tag{18.7}$$

sound pressure can be transformed from the time domain to the frequency domain and we speak of a *sound spectrum*. Here, e is Euler's number, $\mathrm{i} = \sqrt{-1}$ is the imaginary unit, $\omega = 2\pi f$ is the angular frequency and f the frequency. By inverse Fourier transform

$$p(\boldsymbol{x},t) = \frac{1}{2\pi}\int_{\omega=-\infty}^{\infty} P(\boldsymbol{x},\omega)\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}\omega\,. \tag{18.8}$$

the spectrum can be transposed back to the time domain. The wave equation in the frequency domain is known as the *homogeneous Helmholtz equation* and reads

$$\nabla^2 P(\boldsymbol{x},\omega) + k^2 P(\boldsymbol{x},\omega) = 0 \tag{18.9}$$

with *wave number* or *spatial frequency*

$$k = \frac{\omega}{c} = \frac{2\pi}{\lambda}\,,$$

where λ is the wave length. Since the Fourier transform is an integral over time, the homogeneous Helmholtz equation is only valid for stationary signals i. e., periodic vibrations, and not for transients [18.35].

18.2.3 Plane Waves

A general solution of the wave equation is d'Alembert's solution

$$p(\boldsymbol{x},t) = f(\boldsymbol{x}-ct) + f(\boldsymbol{x}+ct)\,. \tag{18.10}$$

The first term describes the propagation of a pressure state in the $\boldsymbol{x}$ direction, the second a propagation in the opposite direction. For waves the principle of superposition applies i. e., they interfere without affecting each other. Assuming the second term to be 0, only one wave in the $\boldsymbol{x}$ direction remains. Other directions can simply be added. One possible solution function $f(\boldsymbol{x}-ct)$ is the function of a plane wave in the time domain

$$p(\boldsymbol{x},t) = A(\omega)\,\mathrm{e}^{-\mathrm{i}(k\boldsymbol{x}-\omega t)} \tag{18.11}$$

and in the frequency domain

$$P(\boldsymbol{x},\omega) = A(\omega)\,\mathrm{e}^{\mathrm{i}k\boldsymbol{x}}\,. \tag{18.12}$$

Here, A is an arbitrary complex amplitude in the form $\hat{A}\mathrm{e}^{\mathrm{i}\phi}$ whose absolute value, the amplitude $\hat{A}$ and argument, the phase ϕ, are individual for each frequency.

$$k^2 = k_x^2 + k_y^2 + k_z^2$$

is the squared wave number in direction $\boldsymbol{x}$, and

$$\lambda^2 = \lambda_x^2 + \lambda_y^2 + \lambda_z^2$$

the wave length in $\boldsymbol{x}$-direction. A plane wave propagates in direction $\boldsymbol{x}$ and phase changes with respect to location. k_x, k_y and k_z are called *trace wavenumbers* [18.7]; λ_x, λ_y and λ_z are *trace wavelengths*. They are projections to the spatial axes. The wave fronts are infinite planes of equal pressure perpendicular to vector $\boldsymbol{x}$.

For a wave with nonnegative k, two formulations for k_y

$$k_y = \begin{cases} \pm\sqrt{k^2 - k_x^2 - k_z^2}, & k^2 \geq k_x^2 + k_z^2 \\ \pm\mathrm{i}\sqrt{-k^2 + k_x^2 + k_z^2}, & k^2 \leq k_x^2 + k_z^2 \end{cases} \tag{18.13}$$

point out two different sorts of wave [18.36]. In the first case all components are real, indicating a propagating plane wave. In the second case k_y is imaginary, leading to an *evanescent wave*. Inserting the second case in (18.11) yields

$$P(\boldsymbol{x},\omega) = A(\omega)\,\mathrm{e}^{\pm\sqrt{-k^2+k_x^2+k_z^2}y}\mathrm{e}^{\mathrm{i}(k_x x + \mathrm{k}_z z)}\,. \tag{18.14}$$

In this case the first exponential term is real, indicating an exponential decay in the y-direction. The term could

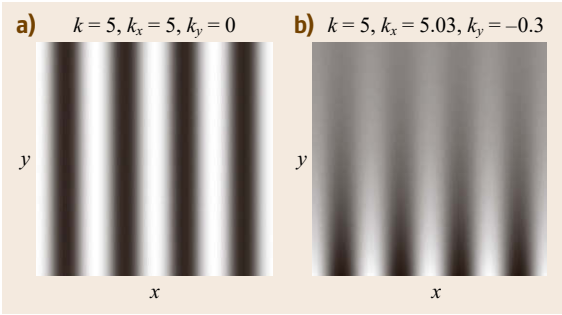


Fig. 18.1a,b Two-dimensional visualization of a propagating plane wave **(a)** and an evanescent wave **(b)**

also indicate an exponential increase, which is ignored since it is nonphysical [18.36]. Both types of waves are illustrated in Fig. 18.1. Note that in this example the propagation direction of the propagating wave and the evanescent wave are the same. Evanescent waves occur at boundaries where the sound velocity in one medium is higher than in the other.

For periodic functions, the motion equation (18.1) yields

$$\nabla p(\boldsymbol{x}, t) = -\mathrm{i}\omega\rho_0 \boldsymbol{v}(\boldsymbol{x}, t) \tag{18.15}$$

and in the frequency domain

$$\nabla P(\boldsymbol{x}, \omega) = -\mathrm{i}k\rho_0 c \boldsymbol{V}(\boldsymbol{x}, \omega)\,, \tag{18.16}$$

where $\boldsymbol{V}$ is the sound velocity in the frequency domain.

18.2.4 Inhomogeneous Wave Equation

The homogeneous wave equation assumes a source-free medium. But every sound field has at least one source that adds acoustic energy to the medium, propagating as a wave pursuant to the wave equation. To account for this, the eighth condition listed in Table 18.1 is dropped and a source term is added to the homogeneous wave equation. Then a solution $p(\boldsymbol{x}, t)$ is sought describing the temporal and spatial behavior of the source signal in the system

$$\nabla^2 p(\boldsymbol{x}, t) - \frac{1}{c^2}\frac{\partial^2 p(\boldsymbol{x}, t)}{\partial t^2} = -4\pi\delta(\boldsymbol{x} - \boldsymbol{x}_0, t - t_0)\,. \tag{18.17}$$

This wave equation is called the *inhomogeneous wave equation*. $\delta(\boldsymbol{x}, t)$ is the Dirac delta function. It is defined as being ∞ at point $\boldsymbol{x}_0$ at time t_0, otherwise it is 0. This means the source term is an impulse that occurs only at one point in time and space. A transformation of the temporal Dirac delta function into the spectral domain

$$\delta(\omega) = \int_{t=-\infty}^{\infty} \delta(t - t_0)\, \mathrm{e}^{\mathrm{i}\omega t}\mathrm{d}t = 1 \tag{18.18}$$

shows that its amplitude for every frequency is 1, i. e., all frequencies have an equal amplitude and are in phase. That means every arbitrary function $p(t)$ can be expressed by weighted and delayed Dirac delta functions $\delta(\boldsymbol{x}, t)$. The amplitude and phase of spectral components $P(\omega)$ of sound signals may be arbitrary so they can be expressed as multiplication of the spectra of the Dirac delta function by frequency-dependent complex amplitudes $A(\omega)$. That conforms to a convolution of a sound signal with the Dirac delta function in the time domain.

18.2.5 Point Sources

One solution for the inhomogeneous wave equation is the point source. A point source is a sound source with no volume. In the simplest case, its radiation is equal in each direction. This is referred to as a *monopole source*. Amplitude and phase are dependent on frequency and distance but independent of direction. Therefore, a formulation in spherical coordinates is meaningful. For spherical coordinates the following holds

$$\begin{aligned}
\boldsymbol{r} &= \begin{bmatrix} r \\ \varphi \\ \vartheta \end{bmatrix} \\
r &= \sqrt{x^2 + y^2 + z^2} \\
\varphi &= \arctan\left(\frac{y}{x}\right) \\
\vartheta &= \arccos\left(\frac{z}{r}\right) \\
\nabla_{\text{spherical}} &\equiv \frac{\partial}{\partial r} + \frac{1}{r}\frac{\partial}{\partial \vartheta} + \frac{1}{r\,\sin\vartheta}\frac{\partial}{\partial \varphi} \\
\boldsymbol{x} &= \begin{bmatrix} x \\ y \\ z \end{bmatrix} \\
x &= r\,\cos\varphi\,\cos\vartheta \\
y &= r\,\sin\varphi\,\cos\vartheta \\
z &= r\,\sin\vartheta \\
\nabla_{\text{Cartesian}} &\equiv \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z}\,.
\end{aligned} \tag{18.19}$$

Here, r is the radius, φ the azimuth angle and ϑ the polar angle. The position vector $\boldsymbol{x}$ is redefined to $\boldsymbol{r}$. The origin of the coordinate system is the source position.

Figure 18.2 illustrates the relations of Cartesian and spherical coordinate systems.

Thus, the inhomogeneous wave equation (18.17) takes the form [18.7]

$$\frac{1}{r}\frac{\partial\left(r^2\frac{\partial p}{\partial r}\right)}{\partial r}+\frac{1}{r^2\sin\vartheta}\frac{\partial\left(\sin\vartheta\frac{\partial p}{\partial\vartheta}\right)}{\partial\vartheta}+\frac{1}{r^2\sin^2\vartheta}\frac{\partial^2 p}{\partial\varphi^2}-\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}=-4\pi\delta(\boldsymbol{x}-\boldsymbol{x}_0,t-t_0)\,. \tag{18.20}$$

Since radiation of a monopole is independent of φ and ϑ the wave equation simplifies to

$$\frac{\partial^2 p(r,t)}{\partial r^2}+\frac{2}{r}\frac{\partial p(r,t)}{\partial r}-\frac{1}{c^2}\frac{\partial^2 p(r,t)}{\partial t^2}=-4\pi\delta(r-r_0,t-t_0)\,, \tag{18.21}$$

and the Helmholtz equation appropriately to

$$\frac{\partial^2 P(r,t)}{\partial r^2}+\frac{2}{r}\frac{\partial P(r,t)}{\partial r}-k^2P(r,t)=-4\pi\delta(r-r_0,\omega)\,. \tag{18.22}$$

The point source solution for this case is

$$p(r,t)=g(r,t)+\tilde{g}(r,t)=A(t)\frac{\mathrm{e}^{-\mathrm{i}(kr-\omega t)}}{r}+\tilde{g}(r,t) \tag{18.23}$$

in the time domain and

$$P(r,\omega)=G(r,\omega)+\tilde{G}(r,\omega)=A(\omega)\frac{\mathrm{e}^{-\mathrm{i}kr}}{r}+\tilde{G}(r,\omega)\,, \tag{18.24}$$

in the frequency domain. It is a *Green's function* comprised of a linear combination of a special solution – $g(r,t)$ and $G(r,\omega)$ – and a general solution – $\tilde{g}(r,t)$ and $\tilde{G}(r,\omega)$ – which are arbitrary solutions of the homogeneous wave equation (18.4) and Helmholtz equation (18.9). It is also called *impulse response* in the time domain and *complex transfer function* in the frequency domain. The impulse response describes the spatiotemporal propagation of the source signal in the medium, the second describes the spatial distribution of the source spectrum. Since the first term of the impulse response is already a complete solution of the inhomogeneous Helmholtz equation, the second term can be assumed to be zero. This case is called *free field Green's function* and describes the radiation of a monopole sound source. The exponential term describes the phase

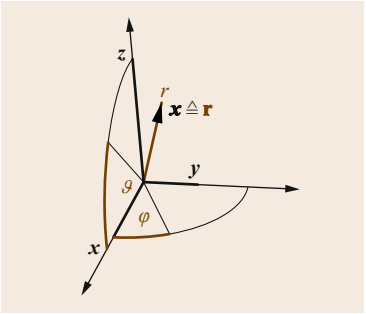


Fig. 18.2 Representation of the position vector $\boldsymbol{x}$, or respectively $\boldsymbol{r}$, via Cartesian coordinates and spherical coordinates

shift per distance of the propagating wave from the source. The fraction represents the pressure amplitude decay per distance, the so-called *inverse distance law* or $1/r$ *distance law*. It states that a local sound pressure amplitude at a receiver point is proportional to the reciprocal of its distance from the source point. This is owed to the fact that the surface of the wave front increases with an increasing sphere radius, so the pressure distributes on a growing area.

The surface of a sphere S is given as

$$S=4\pi r^2 \tag{18.25}$$

so the sound intensity I_0 in the origin of the point source at $r=0$ spreads out to the surface with

$$I(r)=I_0\frac{1}{4\pi r^2}$$

and is thus directly proportional to $1/r^2$. Since I is proportional to p^2, $p(r)$ it is directly proportional to $1/r$ [18.37]

$$I(r)\propto\frac{1}{r^2}\qquad p(r)\propto\frac{1}{r}\,. \tag{18.26}$$

The wave front of a propagating plane wave, in contrast, is assumed to be infinite and thus does not decay. In the far field – i. e., ignoring near-field effects that show a complicated behavior close to the source – any stationary sound source can be simplified by considering it as point source [18.36]. These point sources, however, do not necessarily have to be monopoles. A dependence on direction $\Psi(\omega,\varphi,\vartheta)$ can be introduced ex post by reconsidering $A(\omega)$ as $A(\varphi,\vartheta,\omega)$ or, respectively, $\Psi(\omega,\varphi,\vartheta)A(\omega)$ for the far field

$$p(\varphi,\vartheta,r,t)=g(\varphi,\vartheta,r,t)+\tilde{g}(\varphi,\vartheta,r,t)=\Psi(\omega,\varphi,\vartheta)A(\omega)\frac{\mathrm{e}^{-\mathrm{i}(kr-\omega t)}}{r}+\tilde{g}(r,t) \tag{18.27}$$

which keeps its properties in the frequency domain

$$P(\varphi,\vartheta,r,\omega) = G(\varphi,\vartheta,r,\omega) + \tilde{G}(\varphi,\vartheta,r,\omega)$$
$$= \Psi(\omega,\varphi,\vartheta)A(\omega)\frac{\mathrm{e}^{-\mathrm{i}kr}}{r} + \tilde{G}(r,\omega)\,. \tag{18.28}$$

Due to the complex factor $\Psi(\omega,\varphi,\vartheta)$, the amplitude $A(\omega)$ is modified for any direction. Note, that the Green's function with a direction-dependent radiation factor is not a solution to the inhomogeneous Helmholtz function as such [18.36]. It rather comprises the spherical harmonics, which are a solution to the angular dependencies of the Helmholtz equation in spherical coordinates over a sphere rather than a point. The radiation characteristic of point sources can be any arbitrary function of angles φ and ϑ, which can be composed by a linear combination of mono- and multipoles. In the literature, point sources with a direction-dependent radiation factor are called *complex point sources*, *multipole point sources*, *higher mode radiators* or *point multipoles*; the directivity is called *far-field signature function* [18.36, 38–40].

18.2.6 Huygens' Principle

Every arbitrary radiation from a sound source can be described as integral of elementary sources on its surface. Together, the wave fronts of these elementary sources form the advanced wave front via superposition. This finding is called *Huygens' principle* and is the foundation on which wave field synthesis is based. Figure 18.3 illustrates Huygens' principle. A wave front propagating outwards from a point source is a sphere whose radius increases proportional to the sound velocity c. As stated earlier, such a point source can have individual amplitudes and phases in each direction, indicated here by brightness. At an arbitrary point in time t_1, the wave front can be sampled by a number of points, illustrated as a finite number of dots. These points can be considered as sources of *elementary waves*. The superimposed wave fronts of these elementary sources create the new wave front at a later point in time t_2; an increased sphere. Obviously, the radiation of the elementary waves is not monopole-like. They propagate outwards and set all particles on the increased spherical surface in motion. Inwards, their superposition must yield exactly 0, so no wave front is traveling back to the origin. This radiation characteristic is described by the Kirchhoff–Helmholtz (K–H) integral, discussed in the subsequent subsection.

This principle constitutes the idea of an *acoustic curtain* [18.1, 36]: If an array of microphones records elementary waves, the playback of these recordings via loudspeakers, which are arranged in the exact same way as the microphones, should recreate the original wave field to a certain degree. This idea is a basic concept of wave field synthesis. It is illustrated in Fig. 18.4. The black dots represent pressure microphones that are directly connected to loudspeakers. However, this setup does not provide a thorough sound field capture and synthesis. Errors occur due to the discrete and finite microphone and loudspeaker array as well as due to deficient knowledge of the pressure gradient. How to overcome these issues to increase acoustic control is described in the following section.

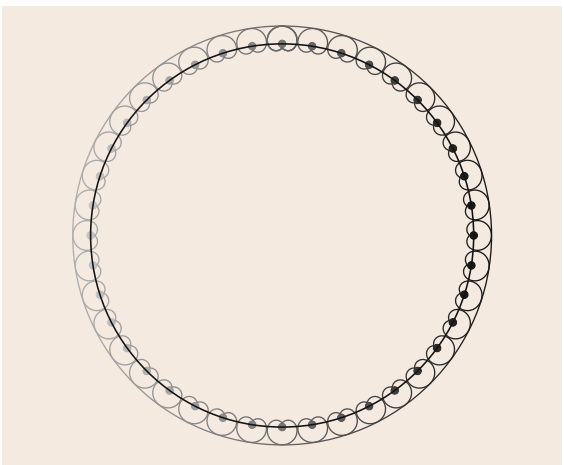


Fig. 18.3 Illustration of Huygens' principle

18.2.7 Kirchhoff–Helmholtz Integral

Gauss' theorem or the *divergence theorem* states that spatial area integrals of a function over a volume V are equal to surface integrals of the normal components $\boldsymbol{n}$ of a function over the volume's surface S

$$\int_V \nabla \boldsymbol{f}\,\mathrm{d}V = \int_S \boldsymbol{f}\boldsymbol{n}\,\mathrm{d}S \tag{18.29}$$

if it has a piecewise smooth boundary and the function $\boldsymbol{f}$ is a steady, differentiable vector function [18.41].

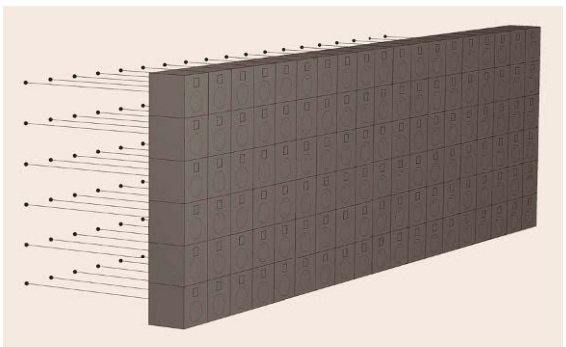


Fig. 18.4 Illustration of an acoustic curtain (after [18.40])

A special case of Gauss' theorem is described by Green's second theorem [18.41]

$$\int_V f\nabla^2 g - g\nabla^2 f \,\mathrm{d}V = \int_S f\nabla g\boldsymbol{n} - g\nabla f\boldsymbol{n} \,\mathrm{d}S\,. \tag{18.30}$$

From Green's second theorem and the Helmholtz equations, (18.9) and (18.22), the Kirchhoff–Helmholtz integral can be derived, which describes the relation between the wave field in a source-free volume V and sources $\boldsymbol{Y}$ on its surface S

$$-\frac{1}{4\pi}\oint_S \left(G(\omega, \Delta\boldsymbol{r}) \frac{\partial P(\omega, \boldsymbol{Y})}{\partial \boldsymbol{n}} - P(\omega, \boldsymbol{Y}) \frac{\partial G(\omega, \Delta\boldsymbol{r})}{\partial \boldsymbol{n}} \right) \mathrm{d}S = \begin{cases} P(\omega, \boldsymbol{X}), & \boldsymbol{r} \in V \\ \frac{1}{2} P(\omega, \boldsymbol{X}), & \boldsymbol{r} \in S \\ 0, & \boldsymbol{r} \notin V \end{cases} \tag{18.31}$$

The K–H integral is a solution to the homogeneous Helmholtz equation with inhomogeneous boundary conditions on the surface. It states that the spectrum $P(\omega, \boldsymbol{X})$ at each point $\boldsymbol{X}$ in a source-free volume V is the integral of the spectra $P(\omega, \boldsymbol{Y})$ at every point $\boldsymbol{Y}$ on the bounding surface S and their propagation function $G(\omega, \Delta\boldsymbol{r})$ in the direction of the normal vector pointing inwards. $G(\omega, \Delta\boldsymbol{r})$ is a Green's function, a solution of the inhomogeneous Helmholtz equation (18.23), and $P(\omega, \boldsymbol{Y})$ is a spectrum, a solution for the homogeneous Helmholtz equation (18.9). $\Delta\boldsymbol{r}$ is the Euclidean distance $||\boldsymbol{Y} - \boldsymbol{X}||_2$. The sources $\boldsymbol{Y}$ on the boundary surface are *secondary sources*, excited by a *primary source* $\boldsymbol{Q}$, which lies in the source volume U. The first term of the closed double contour integral describes a wave that propagates as a monopole since the propagation term

$$G(\omega, \Delta\boldsymbol{r}) = \frac{\mathrm{e}^{-\mathrm{i}k\Delta\boldsymbol{r}}}{\Delta\boldsymbol{r}}$$

is a monopole. From the periodic motion equation (18.15) it emerges that $\partial P/\partial\boldsymbol{n}$ is proportional to sound particle velocity in normal direction $\boldsymbol{V}_\mathrm{n}$. The second term of the integral is a wave that radiates as a dipole, since

$$\frac{\partial G(\omega, \Delta\boldsymbol{r})}{\partial \boldsymbol{n}} = \frac{1 + \mathrm{i}k\Delta\boldsymbol{r}}{\Delta\boldsymbol{r}^2} \cos(\varphi) \mathrm{e}^{-\mathrm{i}k\Delta\boldsymbol{r}} \tag{18.32}$$

is a dipole term. Sound field quantities P and $\boldsymbol{V}$ are convertible into each other after Euler's equation of motion (18.1), so (18.31) is overdetermined and several approaches to a solution exist.

As already stated, the secondary sources on the surface of the source-free medium are monopole and dipole sources. Inwards they radiate in phase and outwards inversely phased. So the radiation doubles inwardly by constructive interference and outwardly becomes 0 by destructive interference. Combined, they create a *cardioid*, also referred to as *kidney* or *heart*. It is illustrated in Fig. 18.5.

The boundary surface could be the wave front around a source and the source-free volume could be the room beyond this wave front. Then, the K–H integral is a quantified formulation of Huygens' principle. But the volume could also be any other arbitrary geometry and the surface a physically existing or nonexisting boundary. This boundary is the separation surface between a source volume, which contains one or more sources, and a source-free volume, which contains the listening area. Any arbitrary closed boundary is conceivable as long as the premises of Gauss' theorem are observed. Figure 18.6 illustrates three examples for a volume boundary that will be considered later in this chapter. Two setup concepts exist: surrounding the listener with secondary sources – as in Fig. 18.6a,b – or surrounding the primary source(s), as illustrated in Fig. 18.6c.

The Kirchhoff–Helmholtz integral describes analytically how spectrum and radiation on a volume surface are related to any arbitrary wave field inside a source-free volume. Therefore, this integral is the core of wave field synthesis [18.3].

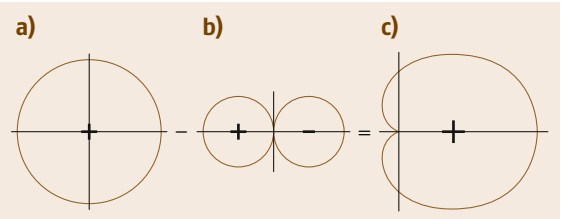


Fig. 18.5a–c Two-dimensional illustration of superposition. **(a)** Monopole- and **(b)** dipole-source form a **(c)** cardioid-shaped radiation (after [18.40])

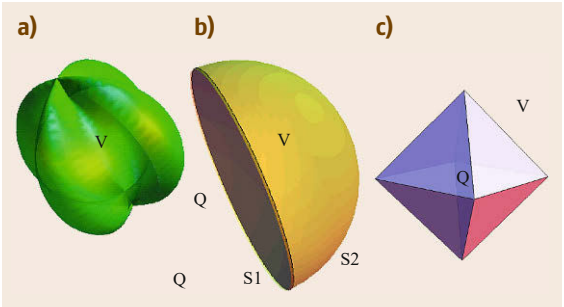


Fig. 18.6a–c Three volumes V with potential source positions Q, **(a)** arbitrary geometry, **(b)** hemisphere, **(c)** octahedron (after [18.40])

18.3 Wave Front Synthesis

The Kirchhoff–Helmholtz integral is a theoretical construct that cannot simply be put into practice by technical means. It demands a continuous distribution of an infinite number of secondary sources with infinitesimal distance, surrounding a volume entirely. That means sound pressure and velocity need to be controllable everywhere on the volume surface, which is hardly possible. However, what we can control is the sound pressure of loudspeakers. But an infinite number of infinitesimally distanced loudspeakers would be required, completely separating a listening area from a source volume and radiating inwards to the listening area but not outwards. This is still not implementable. For a practical realization the reduction of secondary sources to a real number of loudspeakers with discrete distances radiating approximately as monopoles or dipoles is feasible. These have to be fed with the correct *driving functions* [18.5]. Surrounding an entire room with speakers is impracticable as it requires overcoming enormous technical challenges, large computational power and large acquisition and operating costs. Therefore, concepts with plane loudspeaker arrays [18.42] and line arrays [18.25, 43, 44], circular arrays [18.44, 45] and three to four lines surrounding the listening area [18.28, 45] are proposed and in use.

For implementing a wave field synthesis system the K–H integral has to be adjusted to the restrictive circumstances, which leads to errors in the synthesis. Adjustment steps, resulting synthesis errors and their effects as well as compensation methods are discussed in the following.

18.3.1 Adjustments for Implementations

A number of adjustments simplify the K–H integral in a way that allows for a technical implementation of the theory by means of loudspeaker arrays [18.45]. These are listed in Table 18.2. The particular steps will be successively accomplished in the following subsections.

Table 18.2 Adjustments to simplify the Kirchhoff–Helmholtz integral

1.	Enable sources inside the source-free volume
2.	Reduction of the boundary surface to one separation plane between source-free volume and source volume
3.	Restriction to one type of radiator (monopole or dipole)
4.	Reduction of three-dimensional synthesis to two dimensions
5.	Discretization of the surface
6.	Finitization of the surface

18.3.2 Focused Sources

Although the listening volume is required to be source-free, the K–H integral provides us with a tool to create a virtual source located in the inside. This is achieved by creating a concave wave front on the surface, which focuses in one point inside. From this focus point the wave front appears to be convex and cannot be distinguished from a point source located at the focus position. Such sources are called *focused sources*. In the half space behind the focus point the wave front propagation is synthesized correctly. Between the secondary sources and the focus point, however, this is not the case. Here, the wave front does not propagate away from the virtual source position but towards it. Due to this synthesis error, listeners in this region will not localize the sound as coming from the focus point. For focused sources, not the secondary source distribution but the focus point defines the new separation plane between source volume and listening area. So focused sources reduce the listening area to positions behind the focus point. For an implementation a decision has to be made: either the subset of loudspeakers closest to the virtual focused source is used, resulting in the largest possible listening area; or a preferred listening area is defined and those loudspeakers are chosen that create the least erroneous wave fronts inside this area. The derivation of the secondary source signals and further information on these sources can be found, e.g., in [18.29, 46]. Illustrations, Figs. 18.13 and 18.14, are given later in the context of artifacts and compensation methods.

18.3.3 Rayleigh Integrals

Imagine a volume V consisting of a circular plane $S1$ closing a hemisphere $S2$, as illustrated in Fig. 18.6b, whose radius converges to ∞. The influence of the radiation from the secondary sources on $S2$ becomes 0 for the area right in front of $S1$. This coherence satisfies the so-called *Sommerfeld condition*. A separating plane between two half spaces, the source-free volume and source volume, remains. The K–H integral then consists of an integral over the plane $S1$ and thus fulfills the second simplification criterion from Table 18.2.

$$-\frac{1}{4\pi}\iint_{S1}\left(G(\omega,\Delta\boldsymbol{r})\frac{\partial P(\omega,\boldsymbol{Y})}{\partial \boldsymbol{n}} - P(\omega,\boldsymbol{Y})\frac{G(\omega,\Delta\boldsymbol{r})}{\partial \boldsymbol{n}}\right)\mathrm{d}S = \begin{cases} P(\omega,\boldsymbol{x}), & \boldsymbol{x}\in V \\ 0, & \boldsymbol{x}\notin V\,. \end{cases} \tag{18.33}$$

This step reduces the area of secondary sources from a three-dimensional surrounding of a source-free volume to a separation plane.

Since the Green's function that was given in (18.23) is a linear combination of a special solution and a general solution, one term of the integral can be eliminated by adding a carefully chosen general solution to the free-field Green's function. This way the radiation can be restricted to one type of radiator. If the Green's function is chosen to be

$$G_{\mathrm{D}}(\omega, \Delta\boldsymbol{r}) = \frac{\mathrm{e}^{-\mathrm{i}k\Delta r}}{\Delta r} + \frac{\mathrm{e}^{-\mathrm{i}k\Delta r'}}{\Delta r'} \,, \tag{18.34}$$

where $\Delta\boldsymbol{r}'$ is the mirrored position of $\boldsymbol{X}$, mirrored at the tangent of point $\boldsymbol{Y}$ on $S1$. Then $G_{\mathrm{D}}(\omega, \Delta\boldsymbol{r})$ is 0 on the surface $S1$ – which satisfies the *homogeneous Dirichlet boundary condition* [18.47] – and the second term vanishes. This implicitly models the boundary as a rigid surface [18.5], leading to the *Rayleigh I integral* for secondary monopole sources, which is not only applied for wave front synthesis but also for other wave field synthesis approaches as well as near-field acoustic holography and other applications

$$P(\omega, \boldsymbol{X}) = -\frac{1}{2\pi} \iint\limits_{S1} \left(G_{\mathrm{D}}(\omega, \Delta\boldsymbol{r}) \frac{\partial P(\omega, \boldsymbol{Y})}{\partial \boldsymbol{n}} \right) \mathrm{d}S \,. \tag{18.35}$$

Now, considering

$$\frac{\partial P(\omega, \boldsymbol{Y})}{\partial \boldsymbol{n}}$$

the desired driving functions of the secondary sources, an explicit solution can be found, e.g., by means of wave field expansion. This approach is called the *simple source approach* and is the basis of some sound field synthesis methods such as higher-order ambisonics. Here, the sound pressure and pressure gradients are measured (data-based) or calculated (model-based) at a listening position. After decomposing these to a series of *spherical harmonics* – which are orthonormal eigensolutions of the wave equation – an analytical solution for a spherical distribution of secondary sources can be found. The higher the number of measured or calculated pressure gradients, the more the listening point expands to a listening area in which the sound field is synthesized correctly, if enough secondary sources are available. Further information on higher-order ambisonics can be found, e.g., in [18.5–11].

Since the distance $|\Delta\boldsymbol{r}|$ between secondary source position $\boldsymbol{Y}$ and considered position in the source-free volume $\boldsymbol{X}$ equals the distance between the secondary source position and the mirror position $|\Delta\boldsymbol{r}'|$, $G_{\mathrm{D}}(\omega, \Delta\boldsymbol{r})$ is nothing but a doubling of the free-field Green's function $G(\omega, \Delta\boldsymbol{r})$

$$G_{\mathrm{D}}(\omega, \Delta\boldsymbol{r}) = 2G(\omega, \Delta\boldsymbol{r}) \,. \tag{18.36}$$

Assuming

$$\frac{G_{\mathrm{N}}(\omega, \Delta\boldsymbol{r})}{\partial n}$$

to be 0 satisfies the *homogeneous Neumann boundary condition* [18.47] and the first term of (18.33) vanishes. This is accomplished by choosing

$$G_{\mathrm{N}}(\omega, \Delta\boldsymbol{r}) = \frac{\mathrm{e}^{-\mathrm{i}k\Delta r}}{\Delta r} - \frac{\mathrm{e}^{-\mathrm{i}k\Delta r'}}{\Delta r'} \,, \tag{18.37}$$

yielding the *Rayleigh II integral* for secondary dipole sources

$$P(\omega, \boldsymbol{X}) = -\frac{1}{2\pi} \iint\limits_{S1} \left(P(\omega, \boldsymbol{Y}) \frac{\partial G(\omega, \Delta\boldsymbol{r})}{\partial \boldsymbol{n}} \right) \mathrm{d}S \,. \tag{18.38}$$

In both cases the third simplification criterion from Table 18.2 is satisfied. But since the destructive interference outside the source-free volume is missing, $P(\omega, \boldsymbol{X})$ for $\boldsymbol{X} \notin V$ is not 0. A mirrored sound field in the source volume is the consequence. In the case of monopoles, the sound field created by the secondary sources is identical with the one inside the source-free volume. This is illustrated in Fig. 18.7. In the case of dipole sources, the phase in the source volume is the inverse of the phase inside the source-free volume as illustrated in Fig. 18.8. Additionally, the sound pressure, or respectively the particle velocity, duplicate by adding the general solution of the Green's function.

Both formulations do not apply for arbitrary volume surfaces but for separation planes only [18.5]. To ensure that any position around the listening area can be a source position, the listening area has to be surrounded by several separation planes. If (18.35) and

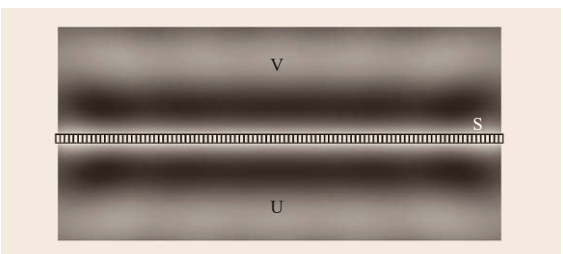


Fig. 18.7 Mirrored sound field due to the exclusive use of secondary monopole sources

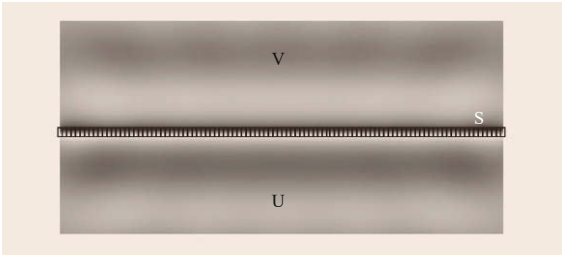


Fig. 18.8 Inversely phased mirrored sound field due to the exclusive use of secondary dipole sources

(18.38) are applied to other geometries, they still deliver approximate results [18.5]. In any case, the source-free volume has to be convex so that no mirrored sound field lies inside the source-free volume, i. e., volume in Fig. 18.6a is inappropriate [18.8]. Since $S1$ is modeled as a rigid surface, reflections occur when a listening area is surrounded by several separation planes. This problem is illustrated in Fig. 18.9. The squares represent loudspeakers that surround the source-free area. Loudspeaker signals from three directions propagate towards the inside. Loudspeakers contribute to the synthesis of the wave front in the half space behind the connection line between them and the virtual source. This half space needs to cover a large portion of the listening area. Otherwise, its mirrored version arrives at the listening area, creating an undesired echo wavefront.

These artifacts can be reduced by a *spatial windowing* [18.5, 43] technique applied to the Rayleigh I integral

$$P(\omega, \boldsymbol{X}) = d(\boldsymbol{Y}) \frac{P(\omega, \boldsymbol{Y})}{\partial \boldsymbol{n}} 2G(\omega, \boldsymbol{Y}) \,,$$
$$d(\boldsymbol{Y}) = \begin{cases} 1, & \text{if } \langle \boldsymbol{Y} - \boldsymbol{Q}, n(\boldsymbol{Y}) \rangle > 0 \\ 0, & \text{otherwise} \end{cases} \qquad (18.39)$$

The variable $d(\boldsymbol{Y})$ is the windowing function for spherical waves, which is 1 if the local propagation direction of the sound of the virtual source at the position of the secondary source has a positive component in normal direction of the secondary source. If the deviation is $\pi/2$ or more, $d(\boldsymbol{Y})$ becomes 0 and the speaker is muted. That means only those loudspeakers whose normal component resembles the tangent of the wave front of the virtual source are active. $G(\omega, \Delta \boldsymbol{r})$ describes the directivity function of the secondary source, i. e., of each loudspeaker. The other terms are the sought-after driving functions D of the loudspeakers [18.5]

$$D(\omega, \boldsymbol{Y}) = 2d(\boldsymbol{Y}) \frac{P(\omega, \boldsymbol{Y})}{\partial \boldsymbol{n}} \,. \qquad (18.40)$$

Figure 18.10 illustrates the same virtual source as Fig. 18.9 but with the applied windowing function, (18.39). The dashed squares represent muted loudspeakers. Here the desired wave front is clearly visible. It is the wave front of a virtual point source located below the loudspeaker array, propagating upwards. Still, strong artifacts occur. The reasons and solutions for these artifacts are discussed next.

Two Dimensions

For applications in which the audience is organized more or less in plane, it is sufficient to recreate the wave field correctly for that listening plane only, rather than in the whole listening volume. Typically, listeners are organized roughly in plane, e.g., in conference rooms, concert halls, cinemas, theaters, in the car, on the couch etc. Furthermore, one or several one-dimensional distributions of loudspeakers are easier implementable than covering a complete room surface with loudspeakers. Reducing the three-dimensional wave field synthesis to two dimensions reduces the separation plane $S1$ to a separation line $L1$. In theory, one could simply reduce the surface integral to a simple integral and the Rayleigh integrals would take the forms

$$P(\omega, \boldsymbol{X}) = \frac{1}{2\pi} \int_{L1} \left(G(\omega, \Delta \boldsymbol{r}) \frac{\partial P(\omega, \boldsymbol{Y})}{\partial \boldsymbol{n}} \right) \mathrm{d}S1 \,, \qquad (18.41)$$

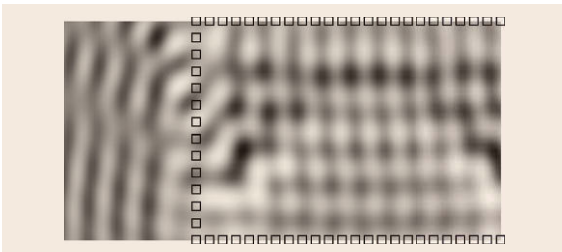


Fig. 18.9 Calculating a surrounding loudspeaker array with Rayleigh I integral implicitly models $S1$ as rigid surfaces creating reflections

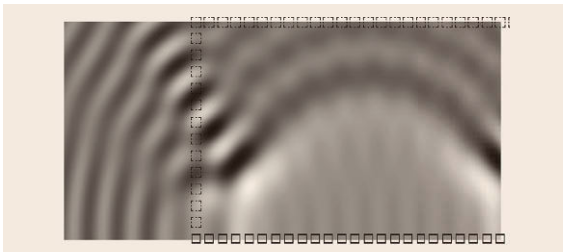


Fig. 18.10 Muting loudspeakers whose normal component deviates strongly from the local propagation direction of the virtual wave front suppresses virtual reflections

and

$$P(\omega, \boldsymbol{X}) = \frac{1}{2\pi} \int_{L1} \left(P(\omega, \boldsymbol{Y}) \frac{\partial G(\omega, \Delta \boldsymbol{r})}{\partial \boldsymbol{n}} \right) \mathrm{d}S1 \,. \tag{18.42}$$

In these cases $\boldsymbol{X}$ is two-dimensional

$$\boldsymbol{X} = \begin{bmatrix} x \\ y \end{bmatrix} . \tag{18.43}$$

This solution was satisfying if no third dimension existed, e.g., if wave fronts of the secondary sources had no spherical but a circular or cylindrical propagation [18.5, 45]. Then, the propagation function $G(\omega, \Delta \boldsymbol{r})$ was different, having an amplitude decay of $1/\sqrt{r}$ instead of $1/r$. This is owed to the fact that the surface S of a circle or cylinder doubles with a doubled circle radius r_{circle}

$$S = 2\pi r_{\mathrm{circle}} \tag{18.44}$$

in contrast to the spherical case in which it squares with the doubled radius as already indicated in (18.26). In this case

$$I \propto \frac{1}{r} \tag{18.45}$$

and thus

$$p \propto \frac{1}{\sqrt{r}} \,. \tag{18.46}$$

So the practical benefit of (18.41) and (18.42) is minor since transducers with a cylindrical radiation in the far field are hardly available. An approximately cylindrical radiation could be achieved with line arrays of loudspeakers, as often applied in PA systems for concerts. But replacing each individual loudspeaker by a line array of speakers contradicts the goal of reducing the number of loudspeakers. Simply replacing cylindrically radiating speakers by conventional loudspeakers that have a spherical radiation function leads to errors in this wave field synthesis formulation due to the deviant amplitude decay.

Huygens' principle states that a wave front can be considered as consisting of infinitesimally distanced elementary sources. An infinite planar arrangement of elementary point sources with a spherical radiation could (re-) construct a plane wave, since the amplitude decay, which is owed to the $1/r$-distance law, is compensated by the contribution of the other sources. Imagining secondary line sources with a cylindrical radiation, a linear arrangement of sources would be sufficient to create a planar wave front. In a linear arrangement of elementary point sources, the contribution of the sources from the second dimension is missing, resulting in an amplitude decay. Therefore, a *2.5-dimensional operator* including a *far-field approximation* that modifies the free-field Green's function to approximate a cylindrical propagation is used [18.5, 48]. This changes the driving function to

$$D_{2.5\mathrm{D}}(\omega, \boldsymbol{Y}) = \sqrt{\frac{2\pi \, |\boldsymbol{Y} - \boldsymbol{X}_{\mathrm{ref}}|}{\mathrm{i}k}} D(\omega, \boldsymbol{Y}) \tag{18.47}$$

with $\boldsymbol{X}_{\mathrm{ref}}$ being a reference point in the source-free volume. This yields the *2.5-dimensional* Rayleigh integral [18.5, 25, 49]

$$P(\omega, \boldsymbol{X}) = -\oint_S D_{2.5\mathrm{D}}(\omega, \boldsymbol{Y}) G(\omega, \Delta \boldsymbol{r}) \,. \tag{18.48}$$

Taking reference points $\boldsymbol{X}_{\mathrm{ref}}$ parallel to the loudspeaker array, the wave field can be synthesized correctly along a reference line. Between the speakers and the reference line, the sound pressures are too high, behind the reference line they are too low.

Until now, free-field conditions are assumed. However, if not installed in the free field, reflections may occur and superimpose with the intended wave field created by the loudspeaker system. Under the term *listening room compensation* a variety of methods are proposed to reduce the influence of reflections. The simplest form is passive listening room compensation, which means that the room is heavily damped. This is an approved method, applied, e.g., in cinemas. However, for some listening rooms, for example living rooms, damping is impractical. Therefore, active solutions are proposed, like adding a filtering function that eliminates the first reflections of the room to the calculated loudspeaker signals [18.50, 51]. *Adaptive wave field synthesis* uses error sensors that measure errors occurring during WFS of a test stimulus emerging, e.g., from reflections or the loudspeaker radiation characteristics [18.43]. Then any WFS solution is modified by a regularization factor that minimizes the squared error. This is of course a vicious circle since compensation signals corrupt the synthesized wave field and are reflected too, adding further errors. Due to an exponentially increasing reflection density it is hardly possible to account for all higher-order reflections. Thus, the approach is limited to first-order reflections.

Discretization

A discretization of the Rayleigh integrals adopts the continuous formulation to discrete secondary source

positions

$$P(\omega, \boldsymbol{X}) = \frac{1}{2\pi} \sum_{\boldsymbol{r}_\mathrm{Y}=-\infty}^{\infty} \left(G(\omega, \Delta\boldsymbol{r}) \frac{\partial P(\omega, \boldsymbol{Y})}{\partial \boldsymbol{n}} \right) \Delta \boldsymbol{r}_\mathrm{Y} \tag{18.49}$$

and

$$P(\omega, \boldsymbol{X}) = \frac{1}{2\pi} \sum_{\boldsymbol{r}_\mathrm{Y}=-\infty}^{\infty} \left(P(\omega, \boldsymbol{Y}) \frac{\partial G(\omega, \Delta\boldsymbol{r})}{\partial \boldsymbol{n}} \right) \Delta \boldsymbol{r}_\mathrm{Y} \,. \tag{18.50}$$

Thereby the Nyquist–Shannon sampling theorem has to be regarded: the sampling frequency has to be at least twice the highest frequency of the signal to be presented for no aliasing to occur. The highest frequency to be represented error-free is the *critical frequency* or *aliasing frequency*. In this case the sampling frequency is spatial; the speaker distance $\Delta \boldsymbol{Y}$ needs to be smaller than half the distance of the largest trace wavelength between the speakers

$$f_\mathrm{max} = \frac{c}{2\Delta \boldsymbol{Y}} \,. \tag{18.51}$$

Typically, secondary sources are equally spaced. The spatial sampling of the secondary source distribution is a process of sampling and interpolation; the interpolator is given by the radiation characteristics of the loudspeakers [18.52]. An adaption of WFS to the radiation characteristic of the loudspeakers is derived in [18.53]. For the trace wavelength between the speakers

$$\lambda_{\Delta \boldsymbol{Y}} = \lambda \, |\sin\alpha| \tag{18.52}$$

is valid, where α is the angle between the normal direction of a loudspeaker and the propagation direction of the virtual source at this loudspeaker position. Respectively, it can be considered as angle between separation line $L1$ and the tangent of the wave front when striking the speaker position. This leads to an adjustment of (18.51) to

$$f_\mathrm{max} = \frac{c}{2\Delta \boldsymbol{Y} \sin\alpha} \,. \tag{18.53}$$

The angle α may vary dependent on position and radiation of the source in a range between $\pi/2$ and $3\pi/2$. Two examples for α are illustrated in Fig. 18.11 to clarify the coherency. The black disk represents the source, the dark and light gray disks the wave front at two different points in time.

Undersampling creates regular erroneous wave fronts above f_max. These erroneous wave fronts, *artifacts*, contain the frequencies above the critical frequency and cause perceivable changes in sound color and disturb the location of the auditory event [18.5, 9]. They can be heard as echoes following and partly superimposing with the synthesized wave fronts. Aliasing artifacts are illustrated in Fig. 18.12. Four periods of a sinusoidal sound are synthesized as a plane wave. Due to undersampling, wave fronts with an unwanted propagation direction superimpose with the synthesized wave front. As a result the synthesized wave fronts are partly overlain and partly followed by unwanted wave fronts. A comb filter effect, well known from stereophonic loudspeaker setups, occurs. Furthermore, the amplitude of the wave front decays with increasing distance to the loudspeaker array, which is not the case for actual plane waves, (18.11) and (18.12). As mentioned earlier, this error occurs due to the missing third dimension.

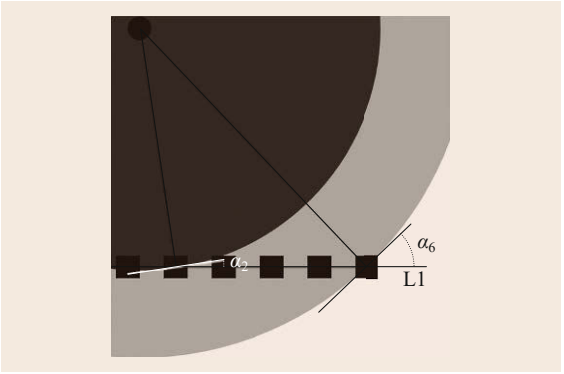


Fig. 18.11 Several incidence angles for one source position (after [18.40])

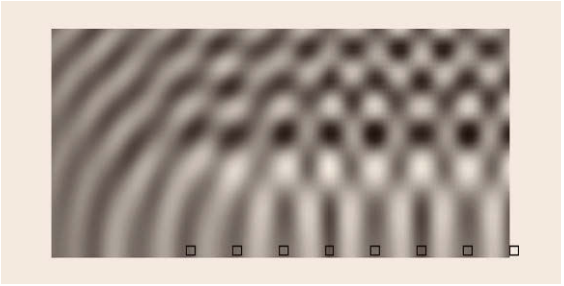


Fig. 18.12 Spatial aliasing due to undersampling with discrete loudspeaker positions (after [18.40])

As long as the condition

$$|\sin\alpha(\omega)| < \frac{c}{2\Delta \boldsymbol{Y} f_\mathrm{max}} = \frac{\pi c}{\Delta \boldsymbol{Y} \omega_\mathrm{max}} \tag{18.54}$$

is satisfied, no aliasing wave fronts will occur.

A focused source with and without aliasing artifacts is illustrated in Figs. 18.13 and 18.14. The virtual source is located the center of the white region.

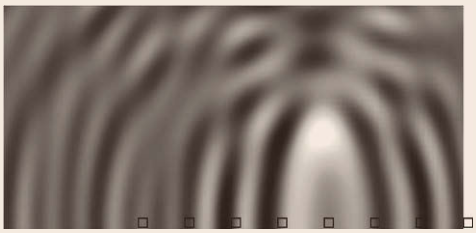


Fig. 18.13 Focused source corrupted by spatial aliasing and truncation errors (after [18.40])

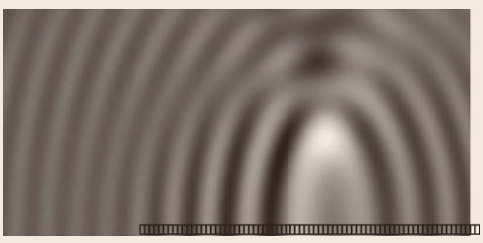


Fig. 18.14 Focused source without spatial aliasing corrupted by truncation errors (after [18.40])

Both have truncation errors due to the finite extent of the loudspeaker array, which will be addressed subsequently. For focused sources, aliasing artifacts occur as pre-echoes preceding the synthesized wave front. Note that the illustrated snapshot shows a point in time at which the loudspeakers are already silent for a while, so there are no wave fronts traveling towards the focus point anymore.

One suggestion to reduce artifacts is to process frequencies above the critical frequency not by means of wave front synthesis but by conventional stereophonic sound between two to three loudspeakers. This method is called *optimized phantom source imaging* (OPSI) [18.5, 48] and combines WFS with conventional intensity panning. Thus, OPSI reintroduces psychoacoustic considerations to WFS. In that manner no aliasing echoes as such occur but the common disadvantages of stereophonic sound become effective: a comb filter effect arises, the display of depth becomes worse and high frequencies are only located correctly in the sweet spot. At other positions two to three wave fronts arrive slightly shifted in time. *J.J. Lopez* et al. [18.54] suggest a related approach, called the *subband approach*, which involves playing frequencies above the aliasing frequency through the one loudspeaker with the most similar direction to the virtual source only. This approach does not bring along the disadvantages of stereo but still a more or less correct localization of virtual sources is only possible in a small part of the listening area. By randomizing the phase of the high frequencies of the loudspeakers artifacts can be smeared [18.5]. The result in illustrated in Fig. 18.15. This minimizes the sound coloration but localization accuracy is reduced [18.48], since amplitude and phase have to be correct for a proper localization via WFS. Also, the resulting wave field does not correspond to the desired one.

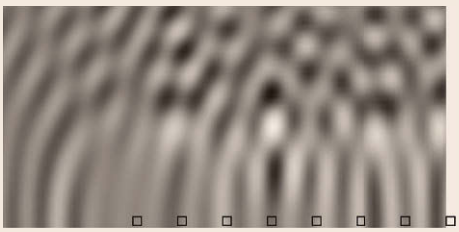


Fig. 18.15 By randomizing the phase of signal components above the aliasing frequency synthesis errors become irregular

In all three cases the signal is divided by the critical frequency into two frequency regions. For the lower frequency region the theory of WFS is applied. For the high frequencies the goal is to retain the natural temporal and spectral properties of the sound wave as well as an approximately correct source position rather than physically correct replicas. The methods are based on the same psychoacoustic considerations:

- Partials of a sound tend to fuse
- Higher frequencies tend to be masked by lower frequencies
- Altogether the audible portion of sound will be integrated into one auditory stream with one group source position.

Then, the lower frequency region – which offers very precise localization cues due to the correct reconstruction of the wave front – is crucial for a distinct and correct localization, and the ambiguous localization cues of higher frequencies are neglected by the auditory system.

Another approach is to increase the aliasing frequency at the expense of the size of the listening area. *Local wave field synthesis* creates several proximate focused sources that act as secondary sources synthesizing high frequency wave fronts correctly for a subspace of the listening area [18.55]. The result is illustrated in Fig. 18.16. The virtual focused sources are marked by an $\times$. Compared to the aliasing case, Fig. 18.12, errors are reduced. This is, however, only true for the area above the line of focused sources. An unintended wave front occurs on the left of the virtual source array. Just as with actual loudspeakers, the finite array of focused point sources introduces truncation errors,

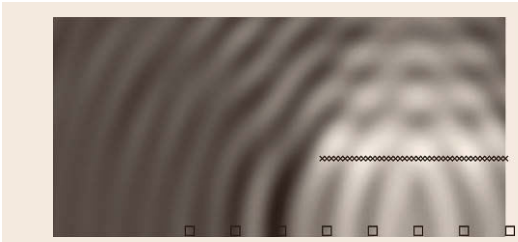


Fig. 18.16 Local wave field synthesis creates plane waves above the aliasing frequency in a subspace of the listening area. Here, the aliasing frequency is increased but truncation errors persist

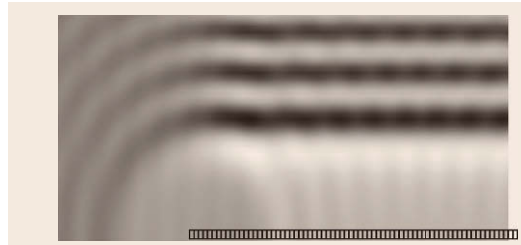


Fig. 18.17 No aliasing but a truncation error due to the finite loudspeaker array length (after [18.40])

which can again be compensated by methods discussed in the following subsection. The computational costs for rendering a high number of virtual sources for each wave front are rather high. Quite a different method is to recreate the wave field not for the discrete loudspeaker positions but for discrete listening positions sampling the listening area. The approach is called *sound field reconstruction* [18.17]. Sampling positions are chosen under the assumption that if a wave field is reproduced correctly on a grid satisfying the Nyquist–Shannon sampling theorem, the wave field is correct everywhere inside the grid. For stationary signals this approach is straightforward. For transient signals psychoacoustic considerations are necessary to achieve a natural sound and a proper source localization [18.19]. The approach can be combined with crosstalk cancellation to create a realistic binaural signal at discrete listening positions [18.18].

Of course, these methods work best if the chosen distance between adjacent speakers is so small that the aliasing frequency is as high as possible. At loudspeaker spacing of 0.1–0.3 m it can be speculated that the influence of the frequencies above the critical frequency is weak concerning sound coloration and localization as perceived by human listeners [18.5].

18.3.4 Finite Extent

Limiting the sum in the discrete Rayleigh integrals (18.49) and (18.50) to a finite number of loudspeaker positions corresponds to the sixth simplification of Table 18.2. This truncates the loudspeaker array. Figure 18.17 shows this scenario. On the left-hand side of the loudspeaker array the synthesized wave front fades to the curved wave front of the outermost speaker. This effect is called *truncation* [18.49]. It appears like diffraction through a gap and has the effect that the wave field cannot be synthesized in the area beyond this border. Furthermore, a spherical wave propagates from the end loudspeaker since the compensatory effect of adjacent speakers is missing. Even behind the synthesized plane wave front this spherical wave front is propagating towards the right. The truncation effect can be compensated by gradually reducing the amplitudes of the outermost loudspeakers, e.g., in terms of a half-Hann function [18.49]

$$\hat{A}(\boldsymbol{Y}_n) = 0.5\left(1-\cos\left(\frac{\pi n}{N}\right)\right),\ n = 0,\ldots,N \tag{18.55}$$

for the nth speaker. The procedure is called *tapering* and the result is illustrated in Fig. 18.18. The amplitude of the loudspeaker signals is indicated by the size of the square. Here, the only errors left are the truncation error arising from the untapered right end of the loudspeaker array. Compensation sources as used in listening room compensation minimize truncation by using speakers with inversely phased signals. However, when surrounding the listening area with three to four speaker line arrays the border effect is inferior [18.49] as the truncation error in corners is weak. This is illustrated in Fig. 18.19. This figure also points out the necessity of heavy damping or listening room compensation: due to constructive interference the signal amplitude outside the source-free volume is higher than the synthesized wave front inside. A reflecting wall would heavily corrupt the sound field inside the source-free volume. This is demonstrated in Fig. 18.20. It shows the same loudspeaker signals as in Fig. 18.19.

Fig. 18.18 A tapering window eliminates the truncation error but reduces the extent of the listening region (after [18.40])

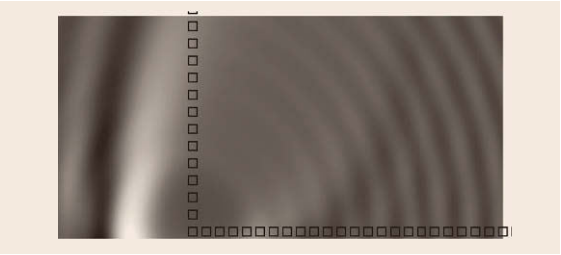


Fig. 18.19 When two loudspeaker line arrays meet, the truncation error is weak

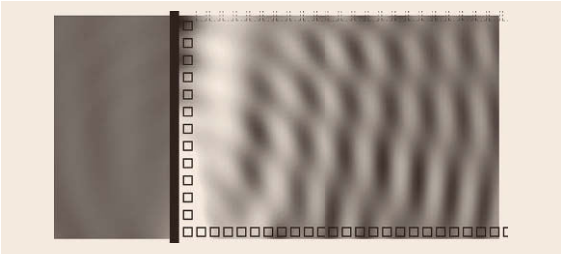


Fig. 18.20 Reflections of the listening room heavily corrupt the synthesized wave field

The black line represents a wall with minor absorption and transmission and without scattering. The reflection corrupts the wave front curvature and propagation direction, and creates a spatial comb filter effect and high amplitudes in the listening area, right in front of the speakers.

18.4 Current Research and Development

Wave field synthesis has reached a state in which the direct sound of virtual monopole sources and plane waves can be synthesized with a high precision in a listening area, which is only limited by the extent of the loudspeaker array and sometimes restricted to a subspace when creating focused sources. Some wave field synthesis systems are already available on the market.

Current research and development projects mainly address two issues. First, the expensive installation and operation of wave field synthesis. Here, one aim is to develop easy accessible software and formats to simulate, create, control, store and play WFS content. The other aim is to make installations more accessible and compatible to interior design by applying flat panel loudspeakers, reducing the number of necessary secondary sources as well as the computational demands. Second, expansions of functionality and the implementation of more acoustical parameters, like sound radiation characteristics and room modes, is still of great interest. Here, the rise of psychoacoustic considerations help to increase control over the perceived sound experience. These two topics are discussed individually in the last subsections.

18.4.1 Applicability

To make wave field synthesis systems more applicable, user interfaces [18.25] and plugins for digital audio workstations [18.31, 56, 57] have been programmed. Here, source signals can be associated with positions and paths in relation to the loudspeaker setup. Also, implementations for mobile devices are being developed [18.26]. Matlab toolboxes help for numerical simulations and visualizations for an easier access to sound field synthesis research and development [18.58, 59]. Because loudspeaker configurations are not standardized, object-based source material is necessary, combining audio tracks with dynamic metadata describing source position and trajectories, source orientation and sound radiation characteristics and information on early reflections and reverberation. Although several formats have been proposed, no standardized wave field synthesis format has established yet [18.25, 36]. The need for object-based source material implies incompatibility with current channel-based audio formats for stereo and 5.1 setups. *M. Cobos* and *J.J. López* evaluated blind audio source separation techniques to use stereo source material for wave front synthesis applications [18.60]. The ISO/MPEG-H 3-D audio standard, which is under development, is supposed to deliver spatial audio object coding that is compatible with stereo, 5.1, 22.2 and higher-order ambisonics [18.61].

Of course, covering a complete living room, concert hall, theater or cinema with a surrounding loudspeaker array is a drastic impairment of the interior decoration. To solve this problem, research currently goes in two directions. The first concept is to apply multiactuator panels (MAPs) [18.24, 27, 62]. Due to their flat, inconspicuous nature, they do not harm interior decoration as much as conventional loudspeakers. MAPs have the additional advantage that they can be arranged continuously, preventing aliasing when sufficient control is reached. However, elaborate filtering is necessary to compensate effects of modes and reverberation within the panels. The second concept is to decrease the number of loudspeakers [18.40, 63, 64]. With psychoacoustic considerations concerning source localization mechanisms and the perception of source width and

spaciousness, the necessary number of loudspeakers can be reduced drastically while maintaining spatial listening experience. Here, all loudspeakers contribute to synthesize the sound field within a defined listening area. Instead of the emanating wave front, a succession of quasistationary sound fields is synthesized in the listening area by solving the discretized Rayleigh I integral, (18.49). This sound field resembles an original sound field as created by a musical instrument including its angle- and frequency-dependent radiation characteristics. In this way, the natural interaural level and phase differences are recreated, giving clues about source direction and extent. However, the spectral cues are not sufficient for a proper localization. As all loudspeaker wavefronts arrive from different angles and at different points in time, source localization is affected by the *precedence effect* [18.65], i. e., the nearest loudspeaker is heard as source position. The result of the precedence effect is unwanted but its initiator can be used. When only one loudspeaker actively plays all note onsets, the perceived source position will coincide with the position of this *precedence loudspeaker*. The signal amplitudes of the other loudspeakers are faded in within several milliseconds. As long as the fading duration is longer than the arrival time difference of the speaker wavefronts and shorter than the *overshoot* phenomenon [18.66] at note onsets, the fading is masked and stays inaudible. This psychoacoustic approach is also applicable with more common loudspeaker setups, such as 5.1 or 22.2, which largely increases compatibility and applicability. The reduced number of loudspeakers also comes along with a reduction of computational demands, which is another topic of current research. *Beckinger* and *Brix* calculated that synthesizing 21 000 rain particles falling in 20 sec with a 64-channel wave front synthesis system would require 26 900% of a typical single-core processor in modern personal computers and suggested an approach that reduces the computational demands to 36% [18.67]. Additionally, ways are investigated to include height with a small number of additional channels [18.68]. This approach is very promising because localization accuracy of human listeners in the median plane is rather poor.

18.4.2 Feature Expansion

So far research has mostly concentrated on synthesizing wave fronts of virtual monopole sources and plane waves. The sound field is synthesized in the spatiotemporal domain. The synthesized wave fronts propagate in a natural manner and deflect around the listeners' heads. Thus, they provide natural localization cues, like interaural time and level differences, wave front curvature and the individual head-related transfer function. Monopoles create a discrete source position localizable by every listener, i. e., they include the source distance. Focused sources are localized within the listening room but not for listeners who are located between the active speakers and the focus point. Plane waves create a common source angle that all listeners agree on. Additional control over virtual sources is gained by motion capture technology [18.26]. By tracking an individual, a singing bird could be placed on both of his shoulders and even stay there while he walks around. However, the left bird is only localized correctly by him and listeners on his right and vice versa. This approach enables 3-D sound for one listener but reintroduces the sweet-spot limitation. For rapidly moving sources, implementing the Doppler effect is important for an authentic sound experience [18.69, 70]. Additionally, several methods have been proposed and implemented to synthesize higher-order radiation patterns, especially the radiation characteristics of musical instruments [18.19, 24, 25, 64, 71]. Effects on perceived timbre, localization, source extent and orientation of the instrument have been reported. As an example, the radiation pattern of a single violin frequency is illustrated in Fig. 18.21. Amplitudes and phases are dependent on direction, therefore wave fronts are no isobars but show certain inhomogeneities. There are also several attempts to include room modes, early reflections and late reverberations, both data-based and model-based [18.23, 72, 73]. It has been found that listeners are very sensitive to diffuseness and overall level of reverberation. An example of virtual room modes is illustrated in Fig. 18.22. The two loudspeaker line arrays synthesize perpendicular plane waves, creating a typical interference pattern. In contrast to a real room, the amplitudes of the the virtual standing waves decrease with increasing distance to the line arrays. This can only partly be compensated by two additional lines.

Many researchers emphasize the potential of psychoacoustics in auralization and wave field synthesis applications [18.22, 64, 74, 75]. The precision of the physical replication of sound fields and thus the amount of data to be processed can be reduced without audible effects. In a psychoacoustic sound field synthesis approach loudspeakers create the desired sound field in a listening area by calculations in the frequency domain [18.19, 64]. Although only valid for stationary signals, calculations are applied for transient sounds. By considering the integration time, precedence effect, critical bands and auditory scene analysis principles physical synthesis errors become inaudible. Another approach based on psychoacoustic considerations is directional audio coding (DirAC) [18.76]. Signals are separated into directional and diffuse components and

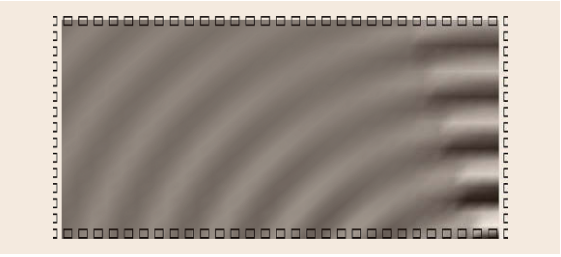


Fig. 18.21 Synthesis of a source with nonuniform radiation pattern

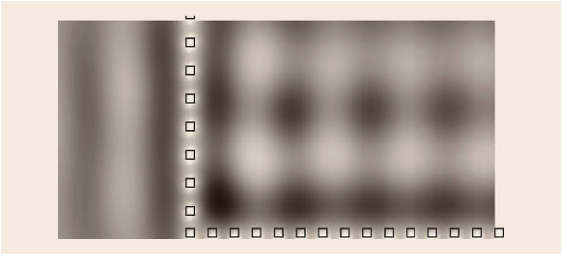


Fig. 18.22 Room modes modeled by orthogonal plane waves

can be treated differently according to those aspects of sound the auditory system is sensitive for. Direct sound and very early reflections affect the perceived source location, orientation and extent, whereas reverberation is characterized by the duration, timbre and degree of diffusion and has an effect on perceived reverberance, spatial impression, listener envelopment and liveness and may modify the perceived loudness, warmth and brilliance of the sound [18.77–79]. Combining limited acoustic control with psychoacoustic control might have the potential to overcome all current issues and restrictions.

References

18.1 J.C. Steinberg, W.B. Snow: Symposium on wire transmission of symphonic music and its reproduction in auditory perspective. Physical factors, Bell Syst. Tech. J. **13**, 239–241 (1934)

18.2 A.J. Berkhout: A holographic approach to acoustic control, J. Audio Eng. Soc. **36**(12), 977–995 (1988)

18.3 A.J. Berkhout, D. de Vries, P. Vogel: Acoustic control by wave field synthesis, J. Acoust. Soc. Am. **93**(5), 2764–2778 (1993)

18.4 A.J. Berkhout, D. de Vries, P. Vogel: Wave front synthesis: A new direction in electroacoustics. In: *Proc. 93rd Audio Eng. Soc. Conv.* (1992) p. 10

18.5 S. Spors, R. Rabenstein, J. Ahrens: The theory of wave field synthesis revisited. In: *Proc. 124th Audio Eng. Soc. Conv.* (2008) p. 5

18.6 J. Ahrens, S. Spors: Analytical driving functions for higher order ambisonics. In: *Proc. IEEE Int. Conf. Acoust. Speech Signal Process., ICASSP* (2008)

18.7 E.G. Williams: *Fourier Acoustics, Sound Radiation and Nearfield Acoustical Holography* (Academic, Cambridge 1999)

18.8 S. Spors, J. Ahrens: A comparison of wave field synthesis and higher-order ambisonics with respect to physical properties and spatial sampling. In: *Proc. 125th Audio Eng. Soc. Conv.* (2008) p. 10

18.9 J. Daniel, R. Nicol, S. Moreau: Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging. In: *Proc. 114th Audio Eng. Soc. Conv.* (2003) p. 3

18.10 D. Menzies, M. Al-Akaidi: Nearfield binaural synthesis and ambisonics, J. Acoust. Soc. Am. **121**(3), 1559–1563 (2007)

18.11 J. Daniel: Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format. In: *Proc. Audio Eng. Soc. Conf. 23rd Int. Conf. Signal Process. Audio Rec. Reprod., May 2003* (2003)

18.12 O. Warusfel, P. Derogis, R. Caussé: Radiation synthesis with digitally controlled loudspeakers. In: *Proc. 103rd Audio Eng. Soc. Conv.* (1997) p. 9

18.13 R. Avizienis, A. Freed, P. Kassakian, D. Wessel: A compact 120 independent element spherical loudspeaker array with programable radiation patterns. In: *Proc. 120th Audio Eng. Soc. Conv.* (2006) p. 5

18.14 F. Zotter: *Analysis and Synthesis of Sound-Radiation with Spherical Arrays*, Ph.D. Thesis (Univ. Music and Performing Arts, Graz 2009)

18.15 T. Ziemer: Wave field synthesis by an octupole speaker system. In: *Proc. SysMus09, Nov. 2009*, ed. by L. Naveda (2009) pp. 89–93

18.16 O. Kirkeby, P.A. Nelson: Reproduction of plane wave sound fields, J. Acoust. Soc. Am. **94**, 2992–3000 (1993)

18.17 M. Kolundzija, C. Faller, M. Vetterli: Sound field reconstruction: An improved approach for wave field synthesis. In: *Proc. 126th Audio Eng. Soc. Conv.* (2009) p. 5

18.18 D. Menzel, H. Wittek, H. Fastl, G. Theile: Binaurale Raumsynthese mittels Wellenfeldsynthese – Realisierung und Evaluierung. In: *Fortschr. Akust. – DAGA, Braunschweig* (2006) pp. 255–256

18.19 T. Ziemer, R. Bader: Implementing the radiation characteristics of musical instruments in a psychoacoustic sound field synthesis system. In: *Proc. 139th Audio Eng. Soc. Conv.* (2015)

18.20 W.-H. Cho, J.-G. Ih, M.M. Boone: Holographic design of a source array achieving a desired sound field, J. Audio Eng. Soc. **58**(4), 282–298 (2010)

18.21 S. Brix, T. Sporer, J. Plogsties: Carrouso – A European approach to 3-D audio (abstract). In: *Proc. 110th Audio Eng. Soc. Conv.* (2001) p. 528

18.22 G. Theile, H. Wittek: Wave field synthesis – A promising spatial audio rendering concept, J. Inst. Image Inf. Telev. Eng. **61**, 638–644 (2007)
18.23 E. Hulsebos: *Auralization Using Wave Field Synthesis*, Ph.D. Thesis (Technical Univ. Delft, Delft 2004)
18.24 E. Corteel: Synthesis of directional sources using wave field synthesis, possibilities and limitations, EURASIP J. Adv. Signal Process. **2007**(1), 090509 (2007)
18.25 M. Baalman: *On Wave Field Synthesis and Electro-Acoustic Music, with a Particular Focus on the Reproduction of Arbitrarily Shaped Sound Sources* (VDM, Saarbrücken 2008)
18.26 W. Fohl: The wave field synthesis lab at the HAW Hamburg. In: *Sound-Perception-Performance*, ed. by R. Bader (Springer, Berlin, Heidelberg 2013) pp. 243–255
18.27 W.P.J. de Bruijn: *Application of Wave Field Synthesis in Videoconferencing*, Ph.D. Thesis (Univ. Technology Delft, Delft 2004)
18.28 S. Spors, A. Kuntz, R. Rabenstein: An approach to listening room compensation with wave field synthesis. In: *Proc. 24th Int. Conf. Multichannel Audio, Audio Eng. Soc. Conf. The New Reality* (2003)
18.29 M. Geier, H. Wierstorf, J. Ahrens, I. Wechsung, A. Raake, S. Spors: Perceptual evaluation of focused sources in wave field synthesis. In: *Proc. 128th Audio Eng. Soc. Conv.* (2010) p. 5
18.30 Fraunhofer IDMT: Audi sound concept, http://www.idmt.fraunhofer.de/content/dam/idmt/de/Dokumente/Publikationen/Referenzflyer/Audi_DE.pdf (2015)
18.31 Sonic Emotion: http://www2.sonicemotion.com (2005)
18.32 F.P. Mechel: General linear fluid acoustics. In: *Formulas of Acoustics*, 2nd edn., ed. by F.P. Mechel (Springer, Berlin, Heidelberg 2008) pp. 5–58
18.33 H. Teutsch: *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition* (Springer, Berlin, Heidelberg 2007)
18.34 W. Wöhe: Grundgleichungen des Schallfeldes und elementare Ausbreitungsvorgänge. In: *Taschenbuch Akustik. Teil 1*, ed. by W. Fasold, W. Kraak, W. Schirmer (Verlag Technik, Berlin 1984) pp. 23–31
18.35 J. Baird, P. Meyer, J. Meyer: Far-field loudspeaker interaction: Accuracy in theory and practice. In: *Proc. 110th Audio Eng. Soc. Conv.* (2001) p. 5
18.36 J. Ahrens: *Analytic Methods of Sound Field Synthesis* (Springer, Berlin, Heidelberg 2012)
18.37 J.G. Roederer: *The Physics and Psychophysics of Music. An Introduction*, 4th edn. (Springer, New York 2008)
18.38 F. Mechel: *Room Acoustical Fields* (Springer, Berlin, Heidelberg 2013)
18.39 M.B.S. Magalhães, R.A. Tenenbaum: Sound sources reconstruction techniques: A review of their evolution and new trends, Acta Acust. United Acust. **90**, 199–220 (2004)
18.40 T. Ziemer: *Implementation of the Radiation Characteristics of Musical Instruments in Wave Field Synthesis Applications*, Ph.D. Thesis (Univ. Hamburg, Staats- und Universitätsbibliothek Hamburg, Hamburg 2014)
18.41 G. Merziger, T. Wirth: *Repetitorium der höheren Mathematik*, 5th edn. (Binomi, Springe 2006)
18.42 H. Oellers: Die virtuelle Kopie des räumlichen Schallfeldes, http://www.syntheticwave.de (2010)
18.43 P.-A. Gauthier, A. Berry: Adaptive wave field synthesis for sound field reproduction: Theory, experiments and future perspectives. In: *Proc. 123rd Audio Eng. Soc. Conv.* (2007) p. 10
18.44 S. Spors: Extension of an analytic secondary source selection criterion for wave field synthesis. In: *Proc. 123rd Audio Eng. Soc. Conv.* (2007) p. 10
18.45 R. Rabenstein, S. Spors, P. Steffen: Wave field synthesis techniques for spatial sound reproduction. In: *Topics in Acoustic Echo and Noise Control. Selected Methods for the Cancellation of Acoustical Echoes, the Reduction of Background Noise and Speech Processing*, Signals and Communication Technology, ed. by E. Hänsler, G. Schmidt (Springer, Berlin, Heidelberg 2006) pp. 517–545
18.46 Y. Kim, S. Ko, J.-W. Choi, J. Kim: Optimal filtering for focused sound field reproductions using a loudspeaker array. In: *Proc. 126th Audio Eng. Soc. Conv.* (2009) p. 5
18.47 T.H. Burns: Sound radiation analysis of loudspeaker systems using the nearfield acoustic holography (nah) and the application visualization system (AVS). In: *Proc. 93rd Audio Eng. Soc. Conv.* (1992)
18.48 H. Wittek: *Perceptual Differences Between Wavefield Synthesis and Stereophony*, Ph.D. Thesis (Univ. Surrey 2007) available at http://www.hauptmikrofon.de/us/helmut-wittek/publications
18.49 E. Verheijen: *Sound Reproduction by Wave Field Synthesis*, Ph.D. Thesis (Tech. Univ. Delft 1997) available at http://repository.tudelft.nl/islandora/object/uuid:9a35b281-f19d-4f08bec7-64f6920a3821
18.50 E. Corteel, R. Nicol: Listening room compensation for wave field synthesis. What can be done? In: *Proc. 23rd Int. Conf. Signal Process. Audio Rec. Reprod., May 2003* (2003)
18.51 S. Spors, H. Buchner, R. Rabenstein, W. Herbordt: Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering, J. Acoust. Soc. Am. **122**(1), 354–369 (2007)
18.52 S. Spors: Investigation of spatial aliasing artifacts of wave field synthesis in the temporal domain. In: *Fortschr. Akust. – DAGA, Dresden* (2008)
18.53 D. de Vries: Sound reinforcement by wave field synthesis: Adaption of the synthesis operator to the loudspeaker directivity characteristics, J. Audio Eng. Soc. **44**(12), 1120–1131 (1996)
18.54 J.J. Lopez, S. Bleda, B. Pueo, J. Escolano: A subband approach to wave-field synthesis rendering. In: *Proc. 118th Audio Eng. Soc. Conv.* (2005) p. 5
18.55 S. Spors, K. Helwani, J. Ahrens: Local sound field synthesis by virtual acoustic scattering and time reversal. In: *Proc. 131st Audio Eng. Soc. Conv.* (2011)
18.56 F. Melchior: Wave field synthesis and object-based mixing for motion picture sound, SMPTE Motion Imaging J. **3**, 53–57 (2010)

18.57 S. Bleda, J.J. Lopez, J. Escolano, B. Pueo: Design and implementation of a compatible wave field synthesis authoring tool. In: *Proc. 118th Audio Eng. Soc. Conv.* (2005) p. 5

18.58 H. Wierstorf, S. Spors: Sound field synthesis toolbox. In: *Proc. 132nd Audio Eng. Soc. Conv.* (2012)

18.59 B. Bernschütz, C. Pörschmann, S. Spors, S. Weinzierl: Sofia – Sound field analysis toolbox. In: *Proc. Int. Conf. Spatial Audio, Detmold* (2011)

18.60 M. Cobos, J.J. López: Resynthesis of sound scenes on wave-field synthesis from stereo mixtures using sound source separation algorithms, J. Audio Eng. Soc. **57**(3), 91–110 (2009)

18.61 A. Murtaza, J. Herre, J. Paulus, L. Terentiv, H. Fuchs, S. Disch: Iso/mpeg-h 3-D audio: Saoc-3-D decoding and rendering. In: *Proc. 139th Audio Eng. Soc. Conv.* (2015)

18.62 B. Pueo, J.J. López, J. Escolano, L. Hörchens: Multiactuator panels for wave field synthesis: Evolution and present developments, J. Audio Eng. Soc. **58**(12), 1045–1063 (2014)

18.63 T. Ziemer: A psychoacoustic approach to wave field synthesis. In: *Proc. 42nd Int. Conf. Semantic Audio, Audio Eng. Soc. Conf., July 2011* (2011) pp. 191–197

18.64 T. Ziemer, R. Bader: Psychoacoustic sound field synthesis for musical instrument radiation characteristics, J. Audio Eng. Soc. **65**(6), 482–496 (2017)

18.65 H. Haas: Einfluss eines Einfachechos auf die Hörsamkeit von Sprache, Acustica **1**, 49–58 (1951)

18.66 E. Zwicker, H. Fastl: *Psychoacoustics. Facts and Models*, 2nd edn. (Springer, Berlin, Heidelberg 1999)

18.67 M. Beckinger, S. Brix: An efficient method to generate particle sounds in wavefield synthesis. In: *Proc. 125th Audio Eng. Soc. Conv.* (2008)

18.68 L. Rohr, E. Corteel, K.-V. Nguyen, H. Lissek: Vertical localization performance in a practical 3-D wfs formulation, J. Audio Eng. Soc. **61**(12), 1001–1014 (2013)

18.69 J. Ahrens, S. Spors: Reproduction of moving virtual sound sources with special attention to the Doppler effect. In: *Proc. 124th Audio Eng. Soc. Conv.* (2008) p. 5

18.70 G. Firtha, P. Fiala: Sound field synthesis of uniformly moving virtual monopoles, J. Audio Eng. Soc. **63**(1/2), 46–53 (2015)

18.71 R. Jacques, B. Albrecht, F. Melchior, D. de Vries: An approach for multichannel recording and reproduction of a sound source directivity. In: *Proc. 119th Audio Eng. Soc. Conv.* (2005) p. 10

18.72 J. Ahrens: Challenges in the creation of artificial reverberation for sound field synthesis, early reflections and room modes. In: *Proc. EAA Joint Symp. Auralization and Ambisonics, Berlin* (2014)

18.73 J. Ahrens: Perceptual evaluation of the diffuseness of synthetic late reverberation created by wave field synthesis at different listening positions. In: *Fortschr. Akust. – DAGA, Berlin* (2015)

18.74 H. Fastl: Praktische Anwendungen der Psychoakustik. In: *Fortschr. Akust. – DAGA, Berlin* (2010) pp. 5–10

18.75 J. Blauert: 3-D-Lautsprecher-Wiedergabemethoden. In: *Fortschr. Akust. – DAGA, Dresden* (2008) pp. 25–26

18.76 V. Pulkki: Spatial sound reproduction with directional audio coding, J. Audio Eng. Soc. **55**, 501–516 (2007)

18.77 J. Blauert: *Spatial Hearing. The Psychophysics of Human Sound Source Localization* (MIT Univ. Press, Cambridge 1997)

18.78 L.L. Beranek: *Concert Halls and Opera Houses: Music, Acoustics and Architecture*, 2nd edn. (Springer, New York 2004)

18.79 W. Kuhl: Räumlichkeit als Komponente des Raumeindrucks (spaciousness (spatial impression) as a component of total room impression), Acustica **40**, 167–181 (1978)

19. Finite-Difference Schemes in Musical Acoustics: A Tutorial

Stefan Bilbao, Brian Hamilton, Reginald Harrison, Alberto Torin

The functioning of musical instruments is well described by systems of partial differential equations. Whether one's interest is in pure musical acoustics or physical modeling of sound synthesis, numerical simulation is a necessary tool, and may be carried out by a variety of means. One approach is to make use of so-called finite-difference or finite-difference time-domain methods, whereby the numerical solution is computed as a recursion operating over a grid. This chapter is intended as a basic tutorial on the design and implementation of such methods, for a variety of simple systems. The 1-D wave equation and simple difference schemes are covered in Sect. 19.1, accompanied by an analysis of numerical dispersion and stability, as well as implementation details via vector-matrix representations. Similar treatments follow for the case of the ideal stiff bar in Sect. 19.2, the acoustic tube in Sect. 19.3, the 2-D and 3-D wave equations in Sect. 19.4, and finally the stiff plate in Sect. 19.5. Some more general nontechnical comments on more complex extensions to nonlinear systems appear in Sect. 19.6.

Systems in musical acoustics, whether they are musical instruments, electromechanical effects, or the 3-D spaces in which they are immersed, can all be described by systems of partial differential equations (PDEs). These equations are accompanied by the regions over which they are defined, coupling and boundary conditions, and forcing terms, which ultimately lead back to the player – whose behavior, as yet, is not well modeled by an equation. Sometimes, as in the case of wave propagation in a room, the equation itself is simple, but the geometry is complex; in other cases, one may have a very complex system of PDEs to solve, as in the case of a gong or a spring reverberation device, but the problem is defined over a relatively simple geometry. In either case, it is almost never possible to obtain an exact analytical solution, and thus simulation is a necessary tool. The ultimate application may be the validation of a particular musical instrument model in a scientific musical acoustics setting, or, perhaps, synthesis: the generation of sound from a numerical model.

There are many approaches to the design of a simulation for a musical instrument. Particularly in musical acoustics, a common approach involves an eigenvalue-eigenmode decomposition, leading to a representation in terms of modes of vibration. Such approaches are appropriate when the system under study is linear and time invariant. Once the modal shapes and their frequencies have been determined, a simulation can be

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_19

very easily implemented – essentially, the system is reduced to a set of uncoupled oscillators. In some cases, such as a stiff string under simply supported boundary conditions, or a rectangular or circular membrane under simply supported conditions, such shapes and frequencies are available in closed form, and the system may be simulated to a very high degree of accuracy. For more complex systems involving, say, irregular geometries, more complex boundary conditions, or perhaps coupling among various linear components, the shapes and frequencies must be determined numerically, through the solution of a (potentially large) eigenvalue problem. When nonlinearity is present, the modal shapes (if they can be thought of as such) are coupled, and the possibility of an efficient solution rapidly becomes remote.

For these reasons, it can be useful to make use of direct grid-based time-domain methods. Such methods are much more generally applicable, but one usually loses from the outset the possibility of an exact solution. Furthermore, new difficulties are introduced, many of which will be described in this chapter. Finite-difference methods are the oldest and perhaps the simplest and most straightforward means of performing a time-domain simulation – they are particularly well suited to problems without enormous geometric complexity, and some key components of musical instruments do fit this requirement nicely. These methods date back at least to the work of *Courant* et al. in the 1920s [19.1], but they were only applied to physical problems with the emergence of computers in the 1940s, and they saw major developments when applied to problems in fluid dynamics in the 1940s and 1950s (see [19.2, 3] and the references therein). Finite difference methods have also seen extensive use in electromagnetics, starting with [19.4] and increasing in popularity in the 1980s [19.5] (in this setting, such methods are often referred to as the finite-difference time-domain method, or FDTD). Finite-difference methods were first employed in musical acoustics (and indeed for sound synthesis) by *Ruiz* [19.6], and *Hiller* and *Ruiz* [19.7], in the case of strings, though earlier work by *Kelly* and *Lochbaum* on speech synthesis [19.8] led to structures that are identical to those which arise when using finite-difference methods. Such studies were followed by more research on strings [19.9], and finally, when computing power reached a level necessary for simulation in a reasonable amount of time, a large body of work emerged, first with *Boutillon* [19.10] and *Chaigne* and *Askenfelt* [19.11], in the case of the hammer string interaction, and then continuing with work on finite-difference and waveguide mesh structures (which can ultimately be viewed as finite-difference methods), for the wave equation in both 2-D [19.12, 13] and 3-D [19.14–16], for membrane vibration and room acoustics applications. See [19.17] for an overview of more recent developments.

This chapter is intended as a basic tutorial on the design and implementation of such methods, for a variety of simple systems. The 1-D wave equation and simple FD schemes are covered in Sect. 19.1, accompanied by an analysis of numerical dispersion and stability, as well as implementation details via vector-matrix representations. Similar treatments follow for the case of the ideal stiff bar, in Sect. 19.2, the acoustic tube in Sect. 19.3, the 2-D and 3-D wave equations in Sect. 19.4, and finally the stiff plate in Sect. 19.5. Some more general nontechnical comments on more complex extensions to nonlinear systems appear in Sect. 19.6.

19.1 The 1-D Wave Equation

The natural starting point for any discussion of numerical methods in musical acoustics has to be the 1-D wave equation [19.18], which may be written as

$$\partial_t^2 u = c^2 \partial_x^2 u \,. \tag{19.1}$$

This is a second-order partial differential equation, in both time t and a spatial coordinate x; $u(x,t)$ is the solution and c is the wave speed, in m/s. Here, ∂_t and ∂_x represent partial differentiation with respect to t and x respectively. Normally, t is defined for $t \in \mathbb{R}^+$, and x is defined over an interval $x \in \mathcal{D} \subset \mathbb{R}$. For practical purposes, it is the interval $\mathcal{D} = \mathcal{D}_L = [0\,, L]$, for some length $L > 0$. For some analysis purposes, it is also useful to examine the solution over an unbounded domain, i. e., $\mathcal{D} = \mathbb{R}$.

The equation above needs to be complemented by initial and boundary conditions. For initial conditions, it is usual to specify the values of u and $\partial_t u$ at time $t = 0$, i. e.,

$$u(x,0) = u_0(x) \quad \partial_t u(x,0) = v_0(x) \,. \tag{19.2}$$

If the domain is a finite interval such as $\mathcal{D}_L$, one boundary condition must be specified at either end of the domain (i. e., at $x = 0$ and $x = L$). Taking just the left end of the domain at $x = 0$ as an example, here are two of the most commonly encountered boundary

conditions

$$u(0,t) = 0 \quad \text{(Dirichlet)}$$
$$\partial_x u(0,t) = 0 \quad \text{(Neumann)} \,. \tag{19.3}$$

There are, of course, many other conditions that one may apply.

The second-order form above arises in different settings, and is normally derived from a more fundamental first-order system. For example, in the case of linear string vibration, the system is

$$\rho \partial_t v = \partial_x f \quad \partial_t f = T \partial_x v \,, \tag{19.4}$$

where here, $v(x,t)$ is transverse string velocity, $f(x,t)$ is vertical force, ρ is the linear mass density of the string in kg/m, and T is the nominal tension. In this case, one has $c = \sqrt{T/\rho}$. As another example, in the case of linear and lossless acoustic field equations in a uniform cylindrical tube, the system is

$$\rho_0 \partial_t v = -\partial_x p \quad \partial_t p = -\rho_0 c^2 \partial_x v \,, \tag{19.5}$$

where here, $v(x,t)$ is particle velocity, $p(x,t)$ is pressure deviation from atmospheric, and ρ_0 is air density in kg/m^3.

First-order systems such as (19.4) and (19.5) are one point of departure for numerical methods, leading to a formulation with staggered grids in time and space [19.2, 4]. In the mechanical setting, second-order forms such as (19.1) are more commonly encountered, and will be the main focus here. In either case, the substitution of one member of the first-order system into the other leads to the 1-D wave equation, in any of the constituent variables.

19.1.1 Behaviour of Solutions

Before moving directly to a simulation setting though, it is always useful to have some insights into the behavior of the solution.

Solution Characteristics: Traveling Waves

It is well known that the 1-D wave equation possesses a solution in terms of traveling waves. Considering the case of an unbounded domain $\mathcal{D} = \mathbb{R}$, the solution may be written directly as

$$u = u_+(x - ct) + u_-(x + ct) \tag{19.6}$$

in terms of two arbitrary functions u_+ and u_-, which represent waves traveling with constant speed c to the right and left respectively. Such a decomposition, which is peculiar to the 1-D wave equation, and not easily extended to more complex systems, has served as the starting point for efficient digital waveguide techniques in sound synthesis applications [19.19]; see Fig. 19.1.

Solution Characteristics: Dispersion Relation

A very useful analysis tool for PDE systems that are linear, and for which there is no variation in the coefficients in either space or time, is dispersion analysis. This is true in the present case of the 1-D wave equation (19.1). The idea is to examine the problem over an unbounded spatial domain, and in particular complex exponential solutions of the form

$$u = \mathrm{e}^{j(\omega t + \beta x)} \,, \tag{19.7}$$

where here, ω is an angular frequency of oscillation in radian/s, and β is a spatial wavenumber (where $2\pi/\beta$ is wavelength, in m). ω and β can be related by the insertion of (19.7) into the wave equation (19.1), to give

$$\omega = \pm c\beta \,. \tag{19.8}$$

This relation (or rather, pair of relations) is referred to as the dispersion relation for the 1-D wave equation – the sign indicates the direction of wave propagation (i. e., to the left or right). From the dispersion relation, one may

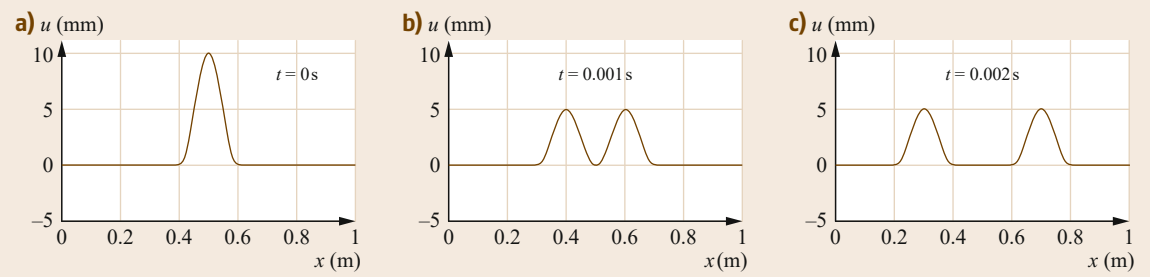


Fig. 19.1a–c Time evolution of a string profile with zero velocity, under an initial condition of a raised cosine distribution, at time instants as indicated. This illustrates the decomposition of the solution into traveling wave components. Here, the string is of length $L = 1$ m, and $c = 100$ m/s

arrive at an expression for the phase velocity, $v_{\text{phase}} = \omega/\beta$, leading to

$$v_{\text{phase}} = \pm c\,. \tag{19.9}$$

All waves thus travel at the same speed, c – this is another way of understanding the lack of distortion of traveling waves.

Solution Characteristics: Modes

For bounded domains, such as, in the present case, $\mathcal{D} = \mathcal{D}_L$, yet another means of analyzing behavior is in terms of natural frequencies or modes. The choice of boundary conditions is of major significance in this case. If the system, including boundary conditions, is lossless, then the solution may be written as

$$u(x,t) = \sum_{p=0}^{\infty} U_p(x)(a_p \sin(2\pi f_p t) + b_p \cos(2\pi f_p t)) \tag{19.10}$$

for some functions $U_p(x)$, and a set of modal frequencies f_p as well as constants a_p and b_p, which depend on initial conditions. (In fact, this is not quite true: for such a representation to hold, it is required, furthermore, that there is no degeneracy of the modal frequencies.)

For example, under Dirichlet conditions (19.3) at both ends of the domain, the series of modal functions and frequencies is given by

$$U_p(x) = \sin\left(\frac{\pi p x}{L}\right), \quad f_p = \frac{cp}{2L}, \quad p = 1, \ldots \tag{19.11}$$

and for Neumann conditions at both ends,

$$U_p(x) = \cos\left(\frac{\pi p x}{L}\right), \quad f_p = \frac{cp}{2L}, \quad p = 0, \ldots\,. \tag{19.12}$$

For Neumann conditions, note that there is an *extra* modal shape when $p = 0$, corresponding to zero-frequency or DC rigid body motion – in terms of a string, for example, when both boundaries are free to move, the string may drift away. This is clearly ruled out under fixed or Dirichlet conditions.

Figure 19.2 shows the frequency response of a string under Dirichlet conditions, and illustrates the equally spaced set of frequency components.

19.1.2 A Grid and Difference Operators

The first step in the construction of a numerical scheme simulating (19.1) is the definition of a grid function u_l^n, representing an approximation to $u(x,t)$ at time instants $t = nk$, for integer $n = 0, \ldots$, and $x = lh$ for integer $l \in d$, for some subset of the integers $d \subset \mathbb{Z}$. Here, k is the time step (and $F_s = 1/k$ is the sample rate in acoustics and audio applications, usually set a priori), and h is the grid spacing. For a grid corresponding to the interval $\mathcal{D}_L$, of length L, the finite discrete domain $d = d_N$ may be defined as $d_N = \{0, \ldots, N\}$, for an integer N such that $h = L/N$. For analysis purposes, the infinite set of grid points $d = \mathbb{Z}$ is also useful. See Fig. 19.3 for a representation of this particular 1-D grid. As will be seen shortly, once a finite-difference scheme is introduced, the grid spacing h and time step k cannot in general be chosen independently!

A useful device in the construction and analysis of finite-difference schemes is the difference operator; such operators can be used in order to represent even relatively complex schemes in a compact manner. Time difference operators can be defined as

$$\begin{aligned}
\delta_{t+} u_l^n &= \frac{1}{k}\left(u_l^{n+1} - u_l^n\right) \\
\delta_{t-} u_l^n &= \frac{1}{k}\left(u_l^n - u_l^{n-1}\right) \\
\delta_{t\cdot} u_l^n &= \frac{1}{2k}\left(u_l^{n+1} - u_l^{n-1}\right).
\end{aligned} \tag{19.13}$$

All are approximations to a first time derivative at time $t = nk$; they are often referred to as forward, backward and centered first difference operators respectively. A second difference operator, approximating a second time derivative, is defined as

$$\begin{aligned}
\delta_{tt} &= \delta_{t+}\delta_{t-} \\
\Rightarrow \delta_{tt} u_l^n &= \frac{1}{k^2}\left(u_l^{n+1} - 2u_l^n + u_l^{n-1}\right).
\end{aligned} \tag{19.14}$$

Similarly, spatial operators may be defined as

$$\begin{aligned}
\delta_{x+} u_l^n &= \frac{1}{h}\left(u_{l+1}^n - u_l^n\right) \\
\delta_{x-} u_l^n &= \frac{1}{h}\left(u_l^n - u_{l-1}^n\right) \\
\delta_{x\cdot} u_l^n &= \frac{1}{2h}\left(u_{l+1}^n - u_{l-1}^n\right)
\end{aligned} \tag{19.15}$$

and

$$\begin{aligned}
\delta_{xx} &= \delta_{x+}\delta_{x-} \\
\Rightarrow \delta_{xx} u_l^n &= \frac{1}{h^2}\left(u_{l+1}^n - 2u_l^n + u_{l-1}^n\right),
\end{aligned} \tag{19.16}$$

which are approximations to first and second spatial derivatives. Notice here that for a grid function u_l^n

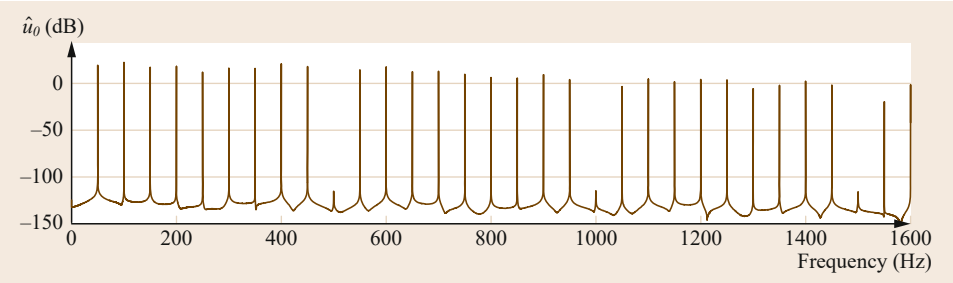


Fig. 19.2 Frequency response of the string described in the caption to Fig. 19.1, obtained by taking the Fourier transform $\hat{u}_o$ of the string displacement at a given grid location (here, approximately 1/5 of the way along the string). Here the string has been initialized with a raised cosine distribution, centered at approximately 0.3 of the length of the string. Note that the response exhibits peaks that are at integer multiples of the string fundamental, which in this case is 50 Hz

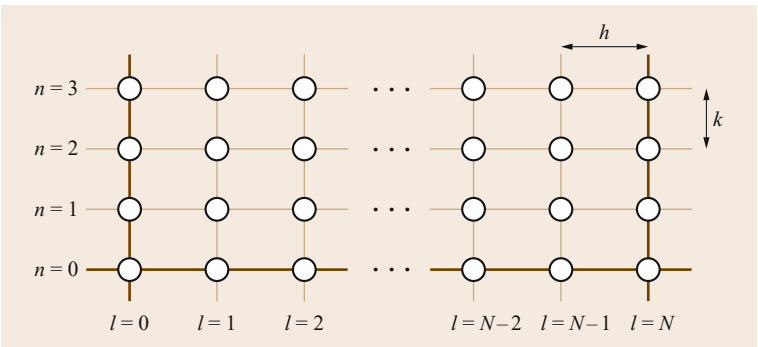


Fig. 19.3 A 1-D grid, with time step k and grid spacing h, represented for $n \geq 0$, and for $l \in d_N = \{0, \ldots, N\}$

defined for $l = 0 \ldots, N$, then at some locations these operators refer to grid points beyond the grid definition – for example, $\delta_{x-} u_0^n = \left(u_0^n - u_{-1}^n\right)/h$. In this case, one will ultimately need a means of relating the value u_{-1}^n at the *ghost* location $l = -1$ back to the domain interior – such a relation comes from an appropriate boundary condition, as will be discussed shortly.

19.1.3 A Simple Finite-Difference Scheme

Consider now the following approximation to the 1-D wave equation (19.1), obtained by approximating the second derivatives ∂_t^2 and ∂_x^2 by δ_{tt} and δ_{xx} respectively.

$$\delta_{tt} u_l^n = c^2 \delta_{xx} u_l^n \tag{19.17}$$

This is certainly the simplest possible scheme for the 1-D wave equation. When the actions of the difference operators are expanded out in full, the scheme may be written as

$$u_l^{n+1} = 2u_l^n - u_l^{n-1} + \lambda^2 \left(u_{l+1}^n - 2u_l^n + u_{l-1}^n\right) \quad \text{where } \lambda = \frac{ck}{h} . \tag{19.18}$$

This is a two-step update: the value u_l^{n+1} may be computed from previously calculated values of the grid function at time steps n and $n-1$. Furthermore, it is explicit: each unknown value of the grid function at time step $n+1$ may be calculated independently of others at the same time step. The scheme as a whole is written in terms of a single dimensionless parameter λ, which is referred to as the Courant number for the scheme.

von Neumann Analysis and Stability

A standard approach to the analysis of stability for finite-difference schemes is a frequency domain method sometimes referred to as von Neumann analysis [19.20]. In general, it yields necessary conditions for stability, and does not directly take into account boundary conditions – it may be viewed as examining the stability of a scheme over the interior of the domain, or, alternatively, when the grid is infinite.

The analysis proceeds much as in the case of dispersion analysis for the continuous problem in Sect. 19.1. Consider now the finite-difference scheme given in the previous section, defined over the infinite grid $d = \mathbb{Z}$. Now, examine the behavior of a test solution of the form

$$u_l^n = z^n \mathrm{e}^{j\beta lh} \quad \text{where} \quad z = \mathrm{e}^{sk} \tag{19.19}$$

for a real wavenumber β and a complex frequency s. Notice that when $s = j\omega$, the solution above is a sampled form of the wavelike solution given in (19.7), at $t = nk$ and $x = lh$. The complex frequency s is introduced here, as the solutions to the difference scheme (19.17) are not necessarily purely oscillatory and may exhibit exponential decay or growth – precisely when the scheme is unstable!

Inserting this solution into the definition (19.17) of the scheme leads to the following characteristic equation relating z and β

$$z + \left(-2 + 4\lambda^2 \sin^2\left(\frac{\beta h}{2}\right)\right) + z^{-1} = 0\,. \qquad (19.20)$$

In order for the scheme to be stable, the solutions must satisfy $|z| \le 1$, for all β. This will be true under the following condition on λ

$$\lambda \le 1 \;\Rightarrow\; h \ge h_{\min} = ck$$
$$\text{(Courant–Friedrichs–Lewy)}\,. \qquad (19.21)$$

If this condition is not respected, then the computed solution will be numerically unstable – this behavior is exhibited as an explosive growth in amplitude of the highest wavenumbers supported by the grid. See Fig. 19.4 for an illustration of such numerical instability, with a value of $\lambda > 1$. Such explosive growth of high frequencies or wavenumbers is typical of instability not just in the case of the wave equation, but for virtually any system in musical acoustics. The condition above, at least in this simple case of the 1-D wave equation, has a nice geometrical interpretation when rewritten as $h \ge ck$: the spacing h between adjacent grid points must be large enough to *track* waves that will have traveled a distance ck in one time step! This argument, due to Courant, is to be viewed as a heuristic in the case of explicit schemes, but is a powerful one indeed.

Numerical Dispersion and Mode Detuning

The characteristic equation (19.20), beyond yielding a stability condition, also indicates something more about the way in which waves propagate in the finite-difference scheme. Under the stability condition (19.21), it is in fact true that not only is $|z| \le 1$, but in fact $|z| = 1$. This means that solutions, when the scheme is stable, are purely oscillatory, and one may write $z = e^{j\omega k}$. The characteristic equation may then be written as

$$\sin^2\left(\frac{\omega k}{2}\right) = \lambda^2 \sin^2\left(\frac{\beta h}{2}\right)$$
$$\Longrightarrow \omega = \pm\frac{2}{k}\sin^{-1}\left(\lambda \sin\left(\frac{\beta h}{2}\right)\right)\,. \qquad (19.22)$$

Notice that this relationship is not the same as that for the continuous problem from (19.8), and consequently the phase velocity given by $v_{\text{phase}} = \omega/\beta$ is not equal to c! In short, wave speed is now frequency dependent. This characteristic is sometimes referred to as numerical dispersion, leading to progressive decoherence of traveling waves, and also to the detuning of modal frequencies; see Fig. 19.5. As illustrated in the figure, the effect becomes progressively worse as λ is decreased away from its limiting value at $\lambda = 1$, and thus it is a good idea to choose h, for a given k, as close to the bound as possible. An interesting, but pathological feature here is that when $\lambda = 1$, the scheme produces an exact solution – a great rarity for a numerical method, and one that has been exploited to great effect in the development of digital waveguide methods [19.19]. For virtually all other systems in musical acoustics, such an exact numerical solution is not available.

Another useful piece of information that the numerical dispersion relation (19.22) provides is a numerical cutoff frequency f_c, which is, in Hz,

$$f \le f_c = \frac{1}{\pi k}\sin^{-1}(\lambda)\,. \qquad (19.23)$$

Notice that the cutoff is below half the sample rate $F_s = 1/k$ – and in fact, substantially so when λ is far

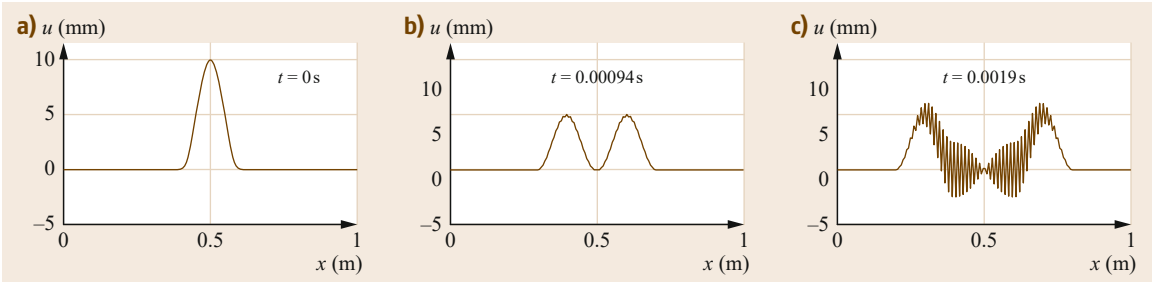


Fig. 19.4a–c Time evolution of a string profile, under conditions as given in the caption to Fig. 19.1, using the simple finite difference scheme, operating at 16 kHz, and with $\lambda = 1.00625$, leading to numerical instability

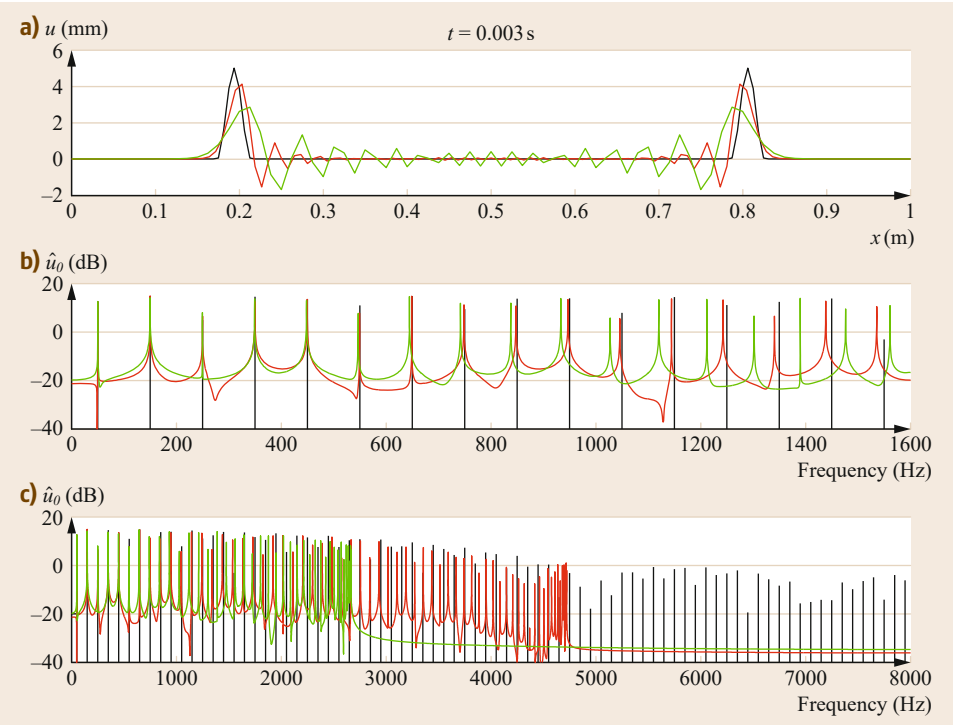


Fig. 19.5a–c Simulation results, for the string of parameters as given in the caption to Fig. 19.1, using scheme (19.17) with different values of the Courant number λ: $\lambda = 1$ (*black*), $\lambda = 0.8$ (*red*), and $\lambda = 0.5$ (*green*). **(a)** Dispersion effects in traveling waves. **(b)** Mode detuning. **(c)** Bandwidth limitation effects

from the limiting value of $\lambda = 1$. This is another good reason for choosing λ as close to the bound as possible – so as not to impose a bandwidth limitation on the scheme, as is easily observed in Fig. 19.5.

Numerical Boundary Conditions

Up to this point, only the case of the wave equation defined over an infinite domain has been treated. In practice, however, one must necessarily work with finite domains such as the $N+1$ point interval $d = d_N = \{0 \ldots, N\}$, where it is to be recalled that $h = L/N$. Now, one is faced with terminating the scheme (19.17) at the endpoints $l = 0$ and $l = N$, requiring the imposition of numerical boundary conditions.

Consider first the Dirichlet type condition for the 1-D wave equation given in (19.3), at the left end point at $x = 0$. In the scheme, this corresponds to the grid location at $l = 0$, and thus the obvious condition to set is

$$u_0^n = 0 \qquad (19.24)$$

permanently for all n, and similarly at the other endpoint at $l = N$. Thus the scheme (19.17) need only be updated at the locations $l = 1, \ldots, N-1$.

Slightly less trivial is the case of a Neumann-type condition, as given in (19.3), again, say, at the left endpoint at $l = 0$. In order to be updated at grid location $l = 0$, the scheme (19.17) requires a value at $l = -1$, which is not part of the grid interior! There are various ways of discretizing the Neumann condition. Two examples are

$$\begin{aligned} u_{-1}^n &= u_0^n \quad \text{(non-centered)} \\ u_{-1}^n &= u_1^n \quad \text{(centered)} \,. \end{aligned} \qquad (19.25)$$

The scheme update (19.17) becomes, at the grid location $l = 0$

$$u_0^{n+1} = 2u_0^n - u_0^{n-1} + \lambda^2 \left(u_1^n - u_0^n\right) \quad \text{(non-centered)} \qquad (19.26)$$

$$u_0^{n+1} = 2u_0^n - u_0^{n-1} + 2\lambda^2 \left(u_1^n - u_0^n\right) \quad \text{(centered)} \,. \qquad (19.27)$$

Both are viable numerical boundary conditions approximating the Neumann condition, and a similar condition may be applied, by symmetry, at the right end grid point at $l = N$ – the conditions are also distinct, meaning that computed solutions will be slightly different. In fact, the centered condition turns out to be a little more accurate.

These boundary conditions are simple – and indeed, all the boundary conditions given here lead to stable numerical solutions. In more involved settings however,

showing stability under a particular choice of numerical boundary condition can become far from trivial. von Neumann analysis may be extended, in the 1-D case, to handle numerical boundary conditions through the theory due to Gustafsson, Kreiss, Sundstrom and Osher (GKSO), which is explained in detail in the book by *Strikwerda* [19.20]. In more general settings in higher dimensions, where geometrical irregularities of the domain may play a role, then other techniques such as the energy method (not discussed in this chapter due to space considerations) are far more powerful. See, e.g., [19.21], and also [19.17].

Vector–Matrix Representation

Now that the scheme has been defined over a finite domain, it is perhaps worth spending some time describing the implementation of such a scheme. A very useful and compact way of representing finite difference schemes such as (19.17) is in a vector-matrix form. To this end, one should consolidate all the values of the grid function u_l^n at time step n that need to be updated into a column vector $\boldsymbol{u}^n$. In the case of Dirichlet conditions at both ends of the domain, the vector will have $N-1$ elements

$$\boldsymbol{u}^n = \left[u_1^n, \ldots, u_{N-1}^n\right]^{\mathrm{T}} .$$

For Neumann conditions at both ends of domain, the vector will contain $N+1$ elements

$$\boldsymbol{u}^n = \left[u_0^n, \ldots, u_N^n\right]^{\mathrm{T}} .$$

The difference operator δ_{xx} acting on the grid function u_l^n can clearly be represented as a square matrix $\mathbf{D}_{xx}$ when applied to the vector $\boldsymbol{u}^n$. Its form, however, depends on the numerical boundary conditions that have been chosen, which alter values in the extreme rows and columns of the matrix representation. Here are three distinct representations of $\mathbf{D}_{xx}$

$$\underbrace{\frac{1}{h^2}\begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -2 \end{bmatrix}}_{\text{Dirichlet}}$$

$$\underbrace{\frac{1}{h^2}\begin{bmatrix} -1 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -1 \end{bmatrix}}_{\text{non-centered Neumann}}$$

$$\underbrace{\frac{1}{h^2}\begin{bmatrix} -2 & 2 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 2 & -2 \end{bmatrix}}_{\text{centered Neumann}} . \qquad (19.28)$$

The matrix corresponding to Dirichlet conditions is of size $(N-1)\times(N-1)$, and those corresponding to Neumann conditions are of size $(N+1)\times(N+1)$.

Regardless of the form of the difference matrix $\mathbf{D}_{xx}$, however, the scheme (19.17) may be written in vector-matrix update form as

$$\begin{aligned} \boldsymbol{u}^{n+1} &= \mathbf{B}\boldsymbol{u}^n - \boldsymbol{u}^{n-1} \\ \mathbf{B} &= 2\mathbf{I} + c^2k^2\mathbf{D}_{xx} , \end{aligned} \qquad (19.29)$$

where $\mathbf{I}$ is an identity matrix of the size of $\mathbf{D}_{xx}$. The update matrix $\mathbf{B}$ can be computed before entering into the runtime loop – furthermore, it is very sparse, with $O(N)$ nonzero entries.

19.2 The Ideal Bar Equation

The 1-D wave equation is very much an idealization – in the case of a string for example, it describes linear, lossless wave propagation and an assumption is that the material making up the string has no inherent stiffness, so that the restoring force is purely due to tensioning.

The opposite case, in which there is no tension in the medium, but a large stiffness, leads to models of metal bars, which are a starting point for models of musical instrument components such as xylophone or marimba bars. In the simplest case of a uniform, lossless and untensioned bar, and assuming that the bar is thin (relative to its length), transverse vibrations may be described by the ideal bar equation

$$\partial_t^2 u = -\kappa^2 \partial_x^4 u . \qquad (19.30)$$

Notice now that the PDE remains second-order in time, but is fourth-order in the spatial variable. The constant

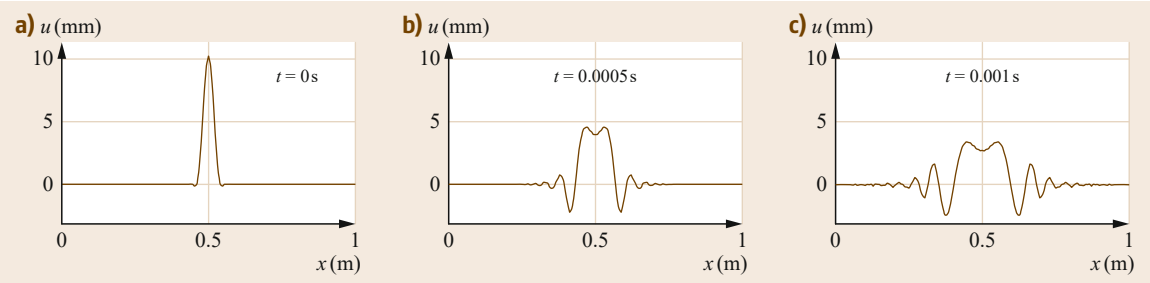


Fig. 19.6a–c Snapshots of the time evolution of the profile of a vibrating bar, at times as indicated. Here, the bar is of length $L = 1$, with $\kappa = 1\,\mathrm{m}^2/\mathrm{s}$, and is initialized with a raised cosine distribution

κ is defined by

$$\kappa = \sqrt{\frac{EI}{\rho}}\,, \tag{19.31}$$

where E is Young's modulus in Pa, ρ is again the linear mass density of the bar in kg/m, and I is the moment of inertia of the bar in m^4, which depends on the geometry of the cross-section. For a bar of circular cross-section and radius r, $I = \pi r^4/4$. Thin bar models have been shown to be more than sufficient for virtually all cases of interest in musical acoustics [19.22].

19.2.1 Solution Characteristics: Ideal Bar Equation

The ideal bar equation requires, as before, two initial conditions, of the form of (19.2). However, if the bar is defined over a finite interval such as $\mathcal{D}_L$, two boundary conditions are required at each endpoint. Examining the left endpoint at $x = 0$, here three sets that are of interest are

$$\begin{aligned} u = \partial_x u = 0 \quad &\text{(Clamped)} \\ u = \partial_x^2 u = 0 \quad &\text{(Pivoting)} \\ \partial_x^2 u = \partial_x^3 u = 0 \quad &\text{(Free)}\,. \end{aligned} \tag{19.32}$$

Solution Characteristics

In contrast to the case of the 1-D wave equation, the ideal bar equation does not possess wave solutions that travel without distortion. Indeed, wave propagation is highly dispersive, as can be seen from the dispersion relation in this case, which is, again examining wavelike solutions of the form of (19.7)

$$\omega = \pm\kappa\beta^2 \quad \Rightarrow \quad v_{\mathrm{phase}} = \kappa\beta = \sqrt{\kappa\omega}\,. \tag{19.33}$$

Here, wave speed is dependent on frequency, with higher frequencies traveling faster than lower frequencies. Figure 19.6 illustrates dispersion in a vibrating bar subject to an initial fixed distribution.

The modal frequencies of the bar are dependent on the boundary conditions in a nontrivial way – in general, they are not available in closed form. Under the pivoting conditions given in (19.32), however, they are available as

$$f_p = \frac{\kappa\pi p^2}{2L^2} \quad \text{for} \quad p = 1, \ldots\,. \tag{19.34}$$

Notice that they are not equally spaced, as in the case of the 1-D wave equation, but become more sparse with increasing frequency – this feature leads (perhaps counterintuitively) to decreases in computational cost for stiff systems in comparison with nonstiff systems; see Fig. 19.7.

19.2.2 Finite-Difference Schemes

Suppose again, for the moment, that the bar is defined over $\mathcal{D} = \mathbb{R}$, so that a corresponding difference scheme will be defined over $d = \mathbb{Z}$, again using a grid function u_l^n.

As before, the operator ∂_t^2 may be approximated by a second time difference operator δ_{tt}. For the fourth spatial derivative ∂_x^4, one may use the operator product $\delta_{xx}\delta_{xx}$, which behaves as

$$\delta_{xx}\delta_{xx}u_l^n = \frac{1}{h^4}\left(u_{l+2}^n - 4u_{l+1}^n + 6u_l^n - 4u_{l-1}^n + u_{l-2}^n\right). \tag{19.35}$$

Under these substitutions, the following scheme results

$$\delta_{tt}u_l^n = -\kappa^2\delta_{xx}\delta_{xx}u_l^n\,. \tag{19.36}$$

When the actions of the difference operators are expanded out in full, the scheme may be written as

$$\begin{aligned} u_l^{n+1} &= 2u_l^n - u_l^{n-1} \\ &\quad - \mu^2\left(u_{l+2}^n - 4u_{l+1}^n + 6u_l^n - 4u_{l-1}^n + u_{l-2}^n\right) \\ \mu &= \frac{\kappa k}{h^2}\,. \end{aligned} \tag{19.37}$$

This scheme is of a different character than (19.17) for the 1-D wave equation, in that updating at a given grid point requires access to values at neighboring grid points up to two grid spacings away; see Fig. 19.8. As one might expect, this has ramifications in terms of setting boundary conditions, but note that the ideal bar equation, as it is fourth-order in space, requires two boundary conditions at each end of the domain. The scheme is dependent on a sole parameter μ, which plays a similar role to that of the Courant number in scheme (19.17) for the wave equation, and particularly in terms of numerical stability!

Frequency Domain Analysis: Stability and Dispersion

Numerical stability analysis may be carried out in a manner similar to the case of the 1-D wave equation. Again examining discrete wavelike solutions of the form of (19.19), the characteristic equation in terms of z and β is now

$$z + \left(-2 + 16\mu^2 \sin^4\left(\frac{\beta h}{2}\right)\right) + z^{-1} = 0\,. \quad (19.38)$$

The solutions $z(\beta)$ are bounded by unity in magnitude under the condition

$$\mu \leq \frac{1}{2} \quad \Rightarrow \quad h \geq h_{\min} = \sqrt{2\kappa k}\,. \quad (19.39)$$

Note that the dependence of the grid spacing on the time step is now with the square root; the implication is that, as the sample rate is increased, the density of grid points increases much more weakly than in the case of the wave equation. This fact dovetails nicely with the notion, discussed earlier, of a reduced density of modes for increasing frequency, as illustrated in Fig. 19.7.

As in the case of the scheme for the 1-D wave equation, under stable conditions (given by (19.39)) one may again derive a numerical relation between frequency ω and wavenumber β which, in this case, is

$$\omega = \frac{2}{k} \sin^{-1}\left(2\mu \sin^2\left(\frac{\beta h}{2}\right)\right). \quad (19.40)$$

It is interesting to plot the numerical phase velocity in this case against that of the ideal bar equation itself, as shown at right in Fig. 19.8. The numerical phase velocity is too slow, and increasingly so at higher frequencies, by more than 30%! This means that, particularly for higher frequencies, modal frequencies will be underapproximated. These plots show dispersion error over the range of wavenumbers supported by the grid. One approach to obtain better accuracy is then to decrease the grid spacing, along with the time step, according to the stability condition for the scheme. But this entails operation at a higher sample rate, and can thus be a costly undertaking. More refined implicit designs can do a better job; see the end of Sect. 19.2 for a simple example.

Vector–Matrix Update Form

As in the case of schemes for the wave equation, scheme (19.36) may be written in a vector-matrix update form in the vector $\boldsymbol{u}^n$, again a column vector containing the values of the grid function u_l^n at time step n. Boundary termination is not discussed here, but as in the case of the scheme for the wave equation, values near the boundary may be omitted depending on the choice of numerical boundary condition. In all cases, however, the scheme may be written as

$$\begin{aligned} \boldsymbol{u}^{n+1} &= \mathbf{B}\boldsymbol{u}^n - \boldsymbol{u}^{n-1} \\ \mathbf{B} &= 2\mathbf{I} - \kappa^2 k^2 \mathbf{D}_{xxxx}\,, \end{aligned} \quad (19.41)$$

where $\mathbf{I}$ is an identity matrix, and where $\mathbf{D}_{xxxx}$ is the square matrix representation of the operator $\delta_{xx}\delta_{xx}$, including numerical boundary conditions.

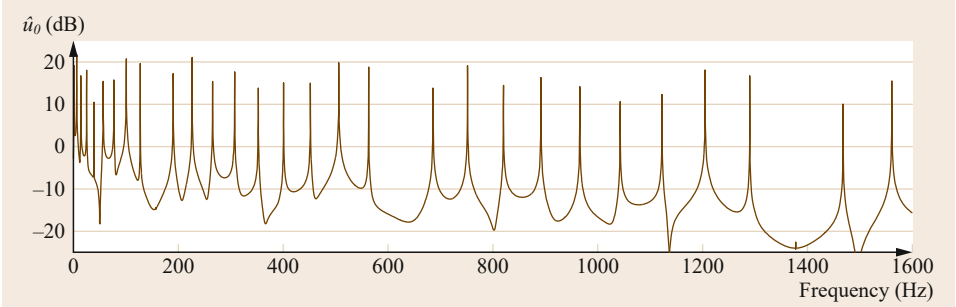


Fig. 19.7 Frequency response of the bar described in the caption to Fig. 19.6, obtained by taking the Fourier transform $\hat{u}_o$ of the bar displacement at a given grid location (here, approximately 0.2 of the way along the bar). Here the bar has been initialized with a raised cosine distribution, centered at approximately 0.3 of the length of the bar. Note the increasing spacing of modes with frequency

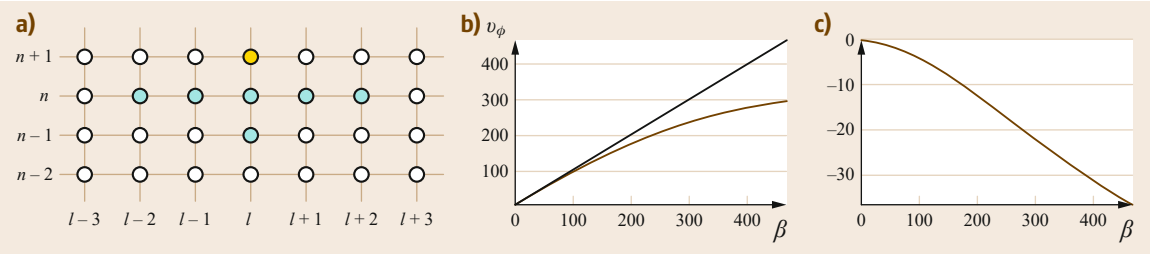


Fig. 19.8 **(a)** Computational grid and stencil for scheme (19.36) for the ideal bar equation, with the update point indicated in *yellow*, and dependency region in *cyan*. **(b)** Phase velocity for the ideal bar, with $\kappa = 1$, in *black*, and numerical phase velocity for scheme (19.36), with $\mu = 1/2$, plotted against wavenumber. **(c)** Percent error in phase velocity

Implicit Schemes

The scheme presented above is what is referred to as explicit – values of the grid function u_l^n at time step $n+1$ may be calculated directly from values at time steps n and $n-1$. An interesting variation of this scheme is the implicit case. Consider the following family of schemes, dependent on the free parameter α

$$\delta_{tt}\left(1+\alpha h^2\delta_{xx}\right)u_l^n = -\kappa^2\delta_{xx}\delta_{xx}u_l^n\,. \tag{19.42}$$

There is now a new term, involving an operation $\delta_{xx}\delta_{tt}$ – notice that because the extra term is scaled by h^2, its effect disappears in the limit of small h or small k. (This notion is referred to as *consistency* with the underlying PDE in the numerical analysis literature [19.20].) The scheme reduces to the explicit scheme when $\alpha = 0$.

Why include this extra term? Indeed, it appears to greatly complicate the update, which is now of the form

$$\begin{aligned}\alpha\left(u_{l+1}^{n+1}+u_{l-1}^{n+1}\right)+(1-2\alpha)u_l^{n+1}\\ =2\alpha\left(u_{l+1}^{n}+u_{l-1}^{n}\right)+(2-4\alpha)u_l^{n}\\ -\alpha\left(u_{l-1}^{n-1}+u_{l+1}^{n-1}\right)-(1-2\alpha)u_l^{n-1}\\ -\mu^2\left(u_{l+2}^{n}-4u_{l+1}^{n}+6u_l^{n}-4u_{l-1}^{n}+u_{l-2}^{n}\right),\end{aligned}$$

so that the unknown values of the grid function to be updated at time step $n+1$ are coupled, and cannot be independently updated; see Fig. 19.9. Furthermore, the stability condition is now altered to

$$\mu \le \frac{\sqrt{1-4\alpha}}{2} \quad \text{and} \quad \alpha \le \frac{1}{4}\,. \tag{19.43}$$

Now, the numerical dispersion relation is given by

$$\omega = \frac{2}{k}\sin^{-1}\left(\frac{2\mu\sin^2\left(\frac{\beta h}{2}\right)}{1-4\alpha\sin^2\left(\frac{\beta h}{2}\right)}\right) \tag{19.44}$$

and depends on α – and strongly as it turns out. Now there is some design flexibility, and one may attempt to optimize the numerical phase velocity of the scheme to match that of the ideal bar equation. The results can be very good indeed – see Fig. 19.9, showing numerical phase velocity and percent error for various choices of the free parameter α, and where μ is chosen at the stability limit given by (19.43). For the special choice of $\alpha = 1/4 - 1/\pi^2 \approxeq 0.1487$, the scheme has a phase velocity error of under 1% over the entire range of wavenumbers. It is probable that such a scheme can be used, without perceptual artifacts, even at a standard audio rate.

The downside of such implicit methods is, of course, that they require more computational work.

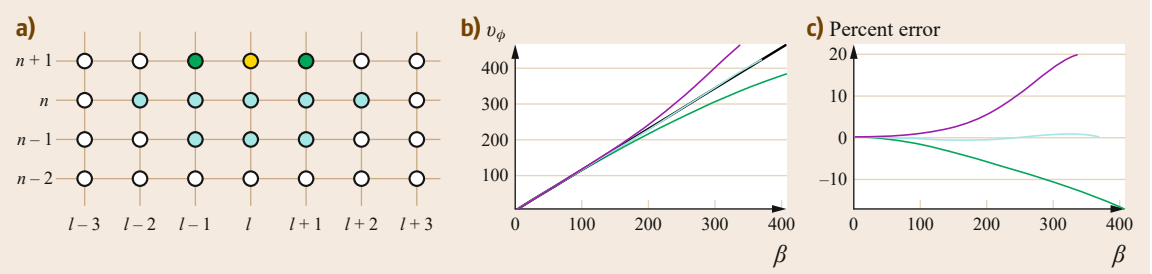


Fig. 19.9 **(a)** Computational grid for scheme (19.42) for the ideal bar equation, with the update point indicated in *yellow*, and dependency region in *cyan* (known values) and *green* (as yet unknown values). **(b)** Phase velocity for the ideal bar, with $\kappa = 1$, in *black*, and numerical phase velocity for scheme (19.42), with $\mu = 1/2$, plotted against wavenumber, for $\alpha = 0.1$ (*green*), $\alpha = 0.1487$ (*cyan*) and $\alpha = 0.18$ (*magenta*). **(c)** Percent error in phase velocity for the three choices of α

Consider the vector-matrix update form which, in contrast with (19.41), is now

$$\mathbf{A}\boldsymbol{u}^{n+1} = \mathbf{B}\boldsymbol{u}^{n} - \mathbf{A}\boldsymbol{u}^{n-1}$$
$$\mathbf{B} = 2\mathbf{A} - \kappa^2 k^2 \mathbf{D}_{xxxx}, \quad \mathbf{A} = \mathbf{I} + \alpha h^2 \mathbf{D}_{xx}\,, \tag{19.45}$$

where $\mathbf{D}_{xx}$ is an approximation to δ_{xx}, including boundary conditions, and is tridiagonal in this case. Thus, the scheme requires the solution of a linear system at each time step. There are many methods for performing such an operation – in this case, if $\mathbf{A}$ is diagonally dominant, a simple method such as Jacobi iteration [19.20] may be employed. A full analysis of the utility of implicit methods must then weigh increased accuracy against the cost of performing such linear system solutions – a typical trade-off in numerical method design.

19.3 Acoustic Tubes

So far in this chapter only spatially uniform systems, such as the 1-D wave equation and the ideal bar equation, have been considered – but there are various musical instrument components for which this not the case. The resonating components of wind instruments (woodwind and brass) and the human voice are characterized as volumes of air enclosed within a duct, or acoustic tube, and are indeed of variable cross-section [19.23]. Such instruments can often be of a complex form, incorporating bends of tubing present in, e.g., a trumpet. For applications in musical acoustics, the geometry can usually be simplified to that of a straight tube, aligned with a spatial coordinate x, and with axially symmetric cross-section $S(x)$, such as that presented in Fig. 19.10. These models are derived under the assumption that the wavelengths of interest are larger than the tube radius – which is true in most brass instruments, except in those with a very wide bell flare.

For lossless, small-amplitude wave propagation in a duct for which wavelengths are significantly larger than the radius, the dynamics of an acoustic tube can be described by *Webster*'s equation [19.24]

$$S\partial_t^2 \Psi = c^2 \partial_x (S \partial_x \Psi)\,, \tag{19.46}$$

where $\Psi(x,t)$ is often called the *acoustic potential* or *velocity potential* [19.18] and is related to the deviation of air pressure from atmospheric $p(x,t)$ and particle velocity $v(x,t)$ by

$$p = \rho \partial_t \Psi, \quad v = -\partial_x \Psi\,. \tag{19.47}$$

These definitions can be used to relate the second-order form of Webster's equation (19.46) to the first-order form (19.5) presented earlier in the case of a cylinder.

Equation (19.46) describes one-parameter waves on isophase surfaces within the tube; it is assumed that its value is constant across the tube area at position x along the tube. In the case of a cylindrical tube, where the tube cross-section is a constant, Webster's equation reduces to the 1-D wave equation. It can also be shown that for a cone described in spherical coordinates, Webster's equation reduces again to the 1-D wave equation.

Due to the spatially varying nature of the system, dispersion analysis cannot be carried out in the same manner as for the 1-D wave equation and the ideal bar equation – the main reason being that a decomposition of the solution into uniform traveling wave components is not very revealing! However, the effects of dispersion can be observed, locally, in the system. See Fig. 19.11 for an example of back-scattering of a wave approaching a constriction inside an acoustic tube.

In this section, Webster's equation is assumed defined for $t \in \mathbb{R}^+$, and $x \in \mathcal{D}_L = [0\,, L]$. As in the case of the 1-D wave equation, one boundary condition is required at each end of the domain.

The simplest realistic boundary conditions for this system are when an end is either open ($p = 0$) or closed ($v = 0$)

$$\partial_t \Psi = 0 \quad \text{(Open)}$$
$$\partial_x \Psi = 0 \quad \text{(Closed)}\,, \tag{19.48}$$

where the open end corresponds to a Dirichlet boundary condition and the closed end to a Neumann boundary

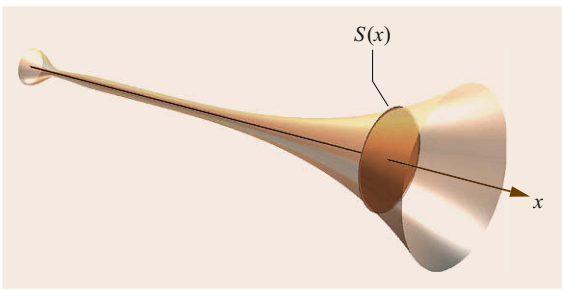


Fig. 19.10 Example of a bore with variable cross-sectional area along its length

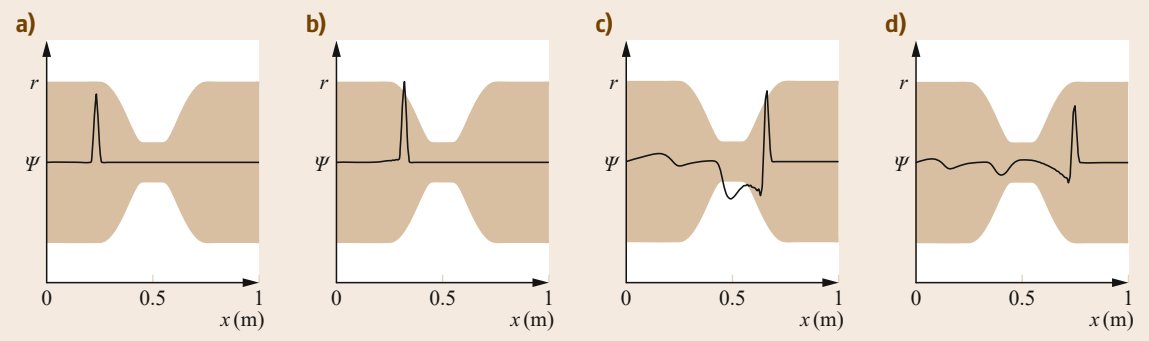


Fig. 19.11a–d Propagation of a pulse through an acoustic tube with a constriction at its center. The wave propagates from left to right and is shown in *black*. The *brown* region shows the tube profile radius $r(x) = \sqrt{S(x)/\pi}$. As the wave encounters the constriction, back-scattering of low frequencies can be seen as disturbances traveling in the opposite direction to that of the impinging wave

condition. Both of these boundary conditions are lossless – the total energy of the system is unchanged upon reflection at the boundaries. In realistic wind instrument modeling more complex boundary conditions are required, in particular at the radiating end where losses due to acoustic radiation must be modeled, and at the mouthpiece where complex coupling to an active device (the player's lips or a reed mechanism, driven by an input pressure) must be modeled as well.

19.3.1 Finite-Difference Schemes

It is rather straightforward to arrive at a finite-difference scheme simulating Webster's equation – note that as the system is not spatially uniform, it is to be expected that this feature will be carried over to the resulting scheme, so that the scheme parameters themselves vary from one grid point to the next. The simplest possible scheme, operating over d_N, is almost certainly

$$\bar{S}_l \delta_{tt} \Psi_l^n = c^2 \delta_{x+} \left(S_{l-1/2} \left(\delta_{x-} \Psi_l^n \right) \right) , \tag{19.49}$$

where δ_{tt}, δ_{x+} and δ_{x-} are the second-order time and first-order forward and backward spatial difference operators defined in Sect. 19.1.2. The scheme is dependent on two fixed grid functions S and $\bar{S}$, which approximate the tube cross-sectional area at interleaved locations. $S_{l-1/2}$ is assumed given, and perhaps sampled from an exact bore profile, $S(x)$, at positions $x = (l-1/2)h$. $\bar{S}_l$ is assumed averaged from the bore cross-sectional area

$$\bar{S}_l = \frac{S_{l+1/2} + S_{l-1/2}}{2} . \tag{19.50}$$

The update form of (19.49) is

$$\Psi_l^{n+1} = \frac{\lambda^2 S_{l+1/2}}{\bar{S}_l} \Psi_{l+1}^n + \frac{\lambda^2 S_{l-1/2}}{\bar{S}_l} \Psi_{l-1}^n + 2(1-\lambda^2)\Psi_l^n - \Psi_l^{n-1} , \tag{19.51}$$

where $\lambda = ck/h$ is again the Courant number for the scheme. Figure 19.12 shows the scheme update and the discretization locations of $S_{l+1/2}$ and $\bar{S}_l$. Due to the

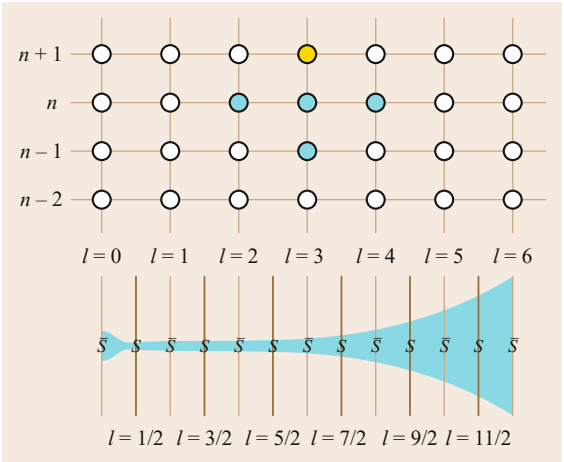


Fig. 19.12 *Top*: Computational grid and stencil for scheme (19.49), with current point in *yellow*, and dependency region in *cyan*. *Horizontal lines* show the time grid at integer time steps, separated by a time step k. *Vertical lines* show the spatial grid at integer spatial steps, separated by a distance h. *Bottom*: acoustic tube cross-section in *cyan* overlaid on a spatial grid. *Solid vertical lines* indicate the spatial grid at half-integer steps, separated by a distance h. Note that the positions of $S_{l+1/2}$ are interleaved with the positions of $\bar{S}_l$

spatial variation of the cross-section in this model, von Neumann analysis cannot be applied to the scheme in order to prove stability. Energy methods may be used for the scheme to show that $\lambda \leq 1$ is again a stability condition [19.17] – such a simplified condition, which is independent of the bore profile, depends on the particular choice of $\bar{S}$ as averaged from S.

Considering boundary conditions at $x = 0$, a numerical open (Dirichlet) condition corresponding to the first of (19.48) may be given simply as

$$\Psi_0^n = 0 \tag{19.52}$$

and for a closed (Neumann) tube corresponding to the second of (19.48) as

$$\begin{aligned} \Psi_0^n &= \Psi_{-1}^n \quad \text{(non-centered)} \\ \Psi_1^n &= \Psi_{-1}^n \quad \text{(centered)} \,. \end{aligned} \tag{19.53}$$

Vector–Matrix Representation

The scheme (19.49) can be written in a vector-matrix form similar to that presented in the case of the 1-D wave equation and ideal bar equation. Gathering all values of the grid function Ψ_l^n into a column vector $\boldsymbol{\Psi}^n$, which has $N-1$ elements for a tube open at both ends: $\boldsymbol{\Psi}^n = [\Psi_1^n, \ldots, \Psi_{N-1}^n]^{\mathrm{T}}$, and $N+1$ elements for a tube closed at both ends: $\boldsymbol{\Psi}^n = [\Psi_0^n, \ldots, \Psi_N^n]^{\mathrm{T}}$, an update can be constructed that has a similar form to that of the scheme for the 1-D wave equation and 1-D bar equation

$$\begin{aligned} \boldsymbol{\Psi}^{n+1} &= \mathbf{B}\boldsymbol{\Psi}^n - \boldsymbol{\Psi}^{n-1} \,, \\ \mathbf{B} &= 2\mathbf{I} + c^2k^2\mathbf{C} \,, \end{aligned} \tag{19.54}$$

where $\mathbf{I}$ is again an identity matrix of the appropriate size. $\mathbf{C}$ is a square matrix that contains the effect of the variation of the bore profile in scheme (19.49). As examples, the forms of this matrix are given below under three different sets of boundary conditions

$$\underbrace{\frac{1}{h^2}\begin{bmatrix} -2 & \frac{S_{3/2}}{\bar{S}_1} & & & & & \\ \frac{S_{3/2}}{\bar{S}_2} & -2 & \frac{S_{5/2}}{\bar{S}_2} & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & \frac{S_{l-1/2}}{\bar{S}_l} & -2 & \frac{S_{l+1/2}}{\bar{S}_l} & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & \frac{S_{N-5/2}}{\bar{S}_{N-2}} & -2 & \frac{S_{N-3/2}}{\bar{S}_{N-2}} \\ & & & & & \frac{S_{N-3/2}}{\bar{S}_{N-1}} & -2 \end{bmatrix}}_{\text{open}}$$

$$\underbrace{\frac{1}{h^2}\begin{bmatrix} -1 & \frac{S_{1/2}}{\bar{S}_0} & & & & & \\ \frac{S_{1/2}}{\bar{S}_1} & -2 & \frac{S_{3/2}}{\bar{S}_1} & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & \frac{S_{l-1/2}}{\bar{S}_l} & -2 & \frac{S_{l+1/2}}{\bar{S}_l} & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & \frac{S_{N-3/2}}{\bar{S}_{N-1}} & -2 & \frac{S_{N-1/2}}{\bar{S}_{N-1}} \\ & & & & & \frac{S_{N-1/2}}{\bar{S}_N} & -1 \end{bmatrix}}_{\text{non-centered closed}}$$

$$\underbrace{\frac{1}{h^2}\begin{bmatrix} -2 & 2\frac{S_{1/2}}{\bar{S}_0} & & & & & \\ \frac{S_{1/2}}{\bar{S}_1} & -2 & \frac{S_{3/2}}{\bar{S}_1} & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & \frac{S_{l-1/2}}{\bar{S}_l} & -2 & \frac{S_{l+1/2}}{\bar{S}_l} & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & \frac{S_{N-3/2}}{\bar{S}_{N-1}} & -2 & \frac{S_{N-1/2}}{\bar{S}_{N-1}} \\ & & & & & 2\frac{S_{N-1/2}}{\bar{S}_N} & -2 \end{bmatrix}}_{\text{centered closed}} \,. \tag{19.55}$$

Input Impedance

In musical acoustics investigations, a wind instrument is often described in terms of its input impedance [19.25], which is a complete characterization of the behavior of the tube in the frequency domain. The peaks of the magnitude of the input impedance curve correspond to resonances, and in the case of musical instruments give an indication of playable notes. The input impedance is defined as

$$\hat{Z}(\omega)_{\mathrm{in}} = \frac{\hat{p}(0,\omega)}{S(0)\hat{v}(0,\omega)} \,, \tag{19.56}$$

where $\hat{v}(0,\omega)$ and $\hat{p}(0,\omega)$ are the Fourier transforms of the particle velocity and acoustic pressure at $x = 0$. $S(0)$ is the cross-sectional area of the acoustic tube opening and ω is angular frequency. To test the validity of a finite-difference scheme modeling an acoustic tube, it is useful to construct an equivalent method for measuring the input impedance of the numerical acoustic tube. Defining

$$\delta_x . \Psi_0^n = -v_{\mathrm{in}}^n \tag{19.57}$$

and substituting into (19.51) for $l = 0$ gives

$$\begin{aligned} \Psi_0^{n+1} = {} & 2\lambda^2\Psi_1^n + 2(1-\lambda^2)\Psi_0^n - \Psi_0^{n-1} \\ & + \frac{2h\lambda^2 S_{-1/2}}{\bar{S}_0} v_{\mathrm{in}}^n \,, \end{aligned} \tag{19.58}$$

which is a usable termination for the scheme. Setting v_{in} as an impulsive time series (i. e., a 1 followed by zeros), and taking the Fourier transform of $p_{in} = \rho\partial_t \Psi_0$ gives a numerical approximation to the input impedance of the tube. Figure 19.13 shows some representative bore and impedance plots.

19.3.2 Energy Losses and Nonlinear Propagation

Although finite-difference schemes including energy losses and nonlinear propagation are not presented in this chapter, it is worth highlighting some modifications that can be made to the model to include these effects.

Viscothermal Losses

Viscothermal losses account for the main source of energy loss within an acoustic tube. These losses occur in the viscous and thermal boundary layers that lie along the interior of the bore with the wave as it propagates along the tube. Although a 3-D effect this can still be modeled in 1-D by using an impedance, Z, and admittance, Y, description of the acoustic tube in the frequency domain

$$\begin{aligned} \partial_x \hat{p} &= -Z\hat{v} \\ \partial_x (S\hat{v}) &= -YS\hat{p}\,. \end{aligned} \tag{19.59}$$

The frequency domain expressions Z and Y require transformation to the time domain, and various approximations are available [19.26–31].

For a second-order form of wave propagation the Webster–Lokshin model has seen extensive applications in the modeling of acoustics tubes, particularly in waveguide methods [19.32, 33].

Nonlinear Wave Propagation

Nonlinearity of wave propagation is responsible for the cuivre, or *brassy*, sound generated by brass instruments when played at high dynamic levels. This is a result of the deformation of the wave shape as it propagates along the tube, which adds additional harmonics to the sound spectrum. To model this effect, a new set of nonlinear equations are required of which there are several to choose from [19.34].

The most fundamental nonlinear model is given by the Navier–Stokes equations. These equations give a full description of the fluid, including effects from viscosity and temperature. However, due to their complexity they prove difficult to solve. The Euler equations are a set of conservation equations concerning mass, momentum and energy; they are also subset of the Navier–Stokes equations neglecting viscosity and heat conduction terms. Burger's equations can also be derived from the Navier–Stokes equations and are written in terms of one variable, often velocity.

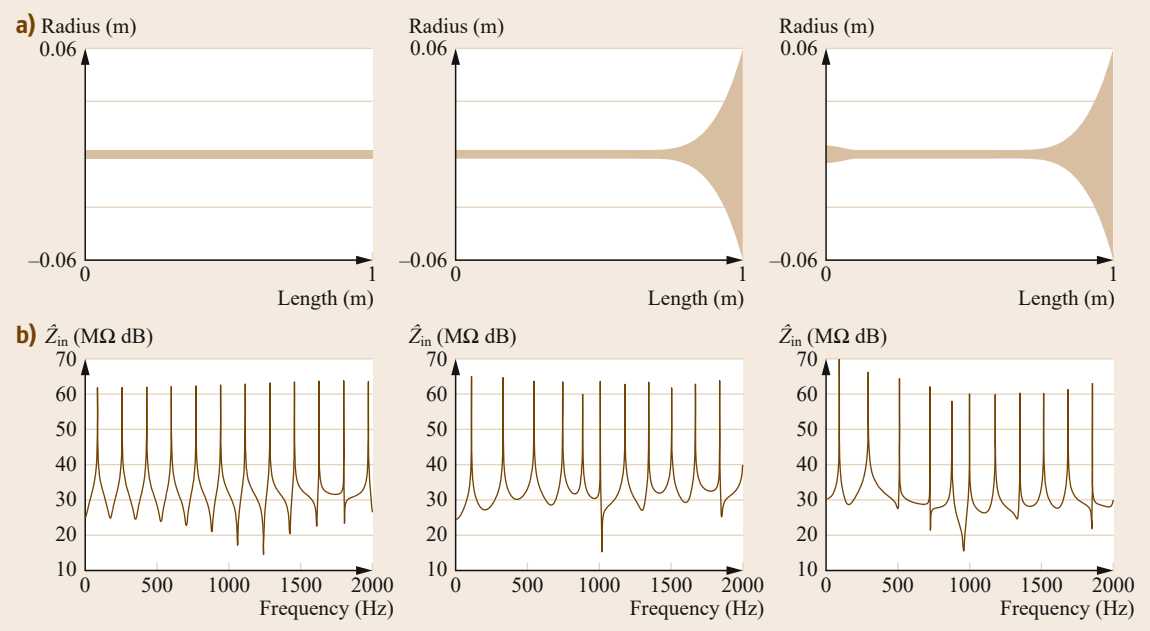


Fig. 19.13 **(a)** Selection of 1 m-long tube profiles. **(b)** Input impedance of the tubes with a closed end at $x = 0$ and an open end at $x = 1$. The tube is excited at $x = 0$ with an impulse and output is taken as the Fourier transform of $p_{in} = \rho\partial_t \Psi_0$. Simulations were run at a sample rate of 44.1 kHz and over a simulation time of 10 s with $c = 343$ m/s

All of the above models demonstrate the wave steepening required for harmonic generation as the wave propagates along the tube. However, stability and numerical distortion provide difficulties when trying to model the equations.

Complex Boundary Conditions

For musically interesting cases, more complex boundary conditions are required than those in (19.48). These new boundaries can either inject energy into the tube (such as a lip or reed model) or allow energy to leave (radiation model).

The simplest reed model is a single mass driven by the pressure difference between the mouth and the entrance of the tube [19.23]. This force causes the mass to move and change the channel height, which affects how much air is injected into the the tube and therefore the pressure difference over the boundary. This means that the motion of the reed is directly coupled to the pressure in the tube and requires the solution of simultaneous equations. Lip models for brass instruments can also be modeled in a similar way with two masses restrained to motion in one or two dimensions [19.35, 36].

The radiation properties of a tube are primarily affected by the surface area of the tube at the interface with the acoustic space but are frequency dependent. *Levine* and *Schwinger* [19.37] derived a frequency domain model for unflanged circular pipes, however this proves difficult to implement in the time domain and requires approximations.

A complete finite difference valved brass instrument synthesis system has been presented recently in [19.38, 39].

19.3.3 Relationship to Other Simulation Techniques

One of the earliest physical models of an acoustic tube is the *Kelly–Lochbaum* speech synthesis algorithm [19.8]. This algorithm assumes that the profile of the acoustic tube can be approximated by a series of joined cylinders, which reduces Webster's equation piecewise to the 1-D wave equation. This allows for traveling wave solutions in pressure and volume velocity, and by considering these values at the boundary of each cylinder segment a scattering matrix can be constructed. These scattering matrices can then be used to model the wave moving through the tube.

Related to scattering methods are digital waveguides, which have been a popular synthesis tool over the last 20 years and have been used in commercial music products such as the Yamaha VL1. Again these methods take advantage of separation of the solution into forward and backwards traveling waves and are therefore most suitable for 1-D linear systems. The traveling wave solutions are stored in two directional delay lines, which and are then shifted along their respective directions. This means that there are no arithmetic operations required, allowing a fast computation time. Excitation mechanisms can be connected to the model using scattering junctions and radiation effects can be added using filters [19.40]. Further modifications have been made to the waveguides models such as using the Webster–Lokshin model to add viscothermal effects [19.32, 33] and modeling quasispherical wavefronts [19.41].

Both the Kelly–Lochbaum and waveguide models of acoustic tubes can be shown to have finite-difference interpretations under specific conditions [19.17]. In the case of modeling the 1-D wave equation (a cylindrical acoustic tube), the digital waveguide can be shown to be equivalent to a finite-difference scheme with $\lambda = 1$. Similarly, the Kelly–Lochbaum algorithm is again a special form of a finite-difference scheme for an acoustic tube with $\lambda = 1$, where the scattering parameters are the update coefficients of the scheme.

19.4 The 2-D and 3-D Wave Equations

The ideal string, ideal bar, and acoustic tube covered in the previous sections are 1-D, meaning that their behavior may be adequately described by a PDE system written in terms of a single spatial coordinate. Such is obviously not the case for many structures of interest in musical acoustics, and certainly not when it comes to the modeling of acoustic wave propagation in real spaces!

A starting point for linear musical systems, such as membranes and also the acoustics of rooms, is the wave equation in spatial dimensions higher than one [19.18]. This second-order partial differential equation is written as

$$\partial_t^2 u = c^2 \Delta u \,. \tag{19.60}$$

Here, $u(\boldsymbol{x}, t)$ is the solution, t represents time, $\boldsymbol{x}$ is a spatial vector in 2-D or 3-D, and c is the wave speed in m/s.

The vector $\boldsymbol{x}$ may be written in terms of individual spatial coordinates as $\boldsymbol{x} = (x, y)$ in 2-D and $\boldsymbol{x} = (x, y, z)$

in 3-D. The operator Δ, usually called the Laplacian, is defined as

$$\Delta = \partial_x^2 + \partial_y^2 \quad \text{(in 2-D)}$$
$$\Delta = \partial_x^2 + \partial_y^2 + \partial_z^2 \quad \text{(in 3-D)}\,. \tag{19.61}$$

An equivalent formulation of (19.60) as a first-order system is also possible [19.18]. If (19.60) models the behavior of an ideal membrane, then u represents the transverse displacement; if it models acoustic wave propagation, then u represents a variable such as pressure or velocity potential.

As in the 1-D case, t is normally defined over $\mathbb{R}^+$, and $\boldsymbol{x}$ is defined in the domain $\mathcal{D}$, i. e., $\boldsymbol{x} \in \mathcal{D}$ where $\mathcal{D} \subseteq \mathbb{R}^2$ in 2-D and $\mathcal{D} \subseteq \mathbb{R}^3$ in 3-D. For practical purposes it is useful to define the rectangular regions $\mathcal{D}_{L_x,L_y} = [0, L_x] \times [0, L_y]$, where $\times$ denotes the Cartesian product, and $\mathcal{D}_{L_x,L_y,L_z} = [0, L_x] \times [0, L_y] \times [0, L_z]$. As before, the domains $\mathcal{D} = \mathbb{R}^2$ or $\mathcal{D} = \mathbb{R}^3$ are also useful for purposes of frequency domain analysis.

As in the 1-D case, the following initial conditions must be specified for the above equation

$$u(\boldsymbol{x}, 0) = u_0(\boldsymbol{x})$$
$$\partial_t u(\boldsymbol{x}, 0) = v_0(\boldsymbol{x})\,. \tag{19.62}$$

An initial condition that will be useful to illustrate the behavior of (19.60) is the following *raised cosine* initial condition

$$u_0 = \begin{cases} 0.5\left(1 + \cos\left(\frac{\pi|\boldsymbol{x}-\boldsymbol{x}_0|}{r_{\text{hw}}}\right)\right) & |\boldsymbol{x}-\boldsymbol{x}_0| \le r_{\text{hw}} \\ 0 & |\boldsymbol{x}-\boldsymbol{x}_0| > r_{\text{hw}} \end{cases}$$
$$v_0 = 0\,, \tag{19.63}$$

where r_{hw} is the half-width of the raised cosine and $\boldsymbol{x}_0$ is its center. Conditions along a boundary with normal $\boldsymbol{n}$ may be of the type

$$u(\boldsymbol{x}, t) = 0 \quad \text{(Dirichlet)}$$
$$\partial_n u(\boldsymbol{x}, t) = 0 \quad \text{(Neumann)}\,, \tag{19.64}$$

where ∂_n denotes a spatial derivative along the direction of $\boldsymbol{n}$. Dirichlet conditions are typically used to terminate the edge of a membrane, whereas Neumann conditions are a simple approximation to reflecting wall conditions in the setting of acoustic wave propagation.

As a first look into the behavior of the 2-D and 3-D wave equations, consider the raised cosine initial condition on the unit square $\mathcal{D} = \mathcal{D}_{1,1}$ and the unit cube $\mathcal{D} = \mathcal{D}_{1,1,1}$ domains, both with Neumann boundary conditions. The time evolution of the 2-D system is shown in Fig. 19.14. It can be seen that the raised cosine spreads out evenly over space, creating a circular wavefront. Figure 19.15 shows the time evolution of the 3-D system. It is slightly more difficult to completely visualize a 3-D field – this would require a fourth spatial dimension! – but one possibility is to view individual 2-D slices of the 3-D field, as illustrated here. Again, a circular wavefront propagates outward from the origin along the 2-D planes (a 2-D slice of a spherical wavefront in 3-D), but the wavefront decreases in amplitude more quickly in 3-D than in the case of the 2-D system.

19.4.1 Solution Characteristics

Another key difference between the behavior of the wave equation in 2-D and 3-D appears in the wake of the spreading wavefront, but this detail can be better

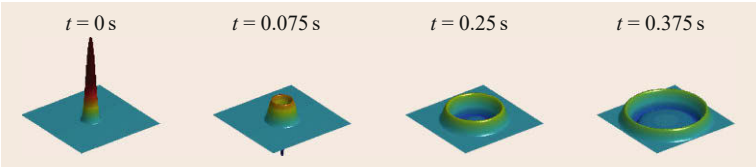


Fig. 19.14 Time evolution of a 2-D acoustical field, under a raised cosine initial condition with $r_{\text{hw}} = 0.1$, $\boldsymbol{x}_0 = (0.5, 0.5)$ and $c = 1\,\text{m/s}$ at times as indicated

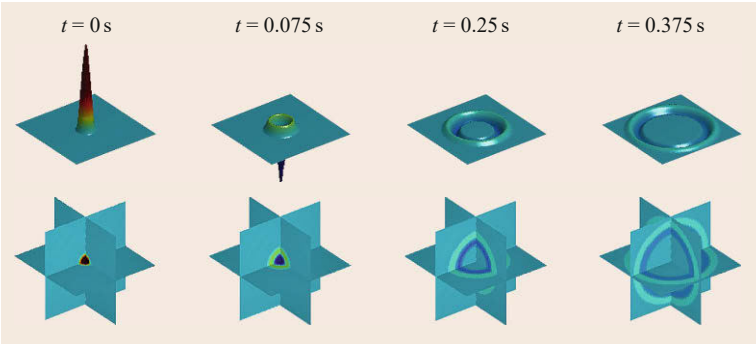


Fig. 19.15 Time evolution of a 3-D acoustic field under a raised cosine initial condition with $r_{\text{hw}} = 0.1$, $\boldsymbol{x}_0 = (0.5, 0.5, 0.5)$, and $c = 1\,\text{m/s}$, at times as indicated. *Top*: viewed along the 2-D-slice $z = 0.5$. *Bottom*: viewed along three 2-D slices: $x = 0.5$, $y = 0.5$, $z = 0.5$

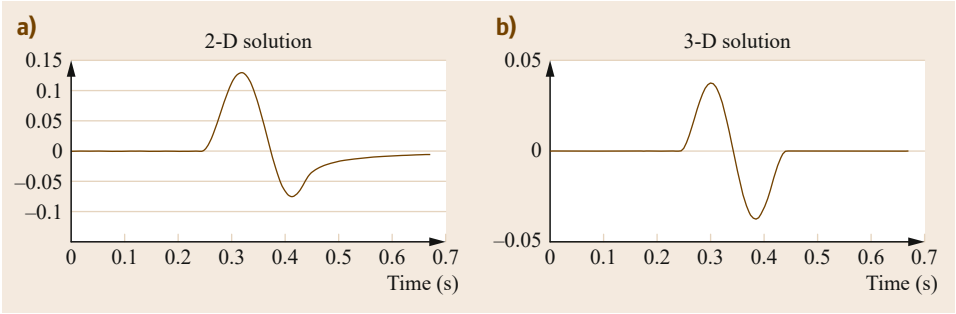


Fig. 19.16 **(a)** Time evolution of the displacement of a 2-D membrane at position $(0.25, 0.25)$ under a raised cosine initial condition with $r_{\text{hw}} = 0.1$ and $\boldsymbol{x}_0 = (0.5, 0.5)$. **(b)** Time evolution of a 3-D acoustic field at position $(0.25, 0.25, 0.5)$ with $r_{\text{hw}} = 0.1$ and $\boldsymbol{x}_0 = (0.5, 0.5, 0.5)$

seen by plotting the solution over time at a single output location. It can be seen in Fig. 19.16 that in the 2-D case the solution dips below zero after the initial pulse, but then fails to return to zero. On the other hand, in the 3-D case the solution returns to zero after the initial pulse. This behavior in 3-D, as well as the behavior in 1-D, is explained by *Huygens' principle*, which only applies in an odd number of spatial dimensions. The trailing *tail* phenomenon seen in the 2-D case is particular to even spatial dimensions. A useful musical analogy is provided in [19.42]:

> *A listener placed in* $\mathbb{R}^3$ *at distance* d *from a musical instrument hears at time* t *the note played at time* $(t-d)$ *and nothing else!* [. . .] *While in* $\mathbb{R}^2$ *he would hear a weighted average of all notes played during the time* $[0, t-d]$.

Solution Characteristics: Dispersion Relation

As in the 1-D case, it is useful to examine the behavior of the 2-D and 3-D wave equations with a plane wave solution of the form

$$u = \mathrm{e}^{j(\omega t + \boldsymbol{\beta}\cdot\boldsymbol{x})} \,. \tag{19.65}$$

Here, $\boldsymbol{\beta}$ is a wave vector in $\mathbb{R}^2$ or $\mathbb{R}^3$ representing a multidimensional spatial frequency, and its Euclidean norm (magnitude) $|\boldsymbol{\beta}|$ is known as the wavenumber, with units of radian/m. Inserting this solution into the wave equation results in the dispersion relation

$$\omega = \pm c|\boldsymbol{\beta}| \tag{19.66}$$

and the phase velocity $v_{\text{phase}} = \omega/|\boldsymbol{\beta}|$

$$v_{\text{phase}} = \pm c \,. \tag{19.67}$$

Thus, all plane waves (in 2-D and 3-D) travel with a constant wave speed c.

Solution Characteristics: Modes

Also as in the 1-D case, solutions to the 2-D and 3-D wave equations on bounded domains may be written as the sum of a set of modal functions

$$u(\boldsymbol{x}, t) = \sum_q U_q(\boldsymbol{x}) \left(a_q \sin(2\pi f_q t) + b_q \cos(2\pi f_q t)\right) . \tag{19.68}$$

Such spatial modes $U_q(\boldsymbol{x})$ may have analytic expressions, but in general these expressions are not easily derived unless the geometry of the domain $\mathcal{D}$ is regular, such as a rectangular or circular-spherical domain, and only when terminated by simple boundary conditions such as those of Neumann or Dirichlet type. On 2-D rectangular domains $\mathcal{D}_{L_x,L_y}$, under Dirichlet conditions at all boundaries, modal functions are separable into x and y components. These modes take the form

$$U_{\boldsymbol{p}}(\boldsymbol{x}) = \sin\left(\frac{p_x \pi x}{L_x}\right) \sin\left(\frac{p_y \pi y}{L_y}\right)$$
$$f_{\boldsymbol{p}} = \frac{c}{2}\left(\left(\frac{p_x}{L_x}\right)^2 + \left(\frac{p_y}{L_y}\right)^2\right)^{1/2} , \tag{19.69}$$

where $\boldsymbol{p}$ is a pair of indices, $\boldsymbol{p} = (p_x, p_y)$ with $p_x, p_y \geq 1$. Some of these spatial modes on the unit square $\mathcal{D}_{1,1}$ are illustrated in Fig. 19.17.

If, instead, Neumann conditions are used at the boundaries, one would have the following set of spatial modes on $\mathcal{D}_{L_x,L_y}$

$$U_{\boldsymbol{p}}(\boldsymbol{x}) = \cos\left(\frac{p_x \pi x}{L_x}\right) \cos\left(\frac{p_y \pi y}{L_y}\right)$$
$$f_{\boldsymbol{p}} = \frac{c}{2}\left(\left(\frac{p_x}{L_x}\right)^2 + \left(\frac{p_y}{L_y}\right)^2\right)^{1/2} , \tag{19.70}$$

where $p_x, p_y \geq 0$. These are illustrated in Fig. 19.18. The mode $\boldsymbol{p} = (0, 0)$ is known as the DC mode, and

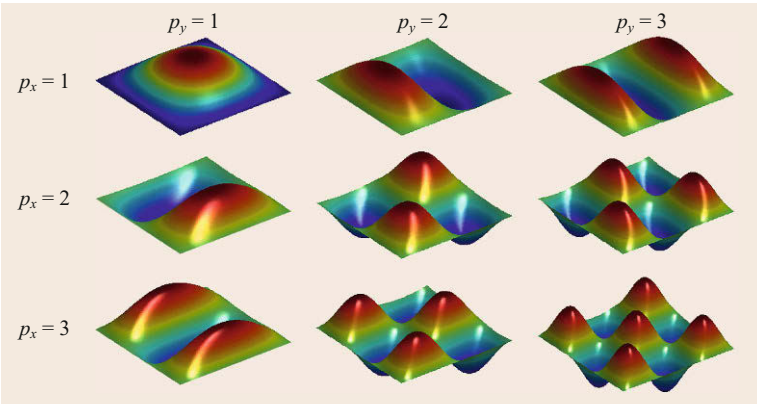


Fig. 19.17 Spatial modes of a 2-D acoustic field on $\mathcal{D}_{1,1}$ with Dirichlet boundary conditions

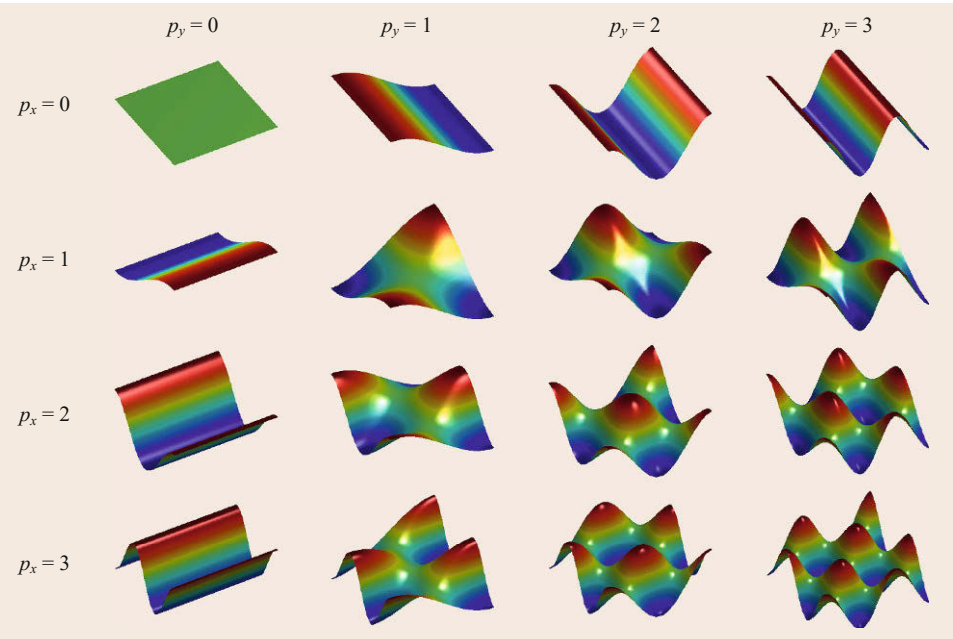


Fig. 19.18 Spatial modes of a 2-D acoustic field on $\mathcal{D}_{1,1}$ with Neumann boundary conditions

it is not permitted when Dirichlet conditions are specified at the boundaries, as well as modes $\boldsymbol{p} = (0, n)$ and $\boldsymbol{p} = (n, 0)$ where n is a positive integer. The DC mode corresponds to a linear drift in the solution, which may not seem physical in the context of musical acoustics, but is strictly valid for the model equation.

In 3-D, the modal shapes on $\mathcal{D}_{L_x,L_y,L_z}$ with respect to Neumann conditions take the form

$$U_{\boldsymbol{p}}(\boldsymbol{x}) = \cos\left(\frac{p_x \pi x}{L_x}\right) \cos\left(\frac{p_y \pi y}{L_y}\right) \cos\left(\frac{p_z \pi z}{L_z}\right)$$

$$f_{\boldsymbol{p}} = \frac{c}{2}\left(\left(\frac{p_x}{L_x}\right)^2 + \left(\frac{p_y}{L_y}\right)^2 + \left(\frac{p_z}{L_z}\right)^2\right)^{1/2}, \tag{19.71}$$

where now $\boldsymbol{p} = (p_x, p_y, p_z)$ with $p_x, p_y, p_z \geq 0$. A selection of modes on the unit cube are illustrated in Fig. 19.19.

Unlike the 1-D case, the temporal frequencies f_p that correspond to these modes are not all multiples of one fundamental frequency. This should be apparent from (19.70), (19.69) and (19.71). This can also be seen in the response of a typical room to an impulsive initial condition. Figure 19.20 shows such a response for a $10\,\mathrm{m} \times 20\,\mathrm{m} \times 30\,\mathrm{m}$ room ($\mathcal{D}_{10,20,30}$) with Neumann boundaries and $c = 340\,\mathrm{m/s}$. It can be seen that the lower frequency modes are sparse and that the density of modes increases at higher frequencies. This is a general trend, which is in contrast to the case of the 1-D wave equation where the density of modes is constant across frequencies, and the case

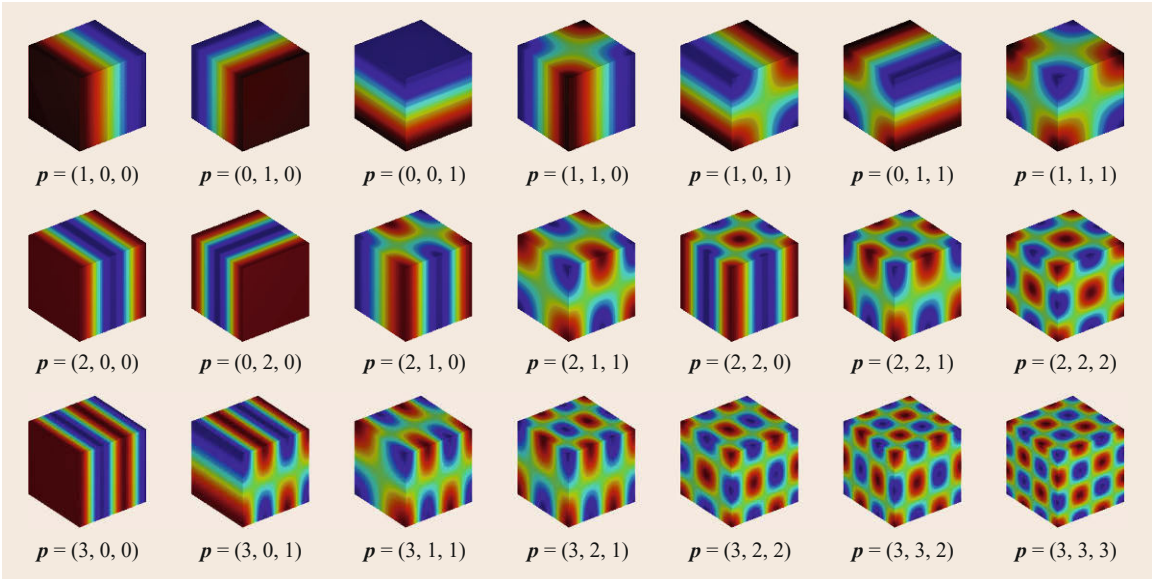


Fig. 19.19 Spatial modes of a 3-D acoustic field on $\mathcal{D}_{1,1,1}$ with Neumann boundary conditions, drawn on the surface of the unit cube

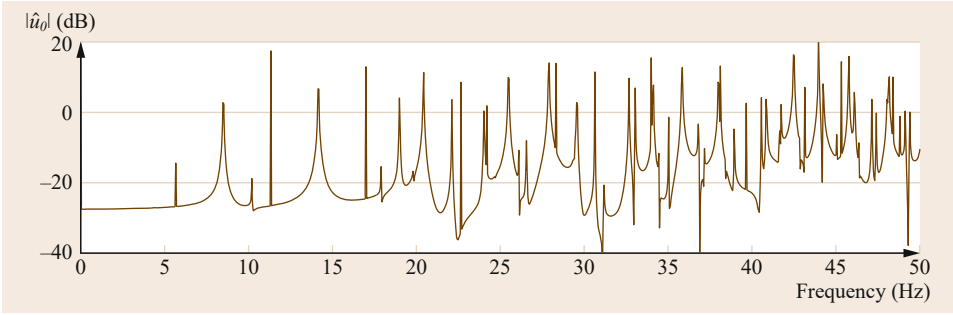


Fig. 19.20 Spectrum of response of a 10 m × 20 m × 30 m room to an impulsive initial condition with $c = 340$ m/s and Neumann boundaries. The output was read at a location near the center of the domain

of the ideal bar where mode density decreases with frequency.

19.4.2 A Grid and Difference Operators

As in the 1-D case, a grid function represents an approximation to the solution of interest – in this case it will be defined over a regular grid. In 2-D and 3-D there are various choices of regular spatial grids (lattices) [19.43], but the simplest and most straightforward choices are the integer lattices $\mathbb{Z}^2$ and $\mathbb{Z}^3$ scaled in each dimension by the spatial step h. The set of integer pairs or triples d represents the discrete analog of the continuous domain $\mathcal{D}$. For example, the rectangular region $\mathcal{D}_{L_x,L_y}$ could be represented by $d = d_{N_x,N_y} = \{0,\ldots,N_x\} \times \{0,\ldots,N_y\}$ where $N_x = L_x/h$ and $N_y = L_y/h$. Likewise, $\mathcal{D}_{L_x,L_y,L_z}$ could be represented by $d = d_{N_x,N_y,N_z} = \{0,\ldots,N_x\} \times \{0,\ldots,N_y\} \times \{0,\ldots,N_z\}$ with $N_z = L_z/h$. For simplicity it is assumed that L_x, L_y, L_z are common multiples of some h. These domains are illustrated in Fig. 19.21.

A grid function in 2-D can then be written as $u^n_{l_x,l_y}$ representing the approximation to $u(\boldsymbol{x},t)$ at $t = nk$, where k is the time step and $n \in \mathbb{Z}^+$, and at $\boldsymbol{x} = (l_x h, l_y h)$ and $(l_x, l_y) \in d \subseteq \mathbb{Z}^2$. Likewise, a grid function in 3-D may be written as $u^n_{l_x,l_y,l_z}$ with $\boldsymbol{x} = (l_x h, l_y h, l_z h)$ and $(l_x, l_y, l_z) \in d \subseteq \mathbb{Z}^3$.

The temporal difference operators δ_{tt} defined previously will be employed here, as will δ_{xx}, which acts on the 2-D and 3-D grid functions in the following manner

$$\delta_{xx} u^n_{l_x,l_y} = \frac{1}{h^2}\left(u^n_{l_x+1,l_y} - 2u^n_{l_x,l_y} + u^n_{l_x-1,l_y}\right)$$
$$\delta_{xx} u^n_{l_x,l_y,l_z} = \frac{1}{h^2}\left(u^n_{l_x+1,l_y,l_z} - 2u^n_{l_x,l_y,l_z} + u^n_{l_x-1,l_y,l_z}\right) . \quad (19.72)$$

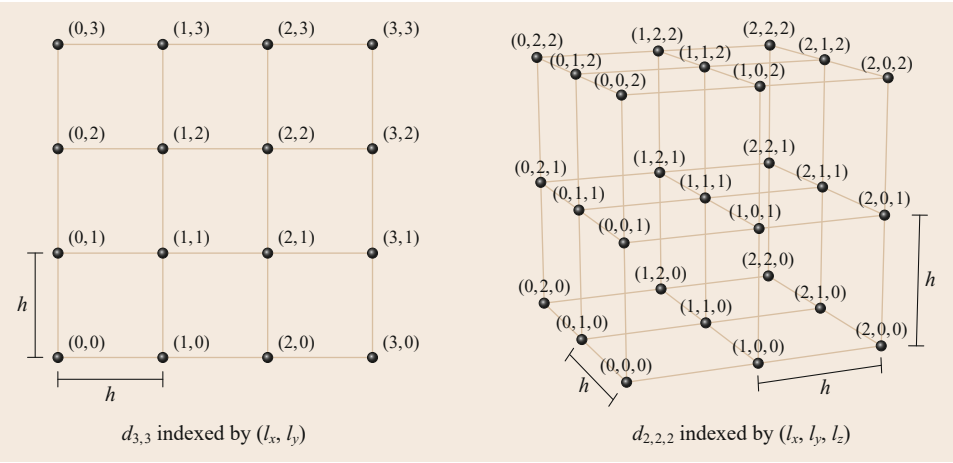


Fig. 19.21 Spatial grids in 2-D and 3-D

Similarly, the operator δ_{yy} in 2-D and 3-D may be defined as

$$\delta_{yy} u^n_{l_x,l_y} = \frac{1}{h^2}\left(u^n_{l_x,l_y+1} - 2u^n_{l_x,l_y} + u^n_{l_x,l_y-1}\right)$$
$$\delta_{yy} u^n_{l_x,l_y,l_z} = \frac{1}{h^2}\left(u^n_{l_x,l_y+1,l_z} - 2u^n_{l_x,l_y,l_z} + u^n_{l_x,l_y-1,l_z}\right) \tag{19.73}$$

and additionally in 3-D, δ_{zz}

$$\delta_{zz} u^n_{l_x,l_y,l_z} = \frac{1}{h^2}\left(u^n_{l_x,l_y,l_z+1} - 2u^n_{l_x,l_y,l_z} + u^n_{l_x,l_y,l_z-1}\right) . \tag{19.74}$$

At boundary locations, these spatial operators will sometimes address a tuple (l_x, l_y) or (l_x, l_y, l_z) that is not in d. Such a point is sometimes referred to as a *ghost point* and the grid function at this point must be set according to the appropriate boundary conditions.

19.4.3 A Simple Finite-Difference Scheme

The simplest finite-difference scheme for the 2-D wave equation is the classic scheme dating back to 1928 [19.1]. It is expressed as

$$\delta_{tt} u^n_{l_x,l_y} = c^2 \delta^{(2)}_{\Delta} u^n_{l_x,l_y} , \tag{19.75}$$

where $\delta^{(2)}_{\Delta} = \delta_{xx} + \delta_{yy}$ is a discrete Laplacian, which operates over a stencil of points. The stencil makes use of five spatial points, so it is called a five-point stencil. The two-step update equation for this scheme, obtained by expanding out the operators in full, is

$$u^{n+1}_{l_x,l_y} = 2u^n_{l_x,l_y} - u^{n-1}_{l_x,l_y} + \lambda^2\left(u^n_{l_x+1,l_y} + u^n_{l_x-1,l_y} + u^n_{l_x,l_y+1} + u^n_{l_x,l_y-1} - 4u^n_{l_x,l_y}\right) , \tag{19.76}$$

where $\lambda = ck/h$ is the Courant number. This update is illustrated in Fig. 19.22.

A straightforward extension of (19.75) gives the simplest scheme for the 3-D wave equation [19.3]

$$\delta_{tt} u^n_{l_x,l_y,l_z} = c^2 \delta^{(3)}_{\Delta} u^n_{l_x,l_y,l_z} , \tag{19.77}$$

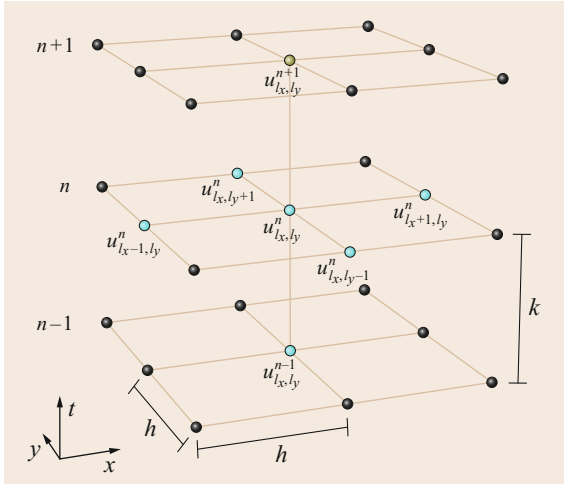


Fig. 19.22 Illustration of 2-D finite-difference update on a space-time grid. The update for the scheme, at the grid point marked in *yellow*, depends on previously calculated values at the grid points colored *cyan*

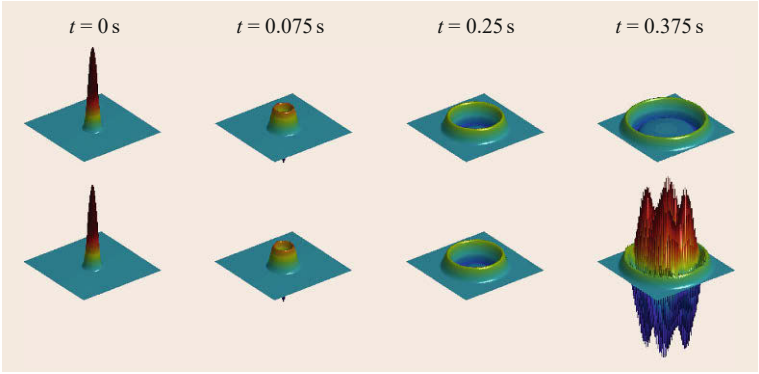


Fig. 19.23 Time evolution of $u^n_{l_x,l_y}$ on the unit square, under a raised cosine initial condition with $r_{\text{hw}} = 0.1\,\text{m}$, $\boldsymbol{x}_0 = (0.5, 0.5)$, $h = 0.01\,\text{m}$, and $c = 1\,\text{m/s}$, at times as indicated. *Top*: with $\lambda = \lambda_{\max}$. *Bottom*: with $\lambda = 1.0077\lambda_{\max}$. Compare with the solution in Fig. 19.14

where now $\delta^{(3)}_{\Delta} = \delta_{xx} + \delta_{yy} + \delta_{zz}$, which operates over a seven-point stencil. The update equation becomes

$$\begin{aligned} u^{n+1}_{l_x,l_y,l_z} &= 2u^n_{l_x,l_y,l_z} - u^{n-1}_{l_x,l_y,l_z} \\ &\quad + \lambda^2 \left(u^n_{l_x+1,l_y,l_z} + u^n_{l_x-1,l_y,l_z} \right. \\ &\quad + u^n_{l_x,l_y+1,l_z} + u^n_{l_x,l_y-1,l_z} + u^n_{l_x,l_y,l_z+1} \\ &\quad \left. + u^n_{l_x,l_y,l_z-1} - 6u^n_{l_x,l_y,l_z} \right) \end{aligned} \tag{19.78}$$

with $\lambda = ck/h$. Another formulation, using a first-order system for (19.60) as a starting point, adapted from the classic *Yee scheme* for electromagnetics [19.4] on staggered grids in space and time [19.44] is also possible. It can be shown that the staggered formulation is equivalent to (19.77) [19.45].

von Neumann Analysis and Stability

Stability analysis for the 2-D scheme over the infinite domain $d = \mathbb{Z}^2$ may be approached using von Neumann analysis, as in the 1-D case. In 2-D, a wavelike test solution (or *ansatz*) is

$$u^n_{l_x,l_y} = z^n \mathrm{e}^{j(\beta_x l_x h + \beta_y l_y h)} \quad z = \mathrm{e}^{sk} \tag{19.79}$$

for real wave vectors $\boldsymbol{\beta} = (\beta_x, \beta_y)$ and complex frequency $s = \sigma + j\omega$. Inserting this solution into (19.76) leads to the following characteristic equation in z

$$z + 4\lambda^2(s_x + s_y) - 2 + z^{-1} = 0\,, \tag{19.80}$$

where $s_x = \sin^2(\beta_x h/2)$ and $s_y = \sin^2(\beta_y h/2)$. The condition $|z| \le 1$ for all real $\boldsymbol{\beta}$ leads to stability of the scheme, and this will be true when

$$\lambda \le \lambda_{\max} = \sqrt{1/2} \quad \Rightarrow \quad h \ge h_{\min} = \sqrt{2}ck\,. \tag{19.81}$$

Numerical instability will arise if these conditions are not satisfied. Figure 19.23 demonstrates the effect of such instabilities – explosive (exponential) growth is quickly seen with λ only slightly larger than $\lambda_{\max}$.

It is straightforward to extend such von Neumann analysis to 3-D. Although omitted for brevity, this analysis leads to the following condition for the 3-D scheme

$$\lambda \le \lambda_{\max} = \sqrt{1/3} \Rightarrow h \ge h_{\min} = \sqrt{3}ck\,. \tag{19.82}$$

Numerical Dispersion and Mode Detuning

As in the 1-D case, numerical dispersion gives rise to a numerical wave speed that deviates from that of the wave equation itself. Inserting a plane wave solution $u = \mathrm{e}^{j(\omega nk + \beta_x l_x h + \beta_y l_y h)}$ into (19.75) gives the numerical dispersion relation for the 2-D scheme

$$\begin{aligned} \sin^2(\omega k/2) &= \lambda^2(s_x + s_y) \\ \Rightarrow \omega &= \pm\frac{2}{k}\sin^{-1}\left(\lambda\sqrt{s_x + s_y}\right). \end{aligned} \tag{19.83}$$

This relates temporal frequencies ω to spatial wave vectors $\boldsymbol{\beta}$ in the finite-difference scheme. The numerical dispersion relation is again dependent on the free parameter λ. As in the 1-D case, fixing the Courant number to $\lambda = \lambda_{\max}$ is generally a good idea, since this will maximize the temporal bandwidth that is produced by the scheme (this follows from (19.83)). This effect is also illustrated in Fig. 19.24, where a square domain is simulated using the 2-D scheme with three different choices of the Courant number. It can be seen that as the Courant number is reduced the scheme reproduces less temporal bandwidth.

It is also worth examining the numerical phase velocity ($v_{\text{phase}} = \omega/|\boldsymbol{\beta}|$) of the scheme in order to see how it deviates from the ideal c. The numerical phase velocity of the scheme is displayed in Fig. 19.25 for three choices of the Courant number. Note that, due to

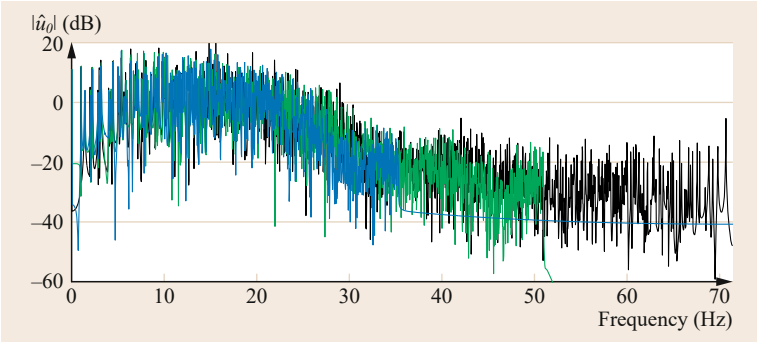


Fig. 19.24 Spectra of outputs from an impulsive initial condition (raised cosine) with Neumann boundaries on the domain $\mathcal{D}_{1,1}$ with $c = 1\,\mathrm{m/s}$ and $h = 0.01\,\mathrm{m}$. The outputs were taken at the center of the domain. The Courant number is set as $\lambda = \lambda_{\max}$ (*black*), $\lambda = 0.9\lambda_{\max}$ (*green*), and $\lambda = 0.7\lambda_{\max}$ (*blue*)

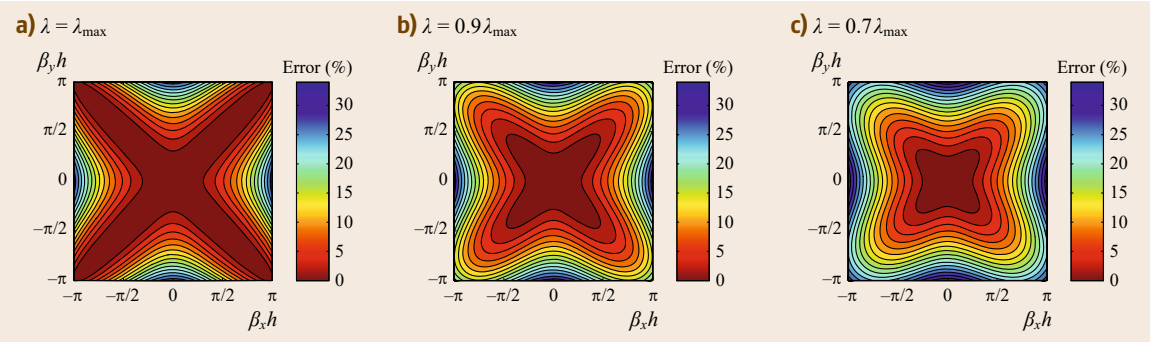


Fig. 19.25a–c Numerical phase velocity v_{phase} of 2-D scheme (19.75) as a function of spatial frequency with 2% contours, for various Courant numbers. **(a)** $\lambda = \lambda_{\max}$, **(b)** $\lambda = 0.9\lambda_{\max}$, **(c)** $\lambda = 0.7\lambda_{\max}$

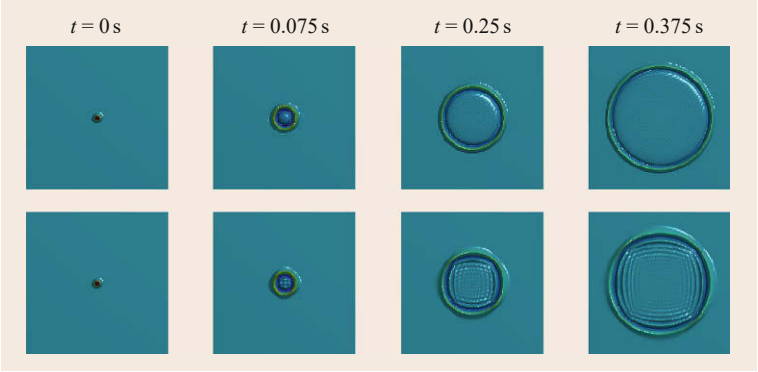


Fig. 19.26 Time evolution of a 2-D acoustic field viewed from above under a raised cosine initial condition with $r_{\text{hw}} = 0.025\,\mathrm{m}$, $\boldsymbol{x}_0 = (0.5, 0.5)$, and $c = 1\,\mathrm{m/s}$. *Top*: exact solution. *Bottom*: finite-difference approximation with $\lambda = \lambda_{\max}$ and $h = 0.01\,\mathrm{m}$

the discrete nature of the spatial grid, the spatial frequencies to consider are within a square cell centered about the origin with sides of length $2\pi/h$. In all three cases featured in Fig. 19.25, v_{phase} is correct at the origin ($|\boldsymbol{\beta}| = 0$, or the DC wavenumber), which follows from the consistency of the numerical scheme with the model equation. Beyond DC ($|\boldsymbol{\beta}| > 0$), v_{phase} exhibits increasing error with wavenumber along any given direction. Furthermore, the numerical phase velocity varies with the angle of propagation for a fixed wavenumber. This phenomenon is known as *anisotropy* of the scheme, as opposed to the ideal dispersion relation, which is isotropic, or independent of direction.

The effect of numerical dispersion on an approximated solution can be seen in Fig. 19.26. The exact solution propagates as a circular wavefront, whereas the approximated solution has ripples that (in this case) trail its circular wavefront due to numerical dispersion.

Similar effects are found in an analysis of the numerical dispersion of the 3-D scheme, which will be left out for the sake of brevity. It should be pointed out, however, that unwanted effects of dispersion are

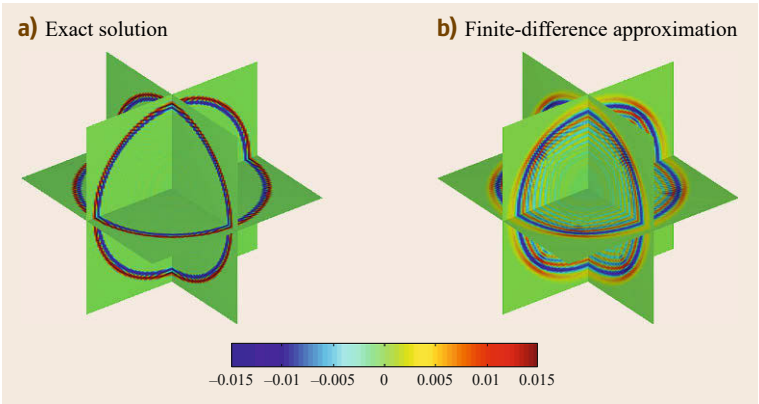


Fig. 19.27a,b Snapshot in time of 3-D acoustic field on $\mathcal{D}_{1,1,1}$ with $c = 1\,\mathrm{m/s}$, under a raised cosine initial condition with $r_{\mathrm{hw}} = 0.025$, $\boldsymbol{x}_0 = (0.5, 0.5, 0.5)$. The finite-difference approximation uses $h = 0.01\,\mathrm{m}$ and $\lambda = \sqrt{1/3}$. The snapshot shows $u^n_{l_x,l_y,l_z}$ after 75 time steps, at $t = 0.43\,\mathrm{s}$. **(a)** Exact solution, **(b)** finite-difference approximation

stronger than in the 2-D case, ultimately due to $\lambda_{\max}$ being smaller in the 3-D case. These effects are illustrated in Fig. 19.27, which compares a snapshot of an exact solution in 3-D to the approximation from (19.77), with $\lambda = \lambda_{\max}$. Again, the effect of numerical dispersion is exhibited as a series of ripples in the wake of the spreading wavefront.

Numerical Boundary Conditions

In order to terminate the grid one must impose numerical boundary conditions. In this section, the scheme (19.75) for the 2-D wave equation is assumed defined over the discrete rectangular region d_{N_x,N_y}.

The simplest such conditions are those of Dirichlet type (19.64), which do not require access to values at any ghost points. Over d_{N_x,N_y} these can be set along the boundary corresponding to $x = 0$ by

$$u^n_{0,l_y} = 0 \quad \text{for} \quad l_y = 0, \ldots, N_y \tag{19.84}$$

and similarly along the boundaries corresponding to $x = L_x$, $y = 0$, and $y = L_y$.

Neumann conditions (19.64) can be implemented in a similar fashion to the 1-D case. Here, along the boundary corresponding to $x = 0$, numerical Neumann conditions can be of the *centered* type or *non-centered* type

$$\begin{aligned} u^n_{-1,l_y} &= u^n_{0,l_y} \quad \text{(non-centered)} \\ u^n_{-1,l_y} &= u^n_{1,l_y} \quad \text{(centered)}\,, \end{aligned} \tag{19.85}$$

where $l_y = 1, \ldots, N_y - 1$ and u^n_{-1,l_y} are the ghost points. Substituting these conditions back into the 2-D update results in the following specialized updates for boundary nodes

$$\begin{aligned} u^{n+1}_{0,l_y} &= 2u^n_{0,l_y} - u^{n-1}_{0,l_y} \\ &\quad + \lambda^2 \left(u^n_{1,l_y} + u^n_{0,l_y+1} + u^n_{0,l_y-1} - 3u^n_{0,l_y} \right) \\ &\quad \text{(non-centered)} \end{aligned} \tag{19.86}$$

$$\begin{aligned} u^{n+1}_{0,l_y} &= 2u^n_{0,l_y} - u^{n-1}_{0,l_y} \\ &\quad + \lambda^2 \left(2u^n_{1,l_y} + u^n_{0,l_y+1} + u^n_{0,l_y-1} - 4u^n_{0,l_y} \right) \\ &\quad \text{(centered)}\,, \end{aligned} \tag{19.87}$$

where $l_y = 1, \ldots, N_y - 1$. For Neumann conditions along adjacent edges of the rectangular domain, corner nodes, such as at $l_x = l_y = 0$, require another specialized update with Neumann conditions; considering, additionally, a Neumann condition along the boundary corresponding to $y = 0$

$$\begin{aligned} u^n_{l_x,-1} &= u^n_{l_x,0} \quad \text{(non-centered)} \\ u^n_{l_x,-1} &= u^n_{l_x,1} \quad \text{(centered)}\,. \end{aligned} \tag{19.88}$$

The update at the corner node $l_x = l_y = 0$ would then be

$$\begin{aligned} u^{n+1}_{0,0} &= 2u^n_{0,0} - u^{n-1}_{0,0} + \lambda^2 \left(u^n_{1,0} + u^n_{0,1} - 2u^n_{0,0} \right) \\ &\quad \text{(non-centered)} \end{aligned} \tag{19.89}$$

$$\begin{aligned} u^{n+1}_{0,0} &= 2u^n_{0,0} - u^{n-1}_{0,0} + \lambda^2 \left(2u^n_{1,0} + 2u^n_{0,1} - 4u^n_{0,0} \right) \\ &\quad \text{(centered)} \end{aligned} \tag{19.90}$$

and similar updates may be worked out by symmetry for the other three corners of the domain d_{N_x,N_y}, as well as extensions to 3-D on d_{N_x,N_y,N_z}.

As in the 1-D case, these boundary conditions lead to stable numerical solutions. For irregular geometries, energy techniques should be employed in order to guarantee numerical stability. Finite volume techniques become very useful in this regard, but must be left out for the sake of brevity – see, e.g., [19.46, 47].

Vector–Matrix Representation

The 2-D and 3-D schemes, which are linear, may be compactly written in terms of matrix updates, as in the 1-D case – but, for such a representation, the multidimensional grid must first be represented as a vector. For

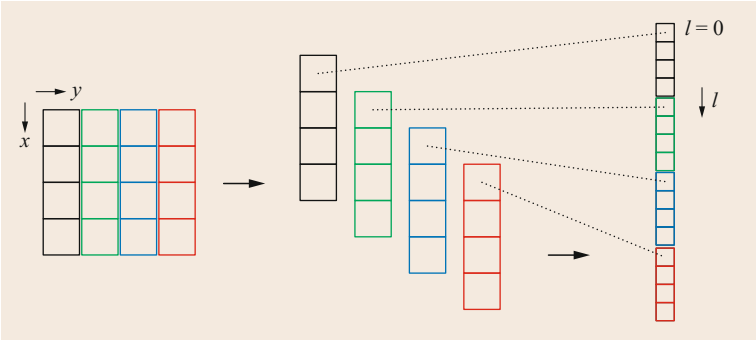


Fig. 19.28 Decomposition of $u^n_{l_x,l_y}$ on $d_{3,3}$ with Neumann conditions into column vector $\boldsymbol{u}^n$

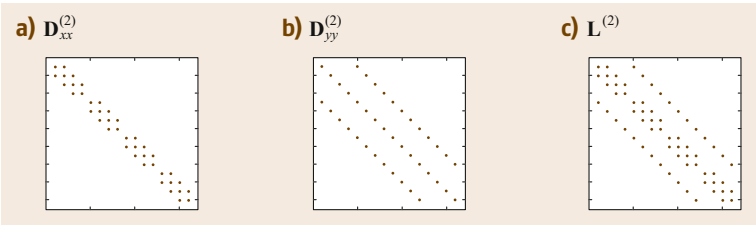


Fig. 19.29a–c Sparsity patterns for matrix operators with Neumann conditions on $d_{3,3}$. *Dots* represent nonzero elements of the matrix. **(a)** $\mathbf{D}^{(2)}_{xx}$, **(b)** $\mathbf{D}^{(2)}_{yy}$, **(c)** $\mathbf{L}^{(2)}$

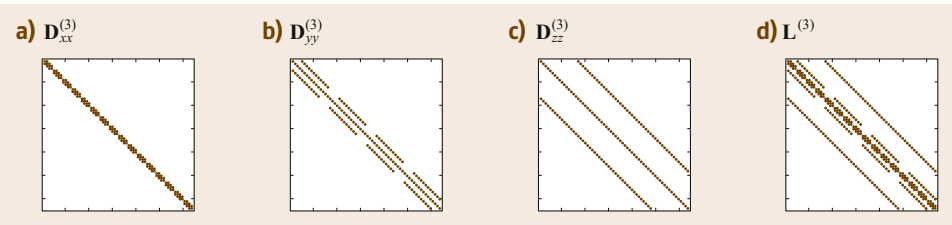


Fig. 19.30a–d Sparsity patterns for matrix operators on $d_{3,3,3}$ with Neumann conditions. **(a)** $\mathbf{D}^{(3)}_{xx}$, **(b)** $\mathbf{D}^{(3)}_{yy}$, **(c)** $\mathbf{D}^{(3)}_{zz}$, **(d)** $\mathbf{L}^{(3)}$

the 2-D grid function $u^n_{l_x,l_y}$, for example, one approach is to decompose the grid into vertical strips and then concatenate them into one column vector $\boldsymbol{u}^n$. In the case of Neumann conditions on $d_{N,N}$, the lth element $\boldsymbol{u}^n_l$ may be defined in terms of the 2-D grid function $u^n_{l_x,l_y}$, as

$$\boldsymbol{u}^n_l = u^n_{l_x,l_y} \quad \text{where} \quad l = l_x + l_y(N+1)\,. \tag{19.91}$$

This vector decomposition is illustrated in Fig. 19.28. Similar decompositions are possible for the 3-D case.

The matrix operators $\mathbf{D}^{(2)}_{xx}$, $\mathbf{D}^{(2)}_{yy}$ for the 2-D scheme can be constructed using the matrix $\mathbf{D}_{xx}$ given in the 1-D case in (19.28). On a square domain $d_{N,N}$, these matrices can be written as

$$\begin{aligned}\mathbf{D}^{(2)}_{xx} &= \mathbf{D}_{xx} \otimes \mathbf{I} \\ \mathbf{D}^{(2)}_{yy} &= \mathbf{I} \otimes \mathbf{D}_{yy}\,,\end{aligned} \tag{19.92}$$

where $\otimes$ denotes the Kronecker product, and where $\mathbf{I}$ is the identity matrix of appropriate size. The Laplacian matrix $\mathbf{L}^{(2)}$, representing the operator $\delta^{(2)}_{\Delta}$ and including boundary conditions on $d_{N,N}$, then becomes

$$\mathbf{L}^{(2)} = \mathbf{D}^{(2)}_{xx} + \mathbf{D}^{(2)}_{yy}\,. \tag{19.93}$$

The sparsity patterns of these matrices are shown in Fig. 19.29.

The update for the 2-D scheme can then be written in the by-now familiar compact form

$$\begin{aligned}\boldsymbol{u}^{n+1} &= \mathbf{B}\boldsymbol{u}^n - \boldsymbol{u}^{n-1} \\ \mathbf{B} &= 2\mathbf{I}^{(2)} + c^2k^2\mathbf{L}^{(2)}\,,\end{aligned} \tag{19.94}$$

where $\mathbf{I}^{(2)} = \mathbf{I} \otimes \mathbf{I}$. Matrices representing the operators δ_{xx}, δ_{yy}, δ_{zz}, $\delta^{(3)}_{\Delta}$ for the 3-D case can be constructed in a similar fashion, leading to the matrices $\mathbf{D}^{(3)}_{xx}$, $\mathbf{D}^{(3)}_{yy}$, $\mathbf{D}^{(3)}_{zz}$, $\mathbf{L}^{(3)}$ with sparsity patterns as shown in Fig. 19.30.

19.4.4 A Family of Implicit Finite-Difference Schemes

Recalling the case of the ideal bar, it may be advantageous to employ an implicit scheme as opposed to an explicit scheme – the point of introducing an implicit character to the scheme is to attempt to reduce unwanted effects of numerical dispersion.

A family of implicit schemes for the 2-D wave equation can be written as

$$\left(1+\alpha h^2 \delta^{(2)}_{\Delta,\gamma}\right) \delta_{tt} u^n_{l_x,l_y} = c^2 \delta^{(2)}_{\Delta,\gamma} u^n_{l_x,l_y} \,, \qquad (19.95)$$

where $\alpha \in \mathbb{R}$ is a free parameter. When $\alpha = 0$ the scheme is explicit, and otherwise implicit. It also helps to use a wider stencil of points, so here $\delta^{(2)}_{\Delta,\gamma}$ operates over a nine-point stencil

$$\delta^{(2)}_{\Delta,\gamma} = \delta_{xx} + \delta_{yy} + \frac{h^2}{2}(1-\gamma)\delta_{xx}\delta_{yy} \,, \qquad (19.96)$$

which has a free parameter $\gamma \in \mathbb{R}$, and which reduces to a five-point Laplacian when $\gamma = 0$ or $\gamma = 1$. It is common to choose the parameter $\gamma = 2/3$ [19.48], which reduces the directional dependence of the approximation error.

Analogous to the update shown in Fig. 19.22, the update for a given node $u^{n+1}_{l_x,l_y}$ is illustrated in Fig. 19.31. Though the analysis is more involved, stability conditions may be worked out in the same manner as for the explicit scheme. von Neumann analysis leads to the following conditions relating γ, λ and α

$$\gamma \geq 0 \quad \lambda \leq \lambda_{\max} = \begin{cases} \sqrt{1-4\alpha} & \text{if } \gamma \leq 1/2 \\ \sqrt{\frac{1}{2\gamma} - 4\alpha} & \text{if } \gamma > 1/2 \end{cases} \,,$$
$$\alpha < \begin{cases} \frac{1}{4} & \text{if } \gamma \leq 1/2 \\ \frac{1}{8\gamma} & \text{if } \gamma > 1/2 \,. \end{cases} \qquad (19.97)$$

The effect of choosing $\gamma = 2/3$ is seen in the numerical dispersion of the explicit scheme ($\alpha = 0$) in Fig. 19.32a. It can be seen that the numerical dispersion is more isotropic than that of the standard scheme as shown in Fig. 19.25.

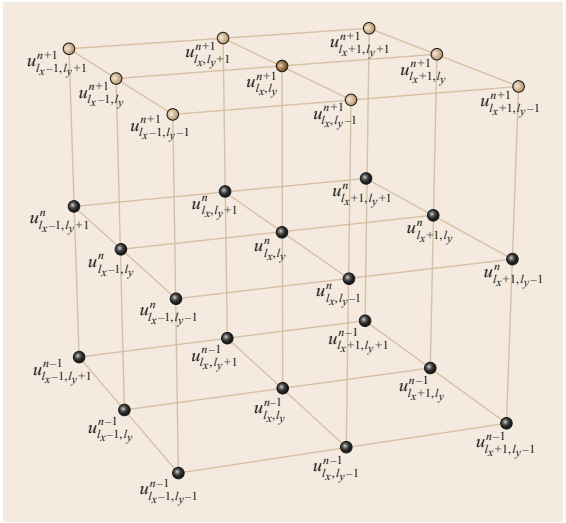


Fig. 19.31 Illustration of the implicit 2-D finite-difference update on a space-time grid. The update for the *highlighted point at the center* (in dark brown) depends on the unknown values of its neighbors (in light brown), as well as the known values (in gray)

For $\gamma = 2/3$, a good choice for an implicit scheme is $\alpha = 1/32$, which results in $\lambda_{\max} = \sqrt{5/8}$. It can be shown that this is in fact a fourth-order accurate scheme, which means that the approximation error is proportional to h^4 (implying, for a fixed $\lambda \leq \lambda_{\max}$, that it is also proportional to k^4), as opposed to the other schemes presented so far, which have an approximation error proportional to h^2 (and are thus second-order accurate). In other words, as the spatial step h is reduced, the approximation error in a fourth-order scheme decreases at a faster rate than in a second-order scheme. In terms of numerical dispersion, the effect of higher-order accuracy is seen mostly near the DC wavenumber, but the improvement in v_{phase} also extends to higher

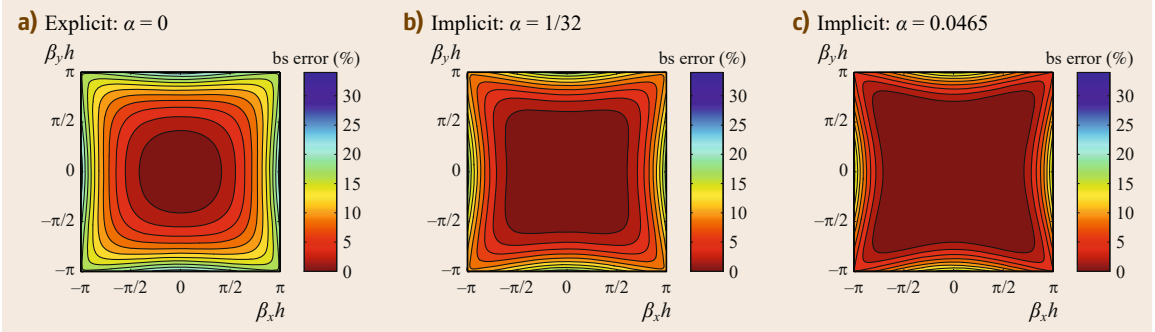


Fig. 19.32a–c Error in numerical phase velocity v_{phase} of 2-D scheme (19.95) with $\gamma = 2/3$ as a function of spatial frequencies with 2% contours. The Courant number is set as $\lambda = \lambda_{\max}$ in each case. **(a)** Explicit $\alpha = 0$, **(b)** implicit $\alpha = 1/32$, **(c)** implicit $\alpha = 0.0465$

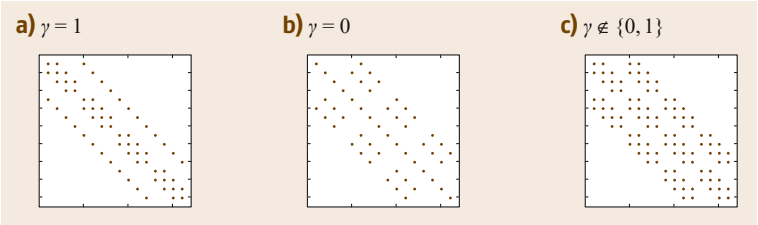


Fig. 19.33a–c Sparsity patterns of $\mathbf{L}_\gamma^{(2)}$ with centered Neumann conditions on $d_{3,3}$, for various γ. **(a)** $\gamma = 1$, **(b)** $\gamma = 0$, **(c)** $\gamma \notin \{0, 1\}$

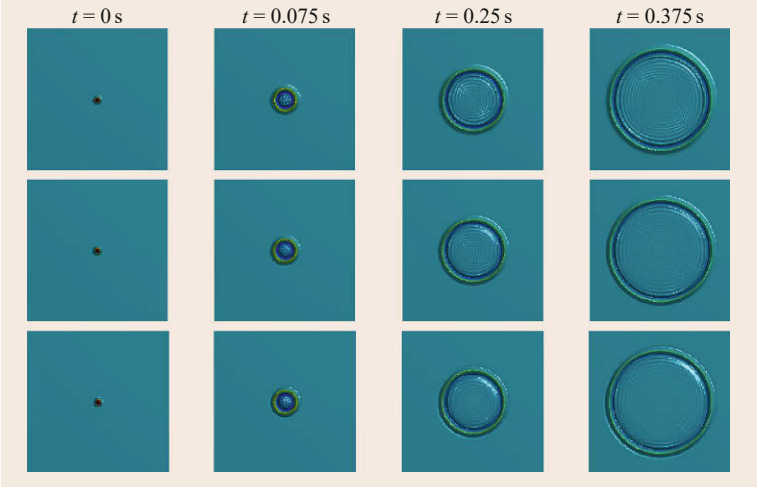


Fig. 19.34 Time evolution of 2-D approximations using a raised cosine initial condition with $r_{\text{hw}} = 0.025$ m, $\boldsymbol{x}_0 = (0.5, 0.5)$, and $c = 1$ m/s. For all schemes, $\gamma = 2/3$, $k = 7$ ms and $\lambda = \lambda_{\max}$. *Top*: $\alpha = 0$, *middle*: $\alpha = 1/32$, *bottom*: $\alpha = 0.0465$. Compare with the exact solution and approximation in Fig. 19.26

wavenumbers, as seen in Fig. 19.32b. It is also possible to choose α such that any improvement in v_{phase} is spread over a wider range of spatial frequencies. A good choice is $\alpha = 0.0465$, which provides less than 2% error in v_{phase} over nearly the entire range of discrete spatial frequencies, as seen in Fig. 19.32c.

As the update for the implicit scheme involves a linear system of equations, it is appropriate to write it in a vector-matrix form. The update on $d_{N,N}$ using centered Neumann conditions is then

$$\mathbf{A}\boldsymbol{u}^{n+1} = \mathbf{B}\boldsymbol{u}^{n} - \mathbf{A}\boldsymbol{u}^{n-1} , \tag{19.98}$$

where

$$\mathbf{A} = \mathbf{I}^{(2)} + \alpha h^2 \mathbf{L}_\gamma^{(2)} \tag{19.99}$$

$$\mathbf{B} = 2A + (c^2k^2 + 2\alpha h^2)\mathbf{L}_\gamma^{(2)} \tag{19.100}$$

$$\mathbf{L}_\gamma^{(2)} = \mathbf{D}_{xx}^{(2)} + \mathbf{D}_{yy}^{(2)} + \frac{h^2}{2}(1-\gamma)\mathbf{D}_{xx}^{(2)}\mathbf{D}_{yy}^{(2)} . \tag{19.101}$$

The sparsity patterns of $\mathbf{L}_\gamma^{(2)}$ with centered Neumann conditions on $d_{3,3}$ are shown in Fig. 19.33.

As in the 1-D implicit scheme (19.45), the system may be solved using the Jacobi method or other iterative methods, leading to a trade-off between accuracy and computational cost. The example featured in Fig. 19.26 is computed now with the three special cases of (19.98). Snapshots in time are displayed in Fig. 19.34.

Similar to the 2-D case, in 3-D a family of implicit-explicit schemes may be defined as

$$\left(1 + \alpha h^2 \delta_{\Delta,\gamma}^{(3)}\right) \delta_{tt} u_{l_x,l_y,l_z}^n = c^2 \delta_{\Delta,\gamma}^{(3)} u_{l_x,l_y,l_z}^n , \tag{19.102}$$

where $\alpha \in \mathbb{R}$ is a free parameter. The discrete Laplacian $\delta_{\Delta,\gamma}^{(3)}$ operates over a 19-point stencil

$$\delta_{\Delta,\gamma}^{(3)} = \delta_{xx} + \delta_{yy} + \delta_{zz} + \frac{h^2}{4}(1-\gamma)(\delta_{xx}\delta_{yy} + \delta_{xx}\delta_{zz} + \delta_{yy}\delta_{zz}) , \tag{19.103}$$

which has a free parameter $\gamma \in \mathbb{R}$, and which reduces to a seven-point stencil when $\gamma = 1$ and a 13-point stencil when $\gamma = 0$. This stencil is illustrated in Fig. 19.35. An isotropic approximation error in the discrete Laplacian can be obtained with the choice $\gamma = 1/3$ [19.49]. Similar to the 2-D case, $\delta_{\Delta,\gamma}^{(3)}$ may be written as a matrix operator, leading to a sparsity pattern as shown

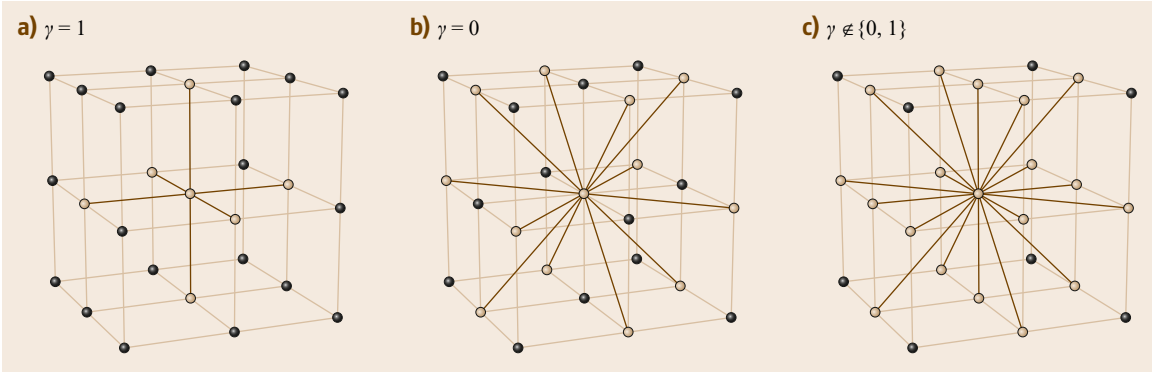


Fig. 19.35a–c Stencil of points for $\delta^{(3)}_{\Delta,\gamma}$. The spatial points used by $\delta^{(3)}_{\Delta,\gamma}$ are highlighted. **(a)** $\gamma = 1$, **(b)** $\gamma = 0$, **(c)** $\gamma \notin \{0, 1\}$

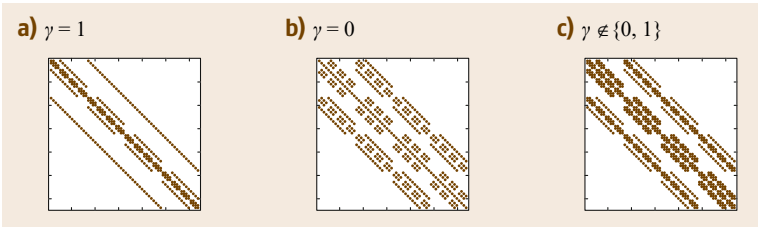


Fig. 19.36a–c Sparsity patterns of the matrix operator corresponding to $\delta^{(3)}_{\Delta,\gamma}$ with centered Neumann conditions on $d_{3,3,3}$, for various values of γ. **(a)** $\gamma = 1$, **(b)** $\gamma = 0$, **(c)** $\gamma \notin \{0, 1\}$

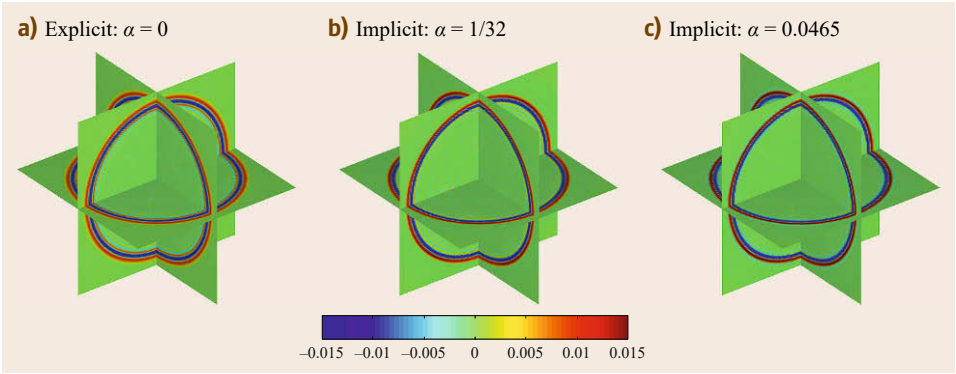


Fig. 19.37a–c Snapshot in time of 3-D approximation from (19.102) using $\gamma = 1/3$ and various values for α. In each case, $\lambda = \lambda_{\max}$ and $k = 0.057\,\mathrm{s}$. The approximation $u^n_{l_x,l_y,l_z}$ is shown after 75 time steps, at $t = 0.43\,\mathrm{s}$. Compare with Fig. 19.27. **(a)** Explicit $\alpha = 0$, **(b)** implicit $\alpha = 1/32$, **(c)** implicit $\alpha = 0.0465$

in Fig. 19.36. See [19.50] for the construction of this matrix.

Using a von Neumann analysis, the stability condition for this scheme is found to be

$$\gamma \geq 0\,,$$
$$\lambda \leq \lambda_{\max} = \begin{cases} \sqrt{\frac{1}{\gamma+1} - 4\alpha} & \text{if } \gamma \leq 1/2 \\ \sqrt{\frac{1}{3\gamma} - 4\alpha} & \text{if } \gamma > 1/2 \end{cases}$$
$$\alpha < \begin{cases} \frac{1}{4(\gamma+1)} & \text{if } \gamma \leq 1/2 \\ \frac{1}{12\gamma} & \text{if } \gamma > 1/2\,. \end{cases} \qquad (19.104)$$

As in the 2-D case, the value $\alpha = 1/32$ leads to a fourth-order accurate scheme, and the value $\alpha = 0.0465$ results in low dispersion error over a wide range of frequencies.

The example featured in Fig. 19.27 is now recalculated using the scheme (19.102) and $\alpha \in \{0, 1/32, 0.0465\}$ and illustrated in Fig. 19.37. An improvement over the simplest approximation in Fig. 19.27 can clearly be seen for all three cases, and the approximations obtained with the implicit schemes $\alpha \in \{1/32, 0.0465\}$ are, visually, close to the exact solution.

For much more on finite difference schemes for the 2-D and 3-D wave equations, see [19.51].

19.5 Thin Linear Plate Vibration

Stiff systems, such as the ideal bar described in detail in Sect. 19.2, require modeling techniques that are somewhat different from those used in the case of, for example, the wave equation. In 2-D, the direct analog of the bar is the plate, which is a thin, flat and stiff two-dimensional structure. When some curvature is present, as in the case of cymbals or gongs, such structures are usually referred to as shells, though shell structures will not be covered in this introductory tutorial. Plate vibration theory is a large research field with many applications in musical acoustics, and it is therefore useful to indicate some subcategories of interest.

While all plate structures are assumed to have a small thickness H relative to the lateral dimensions, a major distinction that can be made at the level of a model is whether the plate is considered to be thick or thin. The equation of motion for a thin plate was first proposed by *Kirchhoff* and *Love* [19.52] under various simplifying assumptions. *Mindlin* and *Reissner* [19.53, 54] subsequently proposed a more general (thick) model. In musical acoustics, thin models are often enough to describe the systems under consideration. These include, for example, gongs and cymbals [19.55], plate reverberation effects [19.56], violin or guitar bodies [19.57] and piano soundboards [19.58]; stiffness can also be included in models of drum membranes for a more realistic description of the dynamics [19.23].

It is the transverse vibrations of plates that are generally of interest, and they schematically fall into three different regimes. When the displacement u is small ($u \ll H$), the transverse motion can be uncoupled from the in-plane motion, and the resulting equations are linear. When $u \approx H$, in-plane and transverse vibrations become coupled and more complex, and nonlinear behavior arises [19.59]. This regime is well described by the von Kármán–Föppl equations [19.60]. For large deflections and rotations in the material, even this model loses validity. Such behavior, however, is generally never found in musical instruments under normal playing conditions.

Two additional properties of the material, namely spatial inhomogeneity and anisotropy, must also be taken into account, as they have major ramifications in terms of simulation. It is not uncommon, in fact, for gongs and cymbals to have a varying thickness, with a central region thicker than the rim. This is a typical example of spatial inhomogeneity. Wooden plates in string instruments or pianos on the other hand, even if spatially uniform, exhibit different behaviors depending on the direction, particularly regarding wave propagation speed, giving rise to anisotropic phenomena.

From a numerical simulation point of view, many possibilities are available. A widely adopted technique in engineering is the finite element method [19.61], which has proven to be very flexible and accurate for a diverse range of applications, including musical acoustics [19.62, 63]. Simulations of gongs and plate-like structures have also been attempted using digital waveguides [19.64]. Modal methods are an attractive option when the modes of vibration are known analytically. This technique has recently also been applied to imperfect plates and shells, i. e., surfaces with any type of defect [19.65], and nonlinear (von Kármán) plates [19.66]. Finite-difference approaches have also been adopted for the simulation of plates [19.67] and are easy to implement when simple geometries are involved. Examples include plate reverberation effects [19.68] and gong and cymbal simulations [19.69].

In this section, only the linear behavior of thin, spatially uniform and isotropic square plates will be considered in detail.

19.5.1 Equations of Motion

Consider a region $\mathcal{D} \subset \mathbb{R}^2$, and let $u(x, y, t)$ be the transverse displacement of the plate at position $(x, y) \in \mathcal{D}$ and time $t \in \mathbb{R}^+$. The equation of motion for a stiff plate is similar to that of the stiff bar equation presented in Sect. 19.2 and is usually referred to as the Kirchhoff model [19.70]

$$\partial_t^2 u = -\kappa^2 \Delta^2 u \,. \tag{19.105}$$

Here, Δ is the Laplacian operator introduced in Sect. 19.4, and $\Delta^2 = \Delta \cdot \Delta = \partial_x^4 + 2\partial_x^2\partial_y^2 + \partial_y^4$ is the biharmonic operator. The constant κ is defined as

$$\kappa = \sqrt{\frac{E H^2}{12\,\rho\,(1-\nu^2)}}\,, \tag{19.106}$$

where E is Young's modulus in Pa, H is the thickness in m, ρ is the density in kg/m^3 and ν is the dimensionless Poisson's ratio.

If the domain $\mathcal{D}$ is finite (such as, e.g., a square region of side length L, or $\mathcal{D} = \mathcal{D}_{L,L}$), it is necessary to complement equation (19.105) with suitable boundary conditions defined over the boundary of $\mathcal{D}$. Many possibilities are at hand, but only three are generally relevant in acoustics applications, namely clamped, simply supported and free conditions. They can be defined as

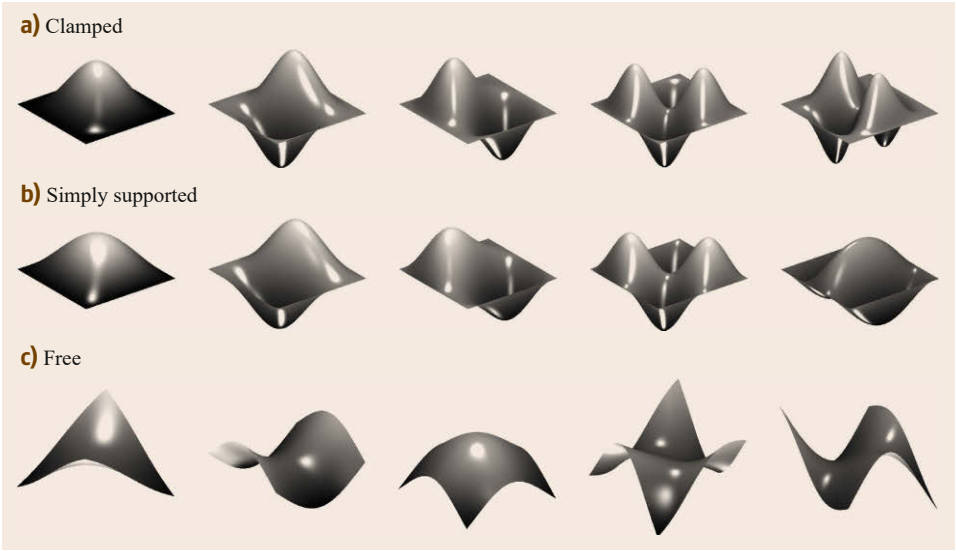


Fig. 19.38a–c First five modes in the modal series for three different boundary conditions: **(a)** clamped, **(b)** simply supported, and **(c)** free. In the free case, the first three modes corresponding to rigid body motion are not represented

follows, over a straight plate edge [19.71]

$$\begin{aligned} &u = \partial_{\boldsymbol{n}} u = 0 \quad \text{(clamped)} \\ &u = \partial_{\boldsymbol{n}}^2 u + \nu\, \partial_{\boldsymbol{s}}^2 u = 0 \quad \text{(simply supported)} \\ &\partial_{\boldsymbol{n}}^3 u + (2-\nu)\, \partial_{\boldsymbol{n}} \partial_{\boldsymbol{s}}^2 u = \partial_{\boldsymbol{n}}^2 u + \nu\, \partial_{\boldsymbol{s}}^2 u = 0 \quad \text{(free)}\,, \end{aligned} \tag{19.107}$$

where $\partial_{\boldsymbol{n}}$ denotes the directional derivative normal to the boundary oriented outwards, and ∂_s is the tangential derivative along the boundary. It is to be noted that two conditions need to be supplied, given that (19.105) is fourth-order in space, as in the case of the ideal bar.

The linearity of equation (19.105) allows a decomposition of the dynamics in terms of modal shapes. These have a simple, closed-form expression only in the simply supported case, in which case the plate equation is separable, and where each mode can be written as a product of sinusoidal functions, similarly to the case of the 2-D membrane. Figure 19.38 illustrates the lowest frequency modes for the three boundary conditions listed above. For the two fixed conditions (clamped and simply supported), the first modes look very similar, although a careful analysis reveals substantial differences at the boundaries, where different conditions on the derivative of the displacement are imposed. In the free case, three zero-frequency modes are also present, and they correspond to rigid body motions of the plate (two rotations and one rigid translation). These are not presented in the figure.

As in the case of the 1-D bar, the system is dispersive. This may be illustrated by inserting a test solution of the form

$$u = \mathrm{e}^{j(\omega t + \boldsymbol{\beta}\cdot\boldsymbol{x})}$$

into (19.105), where $\boldsymbol{\beta} = (\beta_x, \beta_y)$ is the wave vector and $\boldsymbol{x} = (x, y)$. One then obtains

$$\omega = \pm\kappa|\boldsymbol{\beta}|^2 \quad \Longrightarrow \quad v_{\text{phase}} = \kappa|\boldsymbol{\beta}|\,. \tag{19.108}$$

In other words, high-frequency waves travel faster than those at lower frequency. This can be seen in Fig. 19.39, where given an initial raised cosine impulse the disturbance gradually spreads, with the high-frequency components leading those of lower frequency.

It is now possible to derive a rough estimate of the number of modes up to a certain frequency. For simply supported conditions, in fact, the allowed frequencies can be written as [19.23]

$$f_{p_x,p_y} = \frac{\pi\kappa}{2L^2}\left(p_x^2 + p_y^2\right) \quad \text{with} \quad p_x, p_y = 1, \ldots\,. \tag{19.109}$$

Counting the number of modes N_{modes} as the area of the quarter of a circle of radius $|\boldsymbol{p}|$ gives an estimate of the modes below a certain frequency, f, in the limit as $f \to \infty$

$$N_{\text{modes}}(f) = \frac{fL^2}{2\kappa}\,. \tag{19.110}$$

Thus, the number of modes for the system scales linearly with the frequency, and the mode density is therefore constant. This can be seen in Fig. 19.40, where a typical spectrum for a clamped plate is shown. This feature of a very even response, lacking in early reflections, is at the heart of the characteristic sound of plate reverberation devices.

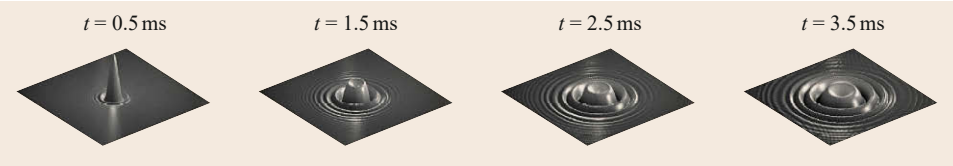


Fig. 19.39 Evolution of the displacement of a simply supported plate at times as indicated. The plate has stiffness coefficient $\kappa = 0.5\,\mathrm{m}^2/\mathrm{s}$ and $L = 1.2\,\mathrm{m}$. The initial condition is a raised cosine distribution at the center of the plate. The dispersive behavior of (19.105) is apparent

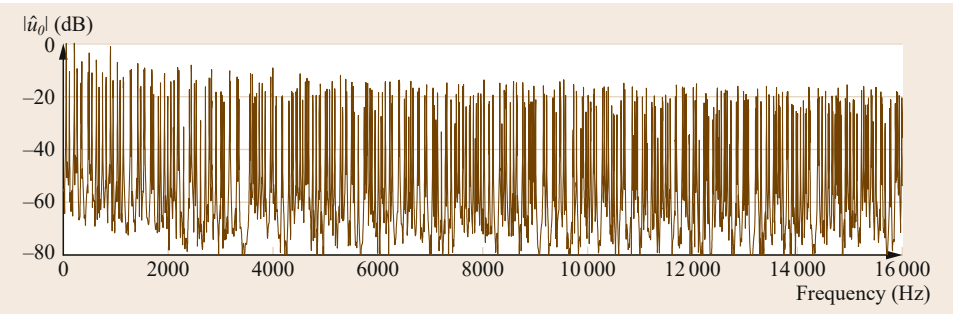


Fig. 19.40 Output spectrum for a square, thin plate with $\kappa = 10\,\mathrm{m}^2/\mathrm{s}$ and $L = 1\,\mathrm{m}$. An even distribution of modes at all frequencies is apparent

19.5.2 A Simple Finite-Difference Scheme

A simple finite-difference scheme for (19.105) can be obtained by introducing a grid function $u^n_{l_x,l_y}$ over a 2-D grid with grid spacing h (see Sect. 19.4.2) and by substituting continuous differential operators with discrete operators. Thus, one can write

$$\begin{aligned} \delta_{tt} u^n_{l_x,l_y} &= -\kappa^2 \delta_{\Delta,\Delta} u^n_{l_x,l_y}\,, \\ \delta_{\Delta,\Delta} &= \delta^{(2)}_{\Delta}\delta^{(2)}_{\Delta}\,, \end{aligned} \tag{19.111}$$

where the discrete biharmonic operator $\delta_{\Delta,\Delta}$ can be obtained by squaring the Laplacian operator $\delta^{(2)}_{\Delta}$ introduced in the previous section.

When expanded at an interior point, (19.111) takes the following form

$$\begin{aligned} u^{n+1}_{l_x,l_y} = {} & \left(2 - 20\mu^2\right) u^n_{l_x,l_y} \\ & + 8\mu^2 \left(u^n_{l_x,l_y+1} + u^n_{l_x,l_y-1} + u^n_{l_x+1,l_y} + u^n_{l_x-1,l_y}\right) \\ & - 2\mu^2 \left(u^n_{l_x+1,l_y+1} + u^n_{l_x-1,l_y+1} + u^n_{l_x+1,l_y-1} \right. \\ & \qquad \left. + u^n_{l_x-1,l_y-1}\right) \\ & - \mu^2 \left(u^n_{l_x,l_y+2} + u^n_{l_x,l_y-2} + u^n_{l_x+2,l_y} + u^n_{l_x-2,l_y}\right) \\ & - u^{n-1}_{l_x,l_y}\,, \end{aligned} \tag{19.112}$$

where $\mu = k\kappa/h^2$ is a dimensionless parameter, which will be restricted by stability conditions, as shown below, in a similar way to the case of the 2-D wave equation. The stencil for the spatial biharmonic operator $\delta_{\Delta,\Delta}$ is shown in Fig. 19.41a, together with the normalized weights of the various grid points, while the space-time update is shown in Fig. 19.41b.

It is possible to apply energy analysis techniques in order to obtain a numerical expression for the boundary conditions (19.107). The interested reader is referred to [19.17] for details. Regardless of the particular choice of boundary conditions, however, the scheme (19.111) can always be written in a vector-matrix form

$$\begin{aligned} \boldsymbol{u}^{n+1} &= \mathbf{B}\boldsymbol{u}^n - \boldsymbol{u}^{n-1}, \\ \mathbf{B} &= 2\mathbf{I} - \kappa^2 k^2 \mathbf{G}\,, \end{aligned} \tag{19.113}$$

where the biharmonic matrix $\mathbf{G}$ can be obtained by multiplying together two Laplacian matrices defined in Sect. 19.4.3. In the free case, particular attention must be paid to modifying the entries of the matrix $\mathbf{G}$ corresponding to the points near the boundaries.

Stability Conditions and Numerical Dispersion

Stability conditions for the scheme (19.111) can be easily obtained via von Neumann analysis techniques. The amplification polynomial can be found by inserting a test function

$$u^n_{l_x,l_y} = z^n\, \mathrm{e}^{j(\beta_x l_x h + \beta_y l_y h)}$$

into (19.111). This gives

$$z + 16\mu^2 (s_x + s_y)^2 - 2 + z^{-1} = 0\,, \tag{19.114}$$

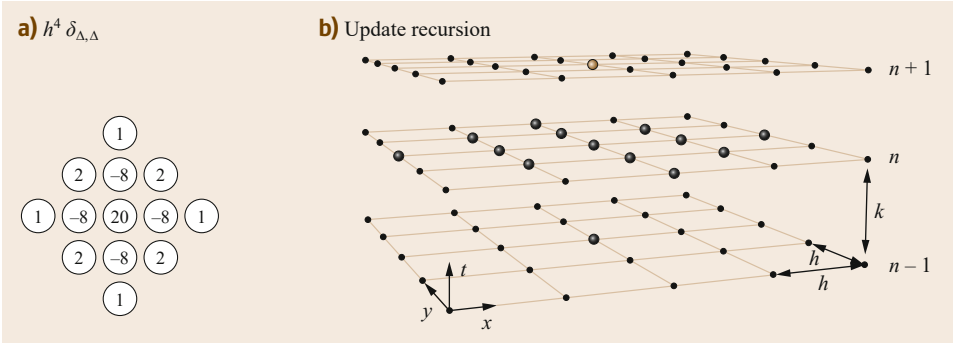


Fig. 19.41 **(a)** Stencil for the biharmonic operator $\delta_{\Delta,\Delta}$, scaled by a factor h^4. **(b)** Recursion stencil for the scheme (19.112). The update for the *highlighted point* at $n+1$ depends on the known values of the *gray* points at n and $n-1$

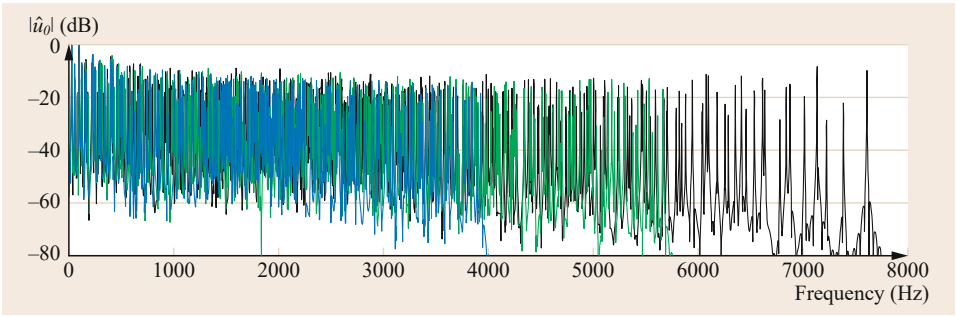


Fig. 19.42 Spectrum of the output given a raised cosine initial velocity at the center of a clamped, square plate with $L = 1\,\mathrm{m}$. The parameter μ was set as $\mu = \mu_{\max}$ (*black*), $\mu = 0.9\mu_{\max}$ (*green*) and $\mu = 0.7\mu_{\max}$ (*blue*). Also visible in the upper part of the spectrum is a decreasing density of modes due to numerical dispersion errors

where $z = \mathrm{e}^{(\sigma+j\omega)k}$, $s_x = \sin^2(\beta_x h/2)$ and $s_y = \sin^2(\beta_y h/2)$. In order for the solutions to be bounded at all times, the following conditions must be imposed

$$0 \le 4\mu^2(s_x + s_y) \le 1\,, \tag{19.115}$$

which ultimately lead to bounds on μ and h

$$\mu \le \mu_{\max} = \frac{1}{4} \Longrightarrow h \ge h_{\min} = 2\sqrt{\kappa k}\,. \tag{19.116}$$

Ideally, under stable conditions, one should recover the continuous relation (19.108) but, in practice, numerical errors always appear. The numerical dispersion relation, in fact, can be written as

$$\omega = \pm\frac{2}{k}\sin^{-1}\left[2\mu\left(s_x + s_y\right)\right]. \tag{19.117}$$

Note that, as in all the systems presented previously, the numerical dispersion relation depends on the parameter μ. Choosing values of μ far from the stability limit $\mu_{\max} = 1/4$ progressively reduces the bandwidth of the scheme. See Fig. 19.42 for an example.

Figure 19.43 shows the relative phase velocity (the ratio between numerical and theoretical values) for the Cartesian scheme. It is apparent that the phase velocity

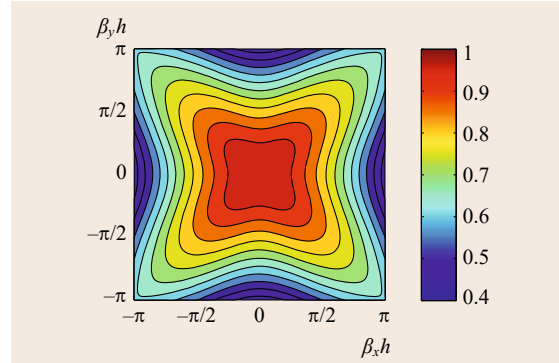


Fig. 19.43 Relative phase velocity for the numerical scheme (19.111) as a function of $\boldsymbol{\beta} \in [-\pi/h, \pi/h] \times [-\pi/h, \pi/h]$, with contours marking 5% deviations

differs from that of the model system even at relatively low frequencies. Note that this numerical dispersion is independent from the original dispersive behavior of the plate equation (19.105).

Another consideration concerns the isotropy of the numerical scheme. As can be noted from Fig. 19.43, contrary to the continuous case, the simple finite-difference scheme (19.111) exhibits a strong dependence on direction, with errors accumulating in partic-

ular along the grid directions. This is easily understood if one considers the anisotropy of the numerical Laplacian, which is now amplified by squaring the operator. One possible way to improve the behavior of the simulation is to use implicit schemes, which require the solution of a linear system, as has been carried out in the case of the wave equation scheme in Sect. 19.4. See [19.72, 73] for additional details.

19.6 Extensions to Nonlinear Systems

This chapter is intended as an introduction to finite-difference schemes for a variety of systems in musical acoustics. The systems covered here are basic, but very much representative – the construction techniques presented here can be applied to much wider classes of systems, incorporating many effects necessary for modeling real-world systems.

The systems examined here are all linear. In recent years, however, nonlinear effects in musical instruments have seen a great increase in interest in a variety of different settings, and time-stepping methods of a form similar to those presented here have been used. It should be obvious, though, that frequency-domain stability analysis along the lines of von Neumann must be ruled out in this case; much more powerful methods are based on the use of numerical energy conservation principles, which have not been discussed in this chapter. In virtually all cases, however, such construction techniques lead, in the nonlinear case, to implicit designs, and thus one can expect more computational work as a result.

One important family of nonlinearities is associated with geometrical nonlinearity in solid structures. Examples include string vibration under so-called tension modulation nonlinearity, allowing for the reproduction of effects such as pitch glides under high amplitude plucking excitation [19.74]. More detailed models of strings include longitudinal-transverse coupling, leading to the production of so-called phantom partials [19.75] in piano tones, which can also be approached using such methods [19.76, 77]. The simulation of nonlinear plate and shell vibration in 2-D can also be carried out using such methods [19.78], allowing for the rendering of effects such as crashes in gongs [19.79].

Another type of nonlinear behavior quite common in musical instrument acoustics involves collision, either of a lumped object with a distributed resonator (as in the case of hammer or mallet excitation) [19.11], or more involved collision between a distributed component with a barrier, as in the case of the sitar or tanpura [19.80] or a guitar string against a fretboard [19.81], and even the collision between two fully distributed components, as in the case of a wire vibrating against a membrane in the case of a snare drum. Finite-difference time-domain methods are also suited to such systems – a great variety of systems and related numerical schemes are described in [19.82]. The main distinguishing feature of numerical methods in this case is the reliance on iterative solvers such as the Newton–Raphson method.

Perhaps the most difficult nonlinearity to deal with, numerically, is that which is associated with high-amplitude playing in brass instruments, leading to increases in brightness due to shock wave formation [19.83]. Numerical methods associated with shock capturing have a long history in the mainstream numerical simulation community [19.84]; finite volume methods, which are closely related to the finite-difference schemes presented here are often used [19.85]. Numerical schemes of relevance in musical acoustics applications have been proposed recently [19.86].

References

19.1 R. Courant, K. Friedrichs, H. Lewy: Über die partiellen Differenzengleichungen der mathematischen Physik, Math. Ann. **100**(1), 32–74 (1928)

19.2 R.D. Richtmyer: *Difference Methods for Initial-Value Problems* (Interscience, New York 1957)

19.3 G.E. Forsythe, W.R. Wasow: *Finite-Difference Methods for Partial Differential Equations* (Wiley, New York 1960)

19.4 K.S. Yee: Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media, IEEE Trans. Antennas Propag. **14**(3), 302–307 (1966)

19.5 A. Taflove: Application of the finite-difference time-domain method to sinusoidal steady-state electromagnetic-penetration problems, IEEE Trans. Electromagn. Compat. EMC **22**(3), 191–202 (1980)

19.6 P. Ruiz: *A Technique for Simulating the Vibrations of Strings with a Digital Computer*, Master's Thesis (Univ. Illinois, Urbana 1969)

19.7 L. Hiller, P. Ruiz: Synthesizing musical sounds by solving the wave equation for vibrating objects: Part II, J. Audio Eng. Soc. **19**(7), 542–550 (1971)

19.8 J. Kelly, C. Lochbaum: Speech synthesis. In: *Proc. 4th Int. Congr. Acoust., Copenhagen* (1962) pp. 1–4
19.9 R. Bacon, J. Bowsher: A discrete model of a struck string, Acustica **41**, 21–27 (1978)
19.10 X. Boutillon: Model for piano hammers: Experimental determination and digital simulation, J. Acoust. Soc. Am. **83**(2), 746–754 (1988)
19.11 A. Chaigne, A. Askenfelt: Numerical simulations of struck strings. I. A physical model for a struck string using finite difference methods, J. Acoust. Soc. Am. **95**(2), 1112–1118 (1994)
19.12 S. van Duyne, J.O. Smith III: Physical modelling with the 2-D digital waveguide mesh. In: *Proc. Int. Comput. Music Conf., Tokyo* (1993) pp. 40–47
19.13 F. Fontana, D. Rocchesso: Physical modelling of membranes for percussion instruments, Acta Acust. united Acust. **84**(3), 529–542 (1998)
19.14 L. Savioja, T. Rinne, T. Takala: Simulation of room acoustics with a 3-D finite-difference mesh. In: *Proc. Int. Comput. Music Conf., Århus* (1994) pp. 463–466
19.15 S. van Duyne, J.O. Smith III: The 3-D tetrahedral digital waveguide mesh with musical applications. In: *Proc. Int. Comput. Music Conf., Hong Kong* (1996) pp. 9–16
19.16 D. Botteldooren: Acoustical finite-difference time-domain simulation in a quasi-Cartesian grid, J. Acoust. Soc. Am. **95**(5), 2313–2319 (1994)
19.17 S. Bilbao: *Numerical Sound Synthesis: Finite Difference Schemes and Simulation in Musical Acoustics* (Wiley, Chichester 2009)
19.18 P. Morse, U. Ingard: *Theoretical Acoustics* (Princeton University Press, Princeton 1968)
19.19 J.O. Smith III: *Physical Audio Signal Procesing* 2004)
19.20 J. Strikwerda: *Finite Difference Schemes and Partial Differential Equations* (Wadsworth Brooks, Pacific Grove 1989)
19.21 B. Gustafsson, H.-O. Kreiss, J. Oliger: *Time Dependent Problems and Difference Methods* (Wiley, New York 1995)
19.22 M. Ducceschi, S. Bilbao: Linear stiff string vibrations in musical acoustics: assessment and comparison of models, J. Acoust. Soc. Am. **140**(4), 2445 (2016)
19.23 N. Fletcher, T. Rossing: *The Physics of Musical Instruments* (Springer, New York 1998)
19.24 A.G. Webster: Acoustical impedance, and the theory of horns and of the phonograph, Proc. Natl. Acad. Sci. U.S.A. **5**(7), 275–282 (1919)
19.25 M. Campbell, C. Greated: *The Musician's Guide to Acoustics* (Oxford University Press, Oxford 1987)
19.26 A.H. Benade: On the propagation of sound waves in a cylindrical conduit, J. Acoust. Soc. Am. **44**(2), 616–623 (1968)
19.27 R. Caussé, J. Kergomard, X. Lurton: Input impedance of brass musical instruments – Comparison between experiment and numerical models, J. Acoust. Soc. Am. **75**(1), 241–254 (1984)
19.28 D.H. Keefe: Acoustical wave propagation in cylindrical ducts: Transmission line parameter approximations for isothermal and nonisothermal boundary conditions, J. Acoust. Soc. Am. **75**(1), 58–62 (1984)
19.29 J. Kergomard, R. Caussé: Measurement of acoustic impedance using a capillary: An attempt to achieve optimization, J. Acoust. Soc. Am. **79**(4), 1129–1140 (1986)
19.30 S. Bilbao, R. Harrison, J. Kergomard, B. Lombard, C. Vergez: Passive models of wave propagation in acoustic tubes, J. Acoust. Soc. Am. **138**, 555–558 (2015)
19.31 S. Bilbao, R. Harrison: Passive time-domain numerical models of viscothermal wave propagation in acoustic tubes of variable cross-section, J. Acoust. Soc. Am. **140**, 728–740 (2016)
19.32 T. Hélie, R. Mignot, D. Matignon: Waveguide modeling of lossy flared acoustic pipes: Derivation of a Kelly–Lochbaum structure for real-time simulations. In: *IEEE Workshop Appl. Signal Process. Audio Acoust., New Paltz* (2007) pp. 267–270
19.33 R. Mignot, T. Hélie, D. Matignon: Digital waveguide modeling for wind instruments: Building a state-space representation based on Webster–Lokshin model, IEEE Trans. Audio Speech Lang. Process. **18**(4), 843–854 (2010)
19.34 O.V. Rudenko, S.I. Soluyan: *Theoretical Foundations of Nonlinear Acoustics* (Consultants Bureau, New York 1977)
19.35 S. Adachi, M. Sato: Time-domain simulation of sound production in the brass instrument, J. Acoust. Soc. Am. **97**(6), 3850–3861 (1995)
19.36 S. Adachi, M. Sato: Trumpet sound simulation using a two-dimensional lip vibration model, J. Acoust. Soc. Am. **99**(2), 1200–1209 (1996)
19.37 H. Levine, J. Schwinger: On the radiation of sound from an unflanged circular pipe, Phys. Rev. **73**(4), 383–406 (1948)
19.38 R. Harrison, S. Bilbao, J. Perry, T. Wishart: An environment for physical modeling of articulated brass instruments, Comput. Music J. **39**(4), 80–95 (2015)
19.39 R. Harrison-Harsley: *Physical Modelling of Brass Instruments Using Finite-Difference Time-Domain Methods*, PhD Thesis (Acoustics and Audio Group, University of Edinburgh, Edinburgh 2017)
19.40 J.O. Smith III: Efficient simulation of the reed-bore and bow-string mechanisms. In: *Proc. Int. Comput. Music Conf., The Hague* (1986) pp. 275–280
19.41 T. Hélie: Unidimensional models of acoustic propagation in axisymmetric waveguides, J. Acoust. Soc. Am. **114**(5), 2633–2647 (2003)
19.42 H. Brezis: *Functional Analysis, Sobolev Spaces and Partial Differential Equations* (Springer, New York 2011)
19.43 J. Conway, N.J.A. Sloane: *Sphere Packings, Lattices and Groups* (Springer, New York 1988)
19.44 D. Botteldooren: Finite-difference time-domain simulation of low-frequency room acoustic problems, J. Acoust. Soc. Am. **98**, 3302–3308 (1995)
19.45 J. Botts, L. Savioja: Integrating finite difference schemes for scalar and vector wave equations. In: *IEEE-ICASSP, Vancouver* (2013) pp. 171–175
19.46 S. Bilbao: Modeling of complex geometries and boundary conditions in finite difference–finite volume time domain room acoustics simulation, IEEE Trans. Audio Speech Lang. Process. **21**(7), 1524–

1533 (2013)
19.47 S. Bilbao, B. Hamilton, J. Botts, L. Savioja: Finite volume time domain room acoustics simulation under general impedance boundary conditions, IEEE Trans. Audio Speech Lang. Process. **24**(1), 161–173 (2016)
19.48 G.D. Smith: *Numerical Solution of Partial Differential Equations: With Exercises and Worked Solutions* (Oxford Univ. Press, Oxford 1965)
19.49 W.F. Spotz, G.F. Carey: A high-order compact formulation for the 3-D Poisson equation, Numer. Methods Partial Differ. Equ. **12**(2), 235–243 (1996)
19.50 B. Hamilton, S. Bilbao, C.J. Webb: Revisiting implicit finite difference schemes for 3-D room acoustics simulations on GPU. In: *DAFx* (Univ. of Erlangen, Erlangen 2014)
19.51 B. Hamilton: *Finite Difference and Finite Volume Methods for Wave-Based Modelling of Room Acoustics*, PhD Thesis (Acoustics and Audio Group, University of Edinburgh, Edinburgh 2016)
19.52 A. Love: The small free vibrations and deformation of a thin elastic shell, Philos. Trans. R. Soc. Lond. A **179**, 491–546 (1888)
19.53 E. Reissner: The effect of transverse shear deformation on the bending of elastic plates, J. Appl. Mech. **12**, 69–77 (1945)
19.54 R. Mindlin: Influence of rotatory inertia and shear on flexural motions of isotropic, elastic plates, J. Appl. Mech. **18**, 31–38 (1951)
19.55 A. Chaigne, C. Lambourg: Time-domain simulation of damped impacted plates I. Theory and experiments, J. Acoust. Soc. Am. **109**(4), 1422–1432 (2001)
19.56 K. Arcas, A. Chaigne: On the quality of plate reverberation, Appl. Acoust. **71**(2), 147–156 (2010)
19.57 E. Jansson: *Acoustics for Violin and Guitar Makers* (Department of Speech, Music and Hearing, Stockholm 2002)
19.58 N. Giordano: Simple model of a piano soundboard, J. Acoust. Soc. Am. **102**(2), 1159–1168 (1997)
19.59 A. Chaigne, C. Touzé, O. Thomas: Nonlinear vibrations and chaos in gongs and cymbals, Acoust. Sci. Technol. **26**(5), 403–409 (2005)
19.60 A.H. Nayfeh, D.T. Mook: *Nonlinear Oscillations* (Wiley, New York 1979)
19.61 K.-J. Bathe: *Finite Element Procedures* (Prentice Hall, Upper Saddle River 1996)
19.62 M.J. Elejabarrieta, A. Ezcurra, C. Santamaria: Vibrational behaviour of the guitar soundboard analysed by the finite element method, Acta Acust. united Acust. **87**(1), 128–136 (2001)
19.63 J. Berthaut, M.N. Ichchou, L. Jezequel: Piano soundboard: Structural behavior, numerical and experimental study in the modal range, Appl. Acoust. **64**(11), 1113–1136 (2003)
19.64 S.A. Van Duyne: *Digital Filter Applications to Modeling Wave Propagation in Springs, Strings, Membranes and Acoustical Space*, Ph.D. Thesis (Center for Computer Research in Music and Acoustic, Stanford Univ., Stanford 2007)
19.65 C. Camier, C. Touzé, O. Thomas: Non-linear vibrations of imperfect free-edge circular plates and shells, Eur. J. Mech. A **28**(3), 500–515 (2009)
19.66 M. Ducceschi, C. Touzé, S. Bilbao, C.J. Webb: Nonlinear dynamics of rectangular plates: Investigation of modal interaction in free and forced vibrations, Acta Mechanica **225**(1), 213–232 (2014)
19.67 C. Lambourg, A. Chaigne, D. Matignon: Time-domain simulation of damped impacted plates. II. Numerical model and results, J. Acoust. Soc. Am. **109**(4), 1433–1447 (2001)
19.68 S. Bilbao: A digital plate reverberation algorithm, J. Audio Eng. Soc. **55**(3), 135–144 (2007)
19.69 S. Bilbao: Percussion synthesis based on models of nonlinear shell vibration, IEEE Trans. Audio Speech Lang. Process. **18**(4), 872–880 (2010)
19.70 S. Timoshenko, S. Woinowsky-Krieger: *Theory of Plates and Shells*, Vol. 2 (McGraw-Hill, New York 1959)
19.71 O. Thomas, S. Bilbao: Geometrically nonlinear flexural vibrations of plates: In-plane boundary conditions and some symmetry properties, J. Sound Vib. **315**(3), 569–590 (2008)
19.72 S. Bilbao, L. Savioja, J.O. Smith III: Parametrized finite difference schemes for plates: Stability, the reduction of directional dispersion and frequency warping, IEEE Trans. Audio Speech Lang. Process. **15**(4), 1488–1495 (2007)
19.73 A. Torin: *Percussion Instrument Modelling In 3D: Sound Synthesis Through Time Domain Numerical Simulation*, PhD Thesis (Acoustics and Audio Group, University of Edinburgh, Edinburgh 2015)
19.74 S. Bilbao, J.O. Smith III: Energy conserving finite difference schemes for nonlinear strings, Acustica **91**, 299–311 (2005)
19.75 H. Conklin: Piano strings and phantom partials, J. Acoust. Soc. Am. **102**, 659 (1997)
19.76 S. Bilbao: Conservative numerical methods for nonlinear strings, J. Acoust. Soc. Am. **118**(5), 3316–3327 (2005)
19.77 J. Chabassier: *Modeling and Numerical Simulation of the Piano Through Physical Modeling*, Ph.D. Thesis (Ecole Polytechnique, Paris 2012)
19.78 S. Bilbao: A family of conservative finite difference schemes for the dynamical von Karman plate equations, Numer. Methods Partial Differ. Equ. **24**(1), 193–216 (2008)
19.79 T. Rossing, N. Fletcher: Nonlinear vibrations in plates and gongs, J. Acoust. Soc. Am. **73**(1), 345–351 (1983)
19.80 C. Vyasarayani, S. Birkett, J. McPhee: Modeling the dynamics of a vibrating string with a finite distributed unilateral constraint: Application to the sitar, J. Acoust. Soc. Am. **125**(6), 3673–3682 (2010)
19.81 S. Bilbao, A. Torin: Numerical modeling and sound synthesis for articulated string/fretboard interactions, J. Audio Eng. Soc. **63**(5), 336–347 (2015)
19.82 S. Bilbao, A. Torin, V. Chatziioannou: Numerical modeling of collisions in musical instruments, Acta Acust. united Acust. **101**(1), 155–173 (2015)
19.83 A. Hirschberg, J. Gilbert, R. Msallam, A. Wijnands: Shock waves in trombones, J. Acoust. Soc. Am. **99**(3), 1754–1758 (1996)
19.84 G. Sod: A survey of several finite difference methods for systems of nonlinear hyperbolic conserva-

tion laws, J. Comput. Phys. **27**(1), 1–31 (1978)
19.85 R. Leveque: *Finite Volume Methods for Hyperbolic Problems* (Cambridge University Press, Cambridge 2002)
19.86 B. Lombard, D. Matignon, Y. Le Gorrec: A fractional Burgers equation arising innonlinear acoustics: Theory and numerics. In: *Proc. 9th IFAC Symp. Nonlinear Contr. Syst., Toulouse* (2013)

20. Real-Time Signal Processing on Field Programmable Gate Array Hardware

Florian Pfeifle

Over the last 50 years, advances in high-speed digital signal processing (DSP) and numerical methods for audio signal processing in general were fueled by the rising processing capabilities of personal computers (PCs). Added to this was the advent of specialized coprocessing platforms like general purpose graphics processing units (GPGPUs), central processing unit (CPU)-based accelerators like Intel's Xeon Phi platforms as well as high-performance digital signal processing (DSP) chips like Analog Devices' TigerSHARC. Still, there are applications that are not realizable on the mentioned devices in real time or even close to real time. This chapter gives an introduction to field programmable gate array (FPGA) hardware, a flexible computing platform with massively parallel logic capability that is applicable for problems of high data throughput, high clock rates and high parallelism. After an introduction to the basic structure of FPGAs, several features that enable high-throughput DSP applications are highlighted. An introduction to development platforms as well as the development methodology is given, along with an overview of current FPGA devices and their specific capabilities. Two application examples and an outlook and summary complete this chapter.

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_20

20.1 Overview

Sometime in October 2013 the number of mobile devices equaled the number of humans on planet Earth. This means we have more than 7.4 billion mobile phones, portable computers or other gadgets on our planet, not including the vast amount of nonmobile digital devices, and this number is constantly rising [20.1]. In every single one of them some sort of digital signal processing (DSP) takes place. Take for instance a mobile phone where we have a microphone that transfers the pressure variations in air to an electric signal, which is then transformed to a digital signal by an analog-to-digital converter (ADC). The digitized signal is then processed, converted again, and sent via the phone's antenna to a mobile phone mast where it is transfered to another cell phone. This example should illustrate how we are constantly surrounded by small or large devices performing DSP operations without our direct interaction and that DSP applications have become an important part of our daily lives.

In the humanities, especially in systematic musicology and related research areas, the application of DSP methods for recording, processing, analyzing and auralizing acoustic music or speech signals has become indispensable and is a central component of a musicologist's toolbox. This is mainly due to the rising computational throughput of standard personal computers and easily accessible accelerators like DSP chips or GPUs. Nonetheless, there are certain restrictions when using classical hardware devices to compute signal processing methods. One huge and very annoying problem is sample latency introduced by a digital system in a music processing tool-chain. Another problem is the high data throughput of modern high-sample-rate, multichannel audio and/or video streams, which stretches the computational capacity of common hardware devices. Field programmable gate arrays, being equipped with massively parallel logic circuitry, can be used to minimize or even dissolve the aforementioned bottlenecks.

This chapter tries to elucidate the usage of field programmable gate array (FPGA) hardware for high-throughput high-sample-rate problems, gives an introduction to FPGAs and shows some design practices that have been shown to yield good results when applied to signal processing algorithms. The focus is put on a real-time applicability of DSP implementations in recent FPGA platforms and hardware and the structural features that enable this functionality, as well as certain design aspects of hardware models for signal processing applications.

20.1.1 Technology Overview

The umbrella term digital signal processing is commonly used as a description for conceptually comparable numerical methods that are mainly concerned with measuring, digitizing, processing and evaluating analog signals using digital hardware systems of varying kinds. DSP techniques and appliances are used in numerous areas of engineering including measurement and control systems [20.2], analysis applications in medical imaging [20.3] and audio processing applications [20.4] as software synthesizer or digital hardware effect units [20.5, 6].

Despite the varied application areas of DSP techniques there are several conjoining properties regarding the respective processing steps. In its basic form, the processing chain in most DSP applications can be summarized as: (1) An analog signal of some sort is (2) digitized using an ADC, which leads to a one- or multidimensional array of (3) discrete values, which represent the analog signal. (4) The digitized values are subject to some kind of numerical processing like filtering or analyzing. (5) After the data-processing step, the values are transformed to a humanly comprehensible representation like a graphic depiction on some scale or (6) are converted back to an analog signal again using a digital-to-analog converter (DAC). This basic processing chain is depicted in Fig. 20.1.

A practical example could be a digital effects unit with an integrated spectral analyzer. Here, a physical value, like the alternating current produced by an electronic guitar pickup, is converted to a digital representation via an ADC. In a further processing step, the frequency content of the digitized guitar signal is analyzed using a spectral transform. Finally the results of the frequency analysis are depicted in a graph on a display, representing the frequency content of the signal.

In this not-so far-fetched example the throughput of the system depends on three factors at least: the sampling rate i. e., the acquisition of discrete values; the latency of the numerical method; and the latency introduced by the data representation method. Depending on the respective DSP application, one of these three stages can act as a bottleneck for data throughput. This is one of the reasons that highly optimized and specialized DSP chips, like the high-throughput series of Analog Devices' TigerSHARC processor or Texas Instruments' C66x series chips, are applied when signal processing operations are used in timing-critical, high-data-throughput applications.

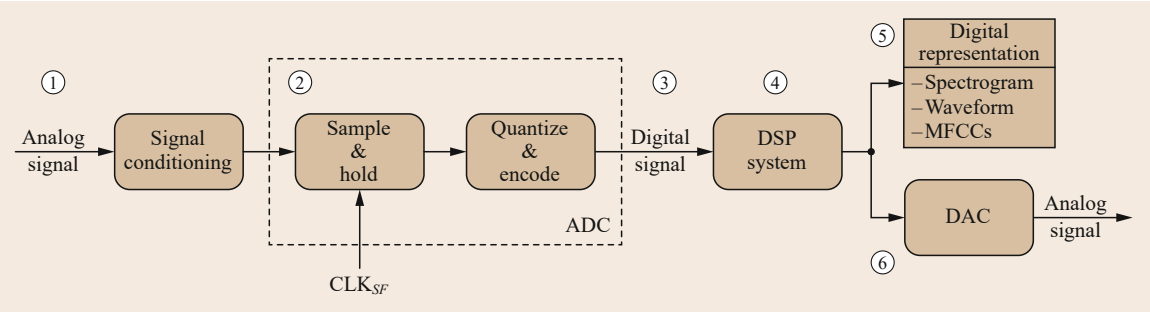


Fig. 20.1 A schematic view of a DSP chain

Getting back to our example, imagine that now 500 channels should be analyzed at once, or the system should be able to sample the input with a sample rate of several MHz. This somewhat contrived example should illustrate that, as electronic systems are getting ever more elaborate, there are applications that are not realizable with feasible effort using conventional DSP chips.

As a subclass of microprocessors with a Harvard memory architecture, DSPs are specifically designed for serial signal processing applications. This means they are especially well suited for high throughput rates in algorithm-intensive computations. In particular, they are very well adapted to perform sequential multiply-and-accumulate (MAC) operations on discrete data streams. This means that the only way to create parallelization would be to use multiple DSPs, but this practice would become unpractical very quickly due to the rising interface complexity and the additional power demands of multiple DSPs.

Using FPGAs – programmable devices with massively parallel hardware logic – for such tasks has been shown to provide a valid alternative for DSP applications where parallel processing power or high sample rates are needed.

FPGAs are applied in various fields of engineering and information technology, as well as other scientific areas for appliances with particular requirements regarding interface flexibility, computational throughput or unusual circuit structure. In integrated hardware designs they are utilized as special function chips for glue-logic (interface-logic), specialized computation cores or design-specific function registers. FPGAs can be found as processing cores in sound cards as a replacement for DSP chips or as processing cores of high-accuracy high-sample-rate oscilloscopes.

By applying FPGA hardware, it is possible to accelerate certain classes of DSP applications, thus making them realizable on a single hardware chip or enhancing their real-time capabilities. This is mainly due to the parallel and flexible hardware layout of FPGA chips. This means that a specialized design can be realized on a FPGA without the constraints of a certain processor architecture like the common *Harvard* architecture used in many DSP chips or the *von Neumann* architecture found in many other microprocessor designs – as a historical side note it is necessary to mention that the *von Neumann* architecture, named after the Hungarian-American scientist John von Neumann, was used in an earlier computer design by *Konrad Zuse* in his Z1 computer [20.7]. Every conceivable form of logic circuit is realizable with a FPGA depending only on the size and the logic capabilities of the FPGA hardware device. When designing for FPGA hardware, it is possible to have control over implementation details from a comparably high abstraction level down to the bare metal implementation on the FPGA chip itself. Unlike other platforms, where certain steps in the development cycle are hardly accessible (like compile-time optimizations) or assessable (like undefined runtime behavior of a software implementation on a PC due to unknown scheduling behavior of an operating system), freely programmable logic devices make it possible to have control over every last part of an algorithm.

Another benefit of FPGA implementations, which is an important deliberation in current high-performance computing applications, is their comparably low power profile. When comparing implementations of algorithms on GPUs, CPUs and FPGAs the number of instructions per Watt are orders of magnitude better in FPGAs [20.8]. On the other side of the performance spectrum, low-power FPGAs could play a central role in small intelligent devices used in the Internet of things or in the growing do-it-yourself (DIY) electronics community where open-source hardware projects are gaining a lot of interest on crowd-funding platforms. An example of a crowd-funded programmable hardware board is the *Snickerdoodle*, which has a FPGA with an included microprocessor on a development board.

20.2 Digital Binary Logic

Binary digital logic is the building block of all logic devices used today ranging from *plain* circuitry like small music chips in children's books, digital watches, or control hardware in cars, to more elaborate designs as found in microprocessors, graphics cards or CPUs of modern personal computers. In this chapter a short introduction and a small amount of background information regarding binary logic is given.

20.2.1 History of Binary Logic

One of the earliest philosophical and mathematical contemplations regarding binary logic in the European science tradition was published by *Gottfried Wilhelm von Leibnitz* (1646–1716) [20.9]. In this treatise he describes a binary number system, using the values 0 and 1, and develops rules for four fundamental algebraic operations: addition, subtraction, multiplication and division. In the same text, he relates this number system to the hexagram figures found in the ancient Chinese scripture *I-Ching* [20.10, pp.117]. A comparable algebraic system as used for the binary number system published by Leibniz was developed almost two centuries later by *George Boole* (1815–1864) in his two publications [20.11, 12]. Contrary to modern belief, it was not exclusively developed for *binary* number systems but was a general description of an algebra of logic [20.13], aimed at extending Aristotelean syllogism [20.14]. Nonetheless, it was shown in the 1930s in the master's thesis of *Claude Shannon* that the formal logic calculus of Boole could be applied to describe and construct electric circuits using algebraic (mathematical) means [20.15]. Nowadays, the binary logic system is applied as the basis for all digital logic devices. In honor of George Boole, its appertaining calculus is called Boolean algebra or, as proposed by Shannon, switching algebra. Similar to a decimal number system, the binary number system is a positional number system that allows the representation of numbers as position-relative digits of powers of two.

Using Boolean algebra, complex algorithms can be represented in digital hardware by using only two discrete states: 1 and 0. When dependent on formal logic they can also be interpreted as two logical truth values: *true* and *false*. In physical terms this can be related to an open or closed switch, a flowing current versus a nonflowing current or a high and low voltage level as shown in [20.15]. In modern digital systems the two logical values 1 and 0 are commonly realized by high and low voltages using specific transistor circuits, which are called logic gates.

20.2.2 Binary Logic

The fundamental algebraic functions that can be expressed in binary logic are logic conjunction *AND*, logic disjunction *OR*, exclusive disjunction *XOR* and logic negation *NOT*. Combining the negation with the other logic functions gives *NAND*, *NOR* and *XNOR*. Written in a truth table, two-input logic gates can produce the outputs given in Table 20.1.

20.2.3 Binary Logic in Hardware

The logical values given in Table 20.1 are the building blocks of digital circuitry of FPGA hardware as well as all other digital hardware chips. Depending on the device and on the vendor, there are different methods for how binary logic is realized in hardware. Early hardware systems used relays as on/off switches, as for instance used in *Konrad Zuses*' Z3 computer [20.7]. Another technique used in the first computers was electron tubes, as for instance in the code deciphering machine *Colossus* [20.16]. During the 1950s the large, error-prone electron tubes were gradually replaced by transistors, which were smaller and required less energy [20.17, pp. 65–75]. Modern integrated chips almost exclusively use transistor circuits to yield the logic values given in Table 20.1. The underlying design may differ depending on the type of the transistor, but their overall yield is the logic values given in Table 20.1. This means that the logic functions of a transistor circuit can be viewed on a more abstract level by only taking the logic value into consideration and moving structural design considerations to a lower abstraction layer.

A logic gate representation of basic Boolean algebra functions in gate logic is given in Fig. 20.2. The plain combination of two logic inputs can be extended

Table 20.1 Logic values for two-input (P,Q) Boolean functions

P	Q	AND (∧)	OR (∨)	XOR (⊕)	NAND	NOR	XNOR
0	0	0	0	0	1	1	1
0	1	0	1	1	1	0	0
1	0	0	1	1	1	0	0
1	1	1	1	0	0	0	1

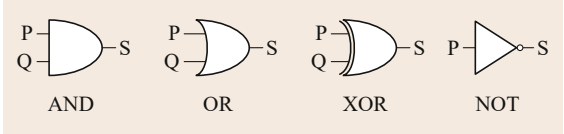


Fig. 20.2 Basic logic operations in gate logic

to more complex circuity if one wants to model an algebraic function of higher order using a data type that is wider than 2 bit. This means that complex functions can be realized by an interconnection of multiple basic gates. Indeed, modern microchips are based on simple logic gates made of transistors but are so complex that they can have up to several billion transistors on one die. Using the equivalent representations of Fig. 20.2 and Table 20.1 logic circuitry can be constructed by applying rules of Boolean algebra. To implement a 2 bit adder one can start by writing down the inputs and the outputs of the respective circuit and truth table.

By inspecting the output values and comparing them with the truth table given in Table 20.1 it is clear that the output signal S can be realized with an *XOR* gate and the output C_{Out} by an *AND*. An equivalent representation of this circuit using logic gates is given in Fig. 20.3.

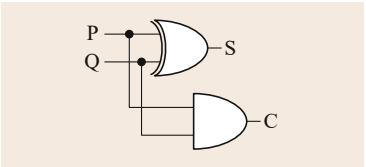


Fig. 20.3 An equivalent logic gate representation of the binary adder given in Table 20.2

Table 20.2 Truth table of a binary adder with input signals P and Q, output S and carry output C

P	Q	S	C_{out}
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

20.2.4 Number Systems in Digital Hardware

Besides the algebraic implementation details of an algorithm in hardware, another important factor that influences a system's throughput is the number representation in the digital domain. As will be shown in Sect. 20.3, the choice of the right data type can speed up and/or optimize the throughput of a signal processing method considerably.

In DSP applications and on dedicated DSP hardware chips there are several standard data types and some not-so-standard data types, which can be used to represent digitized signals. All data types can be divided in two main classes: (a) fixed-point and (b) floating-point. Because FPGAs are freely programmable logic chips, the data type is not fixed upfront but can be chosen according to the respective type of application. On the one hand, this flexibility allows for highly specialized design decisions, but on the other hand it can lead to a heightened development expense because different data types have to be assessed regarding their applicability for a certain problem before the actual design phase.

Fixed-Point Number Representation

Fixed-point data types are commonly used in older DSP hardware as well as in low-power devices, which do not have a floating-point unit. Their strength lies in their simplicity, as the number representation is amenable to representation by binary digital hardware. As a rule of thumb one can state that in comparison to floating-point, fixed-point arithmetic is faster but lacks the dynamic range of floating-point. This means if an application calls for a large data range floating-point should be preferred, otherwise one is most likely to get the better performance out of a fixed-point implementation.

There are several fixed-point data types used to represent numbers in binary code. They can be divided into:

- Unsigned integer representations
- Signed integer representations.

Table 20.3 shows a 4 bit representation of several numbers in signed and unsigned formats.

As given in Table 20.3, the data range of unsigned is 0 to $2^N - 1$, with N the bit width. One's complement (1C) can represent data in the range of $-2^{N-1} + 1$ to $2^{N-1} - 1$; two's complement (2C) in the range -2^{N-1} to $2^{N-1} - 1$.

One's complement numbers have a bit-by-bit complement representation of positive and negative values, and this includes the redundant representation for -0. This means the inversion can be realized in logic by replacing every bit by its logical inverse using a *NOT* gate as illustrated in following short example.

$$\begin{aligned} 1011 &= -4 \\ 0100 &= 4 \end{aligned} \tag{20.1}$$

The inversion of 2C numbers is comparable to the 1C inversion but needs an additional step: an addition of 1 to the inverted number.

$$\begin{aligned} 1011 &= -5 \\ 0100 &= 4 | + 1 \\ 0101 &= 5 \end{aligned}$$

A feature that makes 2C numbers the most widely used fixed-point signed number representation in digital hardware is the property that when summing 2C numbers, all overflows of intermediate computations can be ignored as long as the final sum is in the bit range N.

Table 20.3 Comparison of unsigned, signed one's complement, and two's complement data types

Bit values	Unsigned	One's complement	Two's complement
1000	8	−7	−8
1001	9	−6	−7
1010	10	−5	−6
1110	14	−1	−2
1111	15	−0	−1
0000	0	0	0
0001	1	1	1
0010	2	2	2
0011	3	3	3
0100	4	4	4
0111	7	7	7

Floating-Point Number Representation

In most modern DSP applications that are implemented on a PC, numbers are represented by the floating-point number system. Most recent microprocessors and CPUs have a dedicated floating-point unit, which performs optimized floating-point arithmetic. The number system which is defined in the IEEE-754 (Institute of Electrical and Electronic Engineering) standard was designed to represent numbers over a large range, making it possible to represent values in several magnitudes by using only small bit sizes. One drawback of the this large range is the resolution inside the range, which follows a nonlinear distribution. Smaller numbers have a higher resolution than the larger ones. Floating-point number representation can be imagined as a special type of logarithmic number representation. Computations with logarithms, multiplication, division and exponentiation are comparably easier to perform than additions or subtractions. This is why they are often used for scientific calculations where values from a large range should be representable and the multiplication operation is important. A floating-point number consists of a sign bit S, an exponent E and an unsigned mantissa M and is structured as

20.3 FPGA – A Structural Overview

To fully understand the particular features that enable FPGAs to be the highly versatile, high-performance processors of choice for DSP and other areas of scientific computations today, an overview of structural design aspects is presented in this section. A short overview of their historical evolution is given and the associated design changes that are essential for the understanding of modern FPGA chips are highlighted. A focus is put on hardware features that are convenient for signal processing applications. In particular, features that can be used in real-time audio processing applications are of concern here. Additionally, several input/output (I/O) protocols are presented to illustrate how communication with FPGA hardware can be realized.

20.3.1 History of FPGAs

The historic development of FPGA devices is closely linked to the evolution of integrated digital circuits in the late 1960s, 1970s and early 1980s. The development of logic devices of that time period was mainly driven by the manifold advances in transistor and integrated circuit technologies. For an encompassing overview see for instance [20.18].

Custom logic devices of that time period can roughly be divided into two categories:

1. Programmable logic devices (PLDs)
2. Nonprogrammable devices like application-specific integrated circuits (ASICs) or application-specific standard parts (ASSPs).

The difference between both device classes can be found in their central structure. PLDs are only partially wired, whereas logic functions of ASICs are hardwired in the production process. This means that a specific function of an ASIC is *fixed* and cannot be altered by a designer after the production state, while a PLD can be programmed later.

These fundamental differences lead to differing design and implementation practices and thus to differing fields of applications. In comparison to PLDs, ASICs have a higher logic gate count and are mainly used to implement highly specialized functions, but have the drawback of long production cycles for implementation, prototyping and debugging. If an ASIC has inherent design errors, they often surface *after* a first prototype is manufactured and the appliance is running under realistic conditions in the respective field of application. PLDs and programmable read-only memory (PROMS) devices can be programmed after the production state, which has the advantage that a faulty chip design does not lead to a completely erroneous production charge. Still, compared to most ASICs, PLDs have considerably smaller logic capabilities.

The first programmable devices had a transistor array structure that could be flashed with different logic

circuit designs by adding connections (antifuse technology) [20.18, p. 12], or removing connections (fusible link technology) [20.18, p. 10]. The added or removed links were permanent, so the devices could only be programmed once. Due to this, they were capable of performing specially designed tasks with very high clock rates, but they could not be reconfigured after a design was implemented on the device.

Further advances in the field of programmable integrated circuits led to technologies that made it possible to erase an initial design and reprogram the device. The most prominent technologies among these devices are erasable programmable read-only memory (EPROM) or electrically erasable PROM (EEPROM). Devices incorporating these technologies could be programmed and reprogrammed multiple times by removing the connections of the design on the metal layer by exposing it to ultraviolet light or to a certain voltage respectively.

Up to this point, there always existed a gap between the logic capabilities of the different hardware device classes. On the one hand there were PLDs, which were highly configurable but could only perform a small amount of logical tasks, and on the other hand there were ASICS, which could perform highly complex logic circuits but were not reconfigurable, and were expensive and demanding to develop for.

The next large leap in the evolution of freely programmable hardware was sparked by research done by Ross Freeman and Bernard Vonderschmitt, the founders of Xilinx, who were the first to develop and produce freely programmable hardware-gate logic on a large scale [20.19]. In 1985 the first commercially available programmable logic chip was the XC2064, called a FPGA. Field programmable literally means programmable in the field, outside of the laboratory. The name FPGA is still used today for logic devices of similar design structure and logic capabilities. The first FPGA device had 64 programmable and freely connectible logic blocks (CLBs) and an aggregate gate count of 1200 logic gates, modern FPGAs, like a Virtex-7 for instance, often contain several million logic blocks. FPGAs were developed as a devices that combined the programmability of PLDs, the reconfigurability of EEPROMS, and the high logic-gate count of ASICS [20.18, pp. 49 ff.]. This period of time is often regarded as the starting point for the development of more advanced FPGA devices by Xilinx and other vendors. At the present day, the biggest FPGA chip vendors are Xilinx and Altera among other smaller companies such as Lattice Semiconductor or Microsemi [20.18, pp. 161 ff.].

The development tools provided by the mentioned vendors helped establish modern FPGAs as highly flexible, programmable hardware platforms, ideal for developing custom high-throughput circuit designs. But this is only part of their strength: modern FPGAs have evolved from the plain logic devices of early days so tremendously that they are now able to host multiple high-performance microprocessor architectures on one chip [20.20]. In addition to the flexibility of the freely programmable logic, the mix of on-board memory resources and additional special function logic blocks enhanced the capabilities of FPGA devices but also increased the complexity of the design process. This means that the full potential can only be realized if the special, massively parallel, chip structure is fully utilized in a design.

A more recent development in FPGA technology is the acquisition of Altera, the second largest FPGA vendor, by Intel and the announced cooperation between IBM and Xilinx in 2015. This follows a current trend in high-performance computing and data-center-based cloud computing applications that focuses on flexible, energy-efficient infrastructure, which is planned to be realized by a proposed combination of CPUs and FPGAs in servers as well as in user devices [20.21, 22].

20.3.2 Structural Layout of FPGA Hardware

In modern FPGA devices logic gate functionality is realized by look-up tables (LUTs) (LUTs are also called logic function generators [20.23, p. 2 ff.]), which are, in a sense, addressable function generators. This means that they can be programmed to perform different logic functions on a set of inputs.

These basic logic cells are part of a larger logic conjunction, which is called a *slice* in Xilinx devices. All slices contain one LUT, eight storage elements, multiplexers and carry logic [20.24]. Some Xilinx slices, called *slicem* [20.24], additionally contain functions to store data as distributed random access memory (RAM) and have 32 bit-wide shift registers.

Two slices form a configurable logic block (CLB), as is depicted in Fig. 20.4. It shows a (LUT)-based CLB, as commonly found in Xilinx FPGAs. Earlier Xilinx FPGAs, such as the Virex-II pro, contained LUTs with four inputs. Newer Virtex-7 devices have configurable six-input LUTs.

The LUTs inside a CLB can operate in several different input/output modes depending on the device class and generation. In addition to being connectible as six-input two-output LUTs, they can be configured as shift registers, RAM blocks or first-in/first-out memory. In some CLBs there are additional latches and in others are flip-flop cells [20.24], which can be used as asynchronous or synchronous registers.

The outputs of the single CLBs are connected to a programmable interconnection network, which is at-

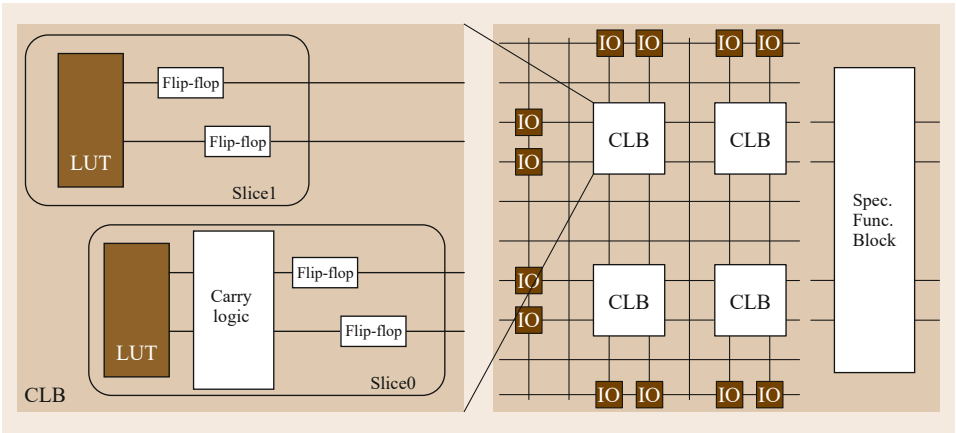


Fig. 20.4 Schematic overview of a FPGA showing an exploded view of a CLB, the interconnect network, in/out ports and special function registers

tached to a multiplexer on the output stages, multiplexing output signals to input stages of other CLBs. This cascading of CLBs allows for more complex logic functions, larger RAM blocks or longer shift registers.

In Altera devices, the large-scale structure of basic logic units is comparable to Xilinx setups, but differs in several stages of internal structure. The equivalent of a Xilinx CLB in a Stratix-V is called a *logic array block*, which contains 10 *adaptive logic modules* (ALMs) [20.25]. Each ALM is composed of two LUTs, each with six inputs. In addition to this, an ALM can contain extra logic, like programmable registers or full adders. How the input/output routing of an ALM is configured depends on the mode of operation of an ALM:

- Normal mode: Two functions in one ALM with up to eight inputs or one function with six inputs are realizable.
- Extended LUT mode: Conditional if-else statements are realizable with seven inputs and one input for register packing.
- Arithmetic mode: The internal adder or other arithmetic functions like counters or accumulators can be implemented.
- Shared arithmetic mode: The ALM functions as a three-input adder or carry bit circuit.

20.3.3 Special Function Blocks

Besides the basic logic cells, FPGAs contain other functional logic blocks that differ from vendor to vendor and from device generation to device generation. Some additional logic blocks that are typically found in most modern FPGAs are memory blocks in the form of random access memory (RAM), first in/first out (FIFO) memory blocks or other kinds of unregistered or registered memory. Logic blocks that extend basic FPGA logic by integrated circuit implementations of arithmetic function are for instance DSP blocks with integrated MAC circuitry. All special function blocks are on the FPGA die in close proximity to the logic gate resources and can be connected to the same high-speed interconnection network the core logic is connected to, as is shown in Fig. 20.4.

In modern Xilinx FPGAs the CLBs and the special function blocks are arranged in a column layout. Figure 20.5 depicts a schematic overview of the arrangement of DSP48e1, block RAM (BRAM) and logic slices.

DSP Blocks

Most modern FPGAs by the vendors Xilinx or Altera have special logic blocks that are designed to perform DSP typical operations. They are implemented on the same structural level as CLBs, which means that the data transfer between gate-based logic and DSP cores can be realized by the high-speed interconnect network.

Figure 20.6 depicts a schematic overview of a DSP48e1 slice, which is part of most modern Xilinx FPGAs. It consists of four inputs, a pre-adder, a 25 bit × 18 bit multiplier, a 48 bit accumulator and a pattern detector that can be used to efficiently compare values.

To derive benefit from the fast MAC operation there are several guidelines when designing with DSP

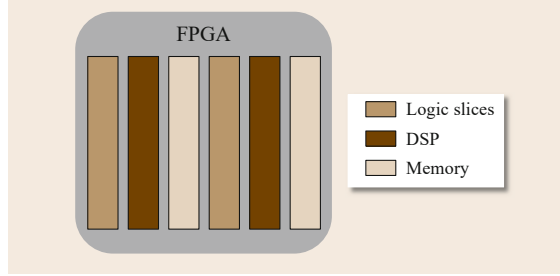


Fig. 20.5 Schematic overview of Xilinx column design

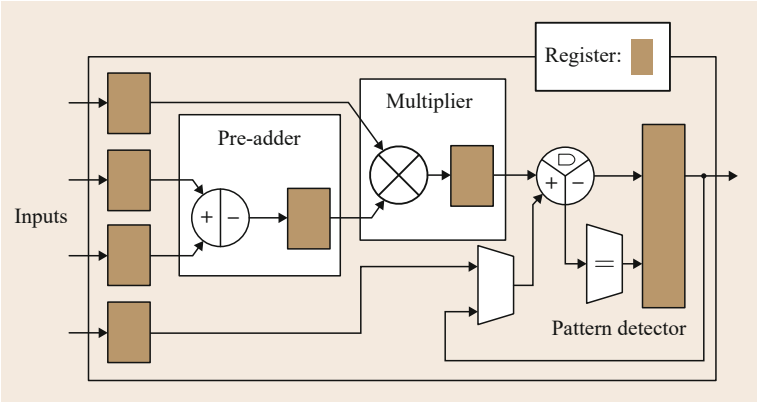


Fig. 20.6 Schematic overview of a Virtex-6 DSP48e1. *Blue blocks* are registers

blocks that, when followed, can lead to highly performant algorithms. To obtain optimal results, several recommendations for using DSP48E1 slices in a design should be followed, which include:

- Filter coefficients or constant multiplicands are best stored in a BRAM block.
- If possible, the design should be pipelined.
- Using a fixed-point signed data type yields the highest throughput.
- Small multipliers should be implemented in CLB logic.
- A time-multiplexed design should be implemented if applicable.

If all these recommendations can be followed and the design makes use of all DSP slices, it is possible to achieve a sustained throughput of several TMACS (tera multiply-and-accumulates per second) on modern FPGA devices.

RAM Blocks

Another important factor when designing DSP applications on FPGAs is the availability of memory resources. In modern Xilinx devices RAM can be implemented by combining several CLBs, which is called distributed RAM or DisRAM. Another sort of dedicated RAM on Virtex FPGAs is called block RAM (BRAM). It can be configured to different sizes and different function modes. Every BRAM block can perform in dual-port mode facilitating 36 Kb of memory. A similar dedicated RAM is present in Altera devices and is called a M20 memory block.

Table 20.4 Comparison of Xilinx FPGA generations

FPGA device	Slices	Logic cells	Max. RAM (kB)	Max. DSP slices
Virtex-II Pro	44 096	99 216	1378	444[a]
Virtex-5	51 840	331 776	3420	192
Virtex-6	118 560	758 584	8280	2016
Virtex-7	305 400	1 954 560	67 680	3600
Virtex-Ultrascale		5 541 000	132 900	5520
Virtex-Ultrascale+		3 763 000	354 500	11 904

[a] Number of integrated 18 × 18 bit multipliers

Table 20.5 Comparison of several Altera FPGA generations

FPGA device	Logic cells	Max. RAM (kB)	Max. DSP blocks	18 × 18 wide mults.
Max 10	50 000	1638	–	144
Cyclone-4	150 000	6480	–	360
Stratix-IV	813 000	33 300	–	1288
Stratix-V	952 000	51 670	1963	3926
Stratix-10	5 510 000	166 000	5760	11 520[a]

[a] The Stratix-10 has 18 × 19 multipliers

Both configurable memory blocks can be used as memory of variable bit width and depth. Similar to the DSP blocks, the special RAM blocks are located on the FPGA chip and are connected to the CLBs via the internal routing network, which makes the communication and data transfer faster compared to communication with a peripheral RAM, which can be implemented on a hardware board connected to the FPGAs I/Os; see Fig. 20.5.

Tables 20.4 and 20.5 show the available DSP block and memory resources of different device generations from Xilinx and Altera.

20.4 Hardware Description Language (HDL)

In the early years of programmable hardware (the 1960s), digital logic circuits were designed, similar to analog electronic circuits, using pencil and paper. Gate-level logic functions, analogous to those shown in Fig. 20.2, were drawn and optimized by hand before being implemented in hardware. An advance in design methodology was the development of hardware description languages (HDLs), which were used to describe hardware functions on a syntactically higher abstraction level. Early versions of HDLs were used in the 1960 to describe functional parts of logic devices on a gate level. It was not until the late 1980s that HDLs began to replace the practice of schematic-based design of the increasingly complex ASIC circuitry of the time [20.26]. During the first stages of their development, the functionality and language design of HDLs were not standardized and most functions were vendor- and application-hardware-specific.

Today, there coexists two commonly applied hardware implementation methodologies for FPGA designs. One is more related to the older design techniques, but instead of using pencil and paper drawings a computer-aided design (CAD) program is used to draw schematic versions of a hardware implementation. This means it is possible to develop a schematic of connected logic gates, logic blocks or special function registers representing a special function of a hardware implementation. The other practice uses a HDL to describe the functionality of the hardware circuit [20.27]. There are two standardized HDLs commonly used in FPGA design: (1) Verilog and (2) VHDL (very high speed integrated circuit hardware description language). Both languages incorporate similar concepts, which can be categorized into *low-level* programming features like:

- Bitwise declaration of signals
- Control over electronic signal levels
- A direct access to electronic signals via physical input and output ports of the hardware development board.

and *high-level* language concepts like for instance:

- Object-oriented programming techniques
- Pointers to data types
- Procedural programming.

among other features.

To describe the complete range of functions of VHDL is far beyond the scope of this chapter and the interested reader is referred to the excellent publications regarding this subject area such as [20.28] or one of the other encompassing introductions to VHDL that can be found in [20.29–31].

Nonetheless, a few basic traits and characteristics shall be explained here. One fundamental differences of a HDL compared to other programming languages is the trait that it enables the developer to directly design hardware functionality using software statements. This means that lexical instruction are represented by gate logic equivalents on the respective hardware platform. Another difference, which is a regular pitfall for engineers with programming experience in a *high-level* language, is the fact that all code is evaluated concurrently. This means that instead of sequential code evaluation like in a compiled high-level language (C, C++, Java, Python, etc.) all instructions in the code are translated into a hardware function and evaluated at the same clock instance, if not specifically designed otherwise. Sequential statements can be implemented by a finite state machine (FSM).

20.4.1 Finite State Machine

A FSM is a clocked sequential circuit that switches its discrete states depending on the value of certain status signals. Therefore, sequential statements that need to be evaluated successively can be controlled in a synchronized way. There are different standard implementations of FSMs commonly used in hardware designs. The two most prominent are known as [20.32]:

1. Moore finite state machine – a state machine where the output only depends on the state
2. Mealy finite state machine – a state machine where the output depends on the state and an input signal.

In a real design implementation both FSM architectures can be mixed and distinguish themselves mostly by their underlying implementation in binary logic.

As an example of a FSM-controlled circuit take for instance the model of a signal generator that is able to generate sine, rectangular or triangular waves controlled by three switches. This means we have three switches that are on (1) or off (0). If two or three switches are simultaneously (1) then the output should be that of the last state. Figure 20.7 shows a FSM depiction of the single states.

20.4.2 VHDL Structure

A hardware model expressed in VHDL consists of several building blocks that follow a certain structure. The topmost level of a VHDL program is called an *entity*. An *entity* is a description of a circuit performing a certain logic function. An *entity* in VHDL can be seen as a box performing a certain logic function. It can consist of several input and output ports and has an embedded functionality defined inside the *entity*. An *entity* declaration of a 2 bit adder as presented in Sect. 20.2 can be formulated as shown in Fig. 20.8.

The example given in Fig. 20.8 shows the basic structure of an *entity* declaration with two input ports P and Q, and two output ports: S, the sum, and C, the carry bit. The functional behavior of an *entity* is defined inside its *architecture*. The body of the architecture consists of instructions that describe the logic functions of an *entity*. When describing the functional behavior of an *entity*, there are two fundamental design strategies, which can be labeled as: (a) structural modeling and (b) behavioral modeling. Approach (a) leads to a design that performs its function as a result of several substructures, which are *entities* as well. These substructures can be abstract components performing certain logic functions; by connecting several substructures a larger logic circuit can be realized. Approach (b) is a series of instructions that perform a certain logic function. It is comparable to functional programming in a high-level language, and in VHDL the functions are realized by several independent processes that perform the specific function. An architecture body following the entity declared in Fig. 20.8 can be constructed using the two logic gates given in Table 20.2. The two gates can be implemented as separate entities following design approach (b). A functional description of an *AND* gate and *XOR* gate can be written as shown in Fig. 20.9.

Combining both logic gates declared in Fig. 20.9 the *architecture* of the 2 bit adder can be realized following design approach (a) as presented in Fig. 20.10

Figure 20.10 connects the declared logic ports to execute the function of a binary adder.

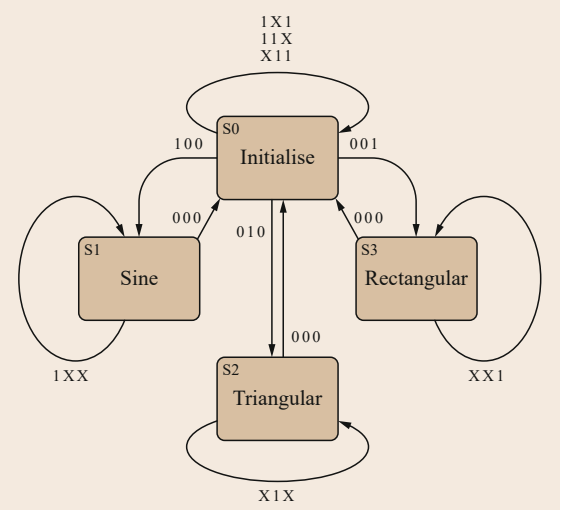


Fig. 20.7 A finite state machine of a simple signal generator. The binary values beside the arrows indicate the position of the three switches. X indicates a *don't care* value, meaning either binary value is possible

20.4.3 Design Flow I: Register Transfer Level (RTL)-Based Designs

Designing a hardware model for a FPGA consists of several steps resulting in a specific design flow. In contrast to the development cycle of *high-level* language implementations there are several significant differences when coding for FPGA hardware. The design flow for hardware models can be divided into six different steps:

1. Design a hardware model in VHDL or by a RTL schematic.
2. Synthesize the model for a certain FPGA hardware chip.
3. Perform a functional simulation of the synthesized system with a VHDL simulation environment.
4. Debug the code using the simulation tool and the synthesis reports. If the model works as expected proceed, otherwise go to step 1.

```
1 entity add2B is
2   port(
3     P : in bit;
4     Q : in bit;
5     S : out bit;
6     C : out bit
7   );
8 end entity add2B;
```

Fig. 20.8 An *entity* declaration for a 2 bit adder in VHDL

```
entity and2 is
   port(
     Pa : in bit;
     Qa : in bit;
     Sa : out bit
);

architecture behavioural of and2 is
begin
   and2Process : process is
   begin
     Sa <= Pa and Qa after 3 ns;
     wait on Pa, Qa;
   end process and2Process;
  end architecture behavioural;

entity xor2 is
   port(
     Px : in bit;
     Qx : in bit;
     Sx : out bit
);

architecture behavioural of xor2 is
begin
   xor2Process : process is
   begin
     Sx <= Px xor Ox after 3 ns;
     wait on Px, Qx;
  end process xor2Process;
end architecture behavioural;
```

Fig. 20.9 An *entity* and *architecture* declaration of an *AND* and *XOR* gate

5. Perform placing and routing of the model for FPGA hardware.
6. Perform a timing simulation. If the timing is as expected proceed, otherwise go to step 4.
7. Create and flash a binary file (bit file) to the specific hardware.

Alongside the functional description of the model, in VHDL source code the place and route process (5) routes external ports to internal buses and vice versa. The declaration of the signal I/O ports is put into a ucf, a *user constraint file*. Here, all input and output signals are routed to the respective hardware addresses of the respective FPGA board.

20.4.4 Design Flow II: Intellectual Property (IP) Core-Based Designs

When following the design process outlined in Fig. 20.11 the efficiency of the resulting hardware model depends to a large extent on the designer's

```
entity add2B is
  port(
     P : in bit;
     Q : in bit;
     S : out bit;
     C : out bit
   );
end entity add2B;

architecture behavioural of add2B is
  component and2 is
    port(
    Pa : in bit;
    Qa : in bit;
    Sa : out bit
   );
  end component and2;

  component xor2 is
   port(
   Px : in bit;
   Qx : in bit;
   Sx : out bit
   );
  end component xor2;

begin
    gateAND : and2
    port map (
        Pa => P,
        Qa => Q,
        Sa => C
        );

    gateXOR : xor2
    port map (
        Px => P,
        Qx => Q,
        Sx => S
        );
end architecture behavioural;
```

Fig. 20.10 An *entity* and *architecture* declaration of a 2 bit adder

knowledge of the FPGA he is designing for. In the early days of FPGA programming, when the devices where smaller and a developer could remember each logic gate by its name, this design approach was state-of-the-art even for *larger* hardware models. Newer device generations have several magnitudes more dedicated logic and additional special functionalities, so this approach is becoming more and more unpractical for larger heterogeneous designs. Therefore modern design approaches rely on core-based implementations.

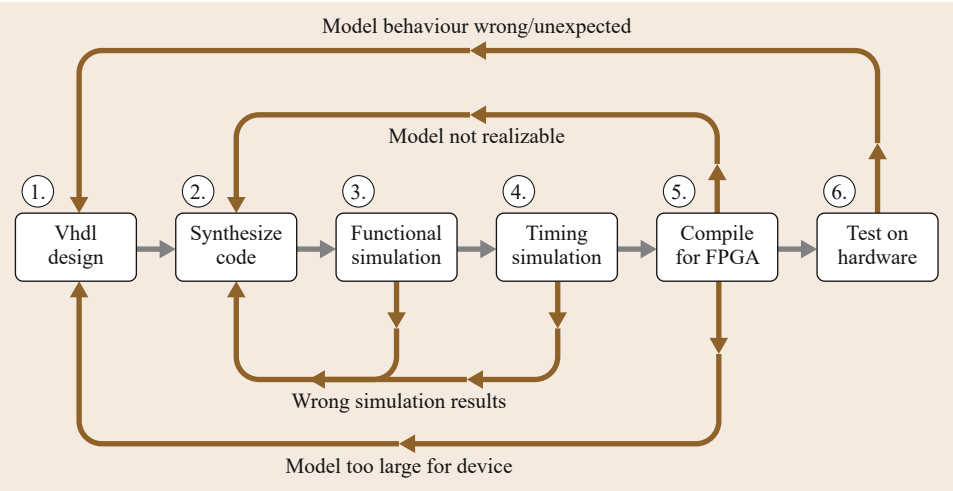


Fig. 20.11 Implementation cycle for FPGA design. The *gray path* indicates successful completion of the preceding step, *brown paths* indicate an error in the respective step

This means, instead of implementing a model from scratch and treating the FPGA as a cluster of freely programmable logic gates, this second design strategy employs reusable HDL blocks called intellectual property (IP) cores, which are *black boxes* accomplishing certain defined functionalities. On a structural level, they are comparable to software libraries of high-level languages, like the Fourier-transform library FFTW (Fastest Fourier Transform in the West) or a linear equation solver library like BLAS (Basic Linear Algebra Subprograms). IP cores facilitate code reutilization and can minimize model design and implementation time, because they are tested and provide defined I/O behavior. There exist IP cores for multiple different problem areas and they are provided by many different vendors. In addition to commercially available IP cores there exist freely available open-source IP cores for various problems. For signal processing applications there are cores like FIR/IIR (finite impulse response/infinite impulse response) filter designs, various transforms like Fourier-transform or Hadamart-transform cores as well as linear algebra cores.

20.5 FPGA Hardware Overview

Alongside a comparable evolution of other digital hardware platforms, the computational throughput and processing power of FPGA chips has evolved drastically over the last 30 years. From the earliest devices, like the Xilinx XC2064-33 containing 64 logic blocks, to modern devices, like the XCVU440, which includes about 4.4 million logic cells and has a throughput in the TMACS range for systolic, symmetric FIR filters using a fixed-point data type, FPGA designs have become a viable solutions in modern high-performance computing (HPC) as well as in data-center applications. This section gives an overview of several FPGA chips and a comparison regarding their applicability to DSP problems. Thereafter, two FPGA hardware development boards are presented with a focus on additional on-board hardware extensions that can be used in DSP design.

20.6 FPGA Chips

In their early years ($\approx$ 1984–1991), FPGA were mainly used as flexible glue logic devices connecting and interfacing logic components in larger, more complex system. Nowadays, FPGA chips are sophisticated systems themselves, which can consist of various hardware extensions, such as distributed arithmetic, soft/hard microprocessor cores, and distributed memory resources, in addition to the basic programmable logic. Thus, a modern FPGA can effectively be seen as a heterogeneous computing platform on one chip.

The structural differences between FPGA chip generations and different chip classes more or less define their general fields of application. For DSP-centric designs there are four logic resources that affect

the throughput of computationally demanding algorithms.

Besides the raw number of logic gates, there are three logic blocks that are of central interest in DSP designs: (1) internal RAM, (2) hard-wired multipliers and (3) DSP logic blocks like the *DSP48* or DSP48e1 slice in Xilinx devices or the *variable precision DSP blocks* in Altera devices. An overview of the availability of these blocks gives a first global view of the capabilities of a FPGA chip for DSP applications.

To this end, an overview of some FPGA devices by the two main vendors, Xilinx and Altera, is given. A focus is put on the processing capabilities of the FPGA devices for DSP-centric applications:

- *Xilinx devices* Table 20.4 gives an overview of several recent and not-so-recent FPGA devices by the manufacturer Xilinx.
- *Xilinx Virtex-II Pro* The Virtex-II Pro was released in 1998 and incorporated several features that were different to other FPGA devices of that time, one being the possibility of programming the chip via an USB (Universal Serial Bus) port from a standard PC with a point-to-point flash protocol JTAG (Joint Tag Action Group). As such, it was the first easily programmable FPGA chip commercially available. The maximal logic resources of a Virtex-II Pro are presented in Table 20.4.
- *Xilinx Virtex-6* A more recent FPGA device, which is part of a design presented in Sect. 20.8, is the Virtex-6 FPGA. It contains multiple integrated hardware components like peripheral component interconnect express (PCIe) blocks, high-speed transceivers and, compared to older devices, enhanced DSP resources. The logic resources available on a Virtex-6 device that are utilized in this work are summarized in Table 20.4.
- *Xilinx Ultrascale+* The Ultrascale+ FPGA is the most powerful Xilinx FPGA device available at the time of writing. As can be seen from Table 20.4 it contains up to 11 904 specialized DSP blocks and a larger amount of high-speed connectible RAM. When running in an optimal design and at maximum clock rates, it is possible to perform ≈ 21 TMACS.
- *Altera Devices* The devices listed in Table 20.5 give an overview of several generations of FPGA devices by Altera. There are only minor differences regarding the nomenclature of FPGA parts of this manufacturer.

  The largest of the Altera devices, the Stratix-10, has a computational throughput of 23 TMACS for fixed-point data and 9.2 TFLOPS (tera floating point operations per second) for a single precision floating-point data type. From Tables 20.4 and 20.5, it can be seen that the number of dedicated DSP slices rises with each device generation, indicating the importance of the high-performance signal processing applicability of FPGA devices. Extrapolating this trend onto future devices and future hardware generations, FPGAs will continue to be viable alternatives for high-performance DSP applications.

20.6.1 Development Boards

In this section an overview of FPGA development boards of different vendors is given. A focus is put on the applicability of the respective board to DSP-centric applications. Hardware development boards exist for most FPGA chips of different generations. They are constructed to test implementations of hardware models on a FPGA including surrounding hardware, because, in addition to the chip itself, there are different peripheral hardware devices available on a development board. The number and type of external hardware on a development board depends on the respective vendor and on the application for which this board is manufactured.

Available in most boards is some sort of on-board flash memory, several user programmable inputs/outputs, a clock circuitry and a communication protocol interface to interact with a FPGA. Hardware development boards include some sort of programming interface to a flash FPGA from a computer or from an internally available flash ROM card. Additional on-board hardware can be high-speed input/output connectors, special clock circuitry, DA/AD converters or high-speed communication protocol ports, a structural overview of a FPGA development board is depicted in Fig. 20.12. The development boards presented here are only a small selection of the vast range of different hardware development test boards and acts as an overview of possible configurations of a hardware system. The Nexys video trainer board is a FPGA development board that includes an on-board 24 bit DA/AD converter. It consists of an Artix-7 FPGA chip extended by multiple hardware components like an USB host controller or an Ethernet connection. It can be programmed via an USB connection and has several freely programmable switches and buttons. A development board of the Virtex-6 family is the *ML-605 evaluation board*, which is a development platform for high-speed communication and signal processing appliances built around a Virtex-6 chip. The on-board features include an PCIe controller, double data rate (DDR) RAM, a small display and several I/O proto-

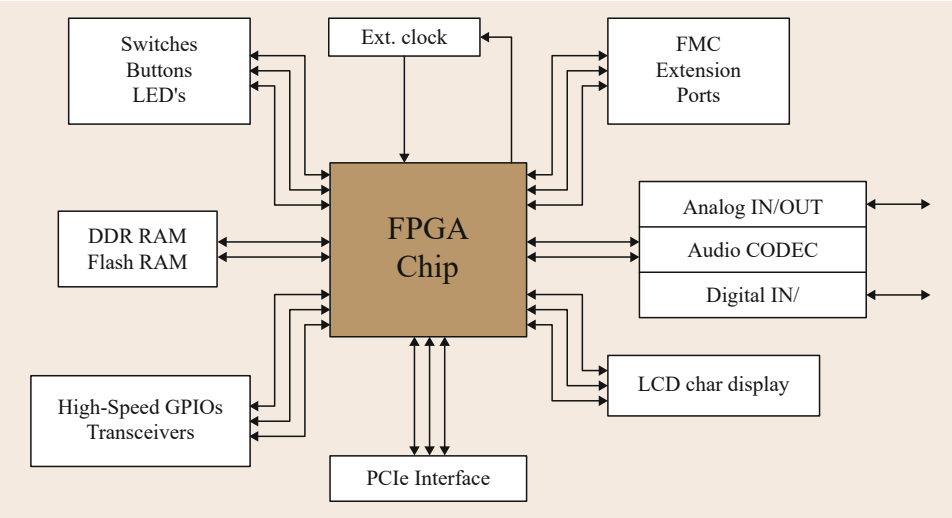


Fig. 20.12 Schematic overview of a FPGA development board

Table 20.6 Overview of several development boards

Name	FPGA device	On-board hardware	Interconnections
Nexys video	Artix-7	24 bit audio codec[a], DDR RAM	USB, ethernet, switches
ML-605	Virtex-6 XC6VLX240T	DDR RAM	FMC Interface, PCIe, ethernet
Profpga quad v7	Up to 4 Virtex-7	Extendable by daughter boards	USB, PCIe, ethernet
Stratix V ASD kit	2 Stratix V	DDR RAM	FMC interface, PCIe 3.0, USB

[a] codec = coder decoder

col ports. It can be extended by FPGA mezzanine card (FMC) daughter boards with other hardware devices, Table 20.6 summarizes features of several development boards.

20.7 Interfacing With a FPGA

The flexibility of FPGA devices is due to their free programmability and their flexible input/output capabilities realized through general purpose I/O ports. Depending on the architecture and the generation of the FPGA hardware, I/O ports are connected to the logic gates of the chip and can be accessed directly from a hardware design. They can either be routed to external devices or to other hardware blocks on a development board. In modern Xilinx devices the I/O ports are called *SelectIO*s and are configurable regarding their output voltage, slew rate and on-chip termination. The I/Os are divided into high-performance and high-range I/Os. The former can be used to interface with high-speed devices like memory or chip-to-chip connections; the latter facilitates a higher voltage range. Modern Xilinx devices facilitate interconnections that transmit data in the gigabytes per second range allowing for connections to PCIe 3.0 devices or 100 Gb LAN (local-area network) standards as well as many others. Other transceiver standards are realizable as well allowing connections to SPI (serial peripheral interface) buses, I2S (inter-IC sound) devices, or SATA (serial advanced technology attachment) connections.

20.7.1 LM4550-AC'97 CODEC

For DSP applications with an analog audio input or output an audio chip must be interfaced with the FPGA hardware. The Texas Instruments LM4550 chip provides an AC'97 protocol-compliant converter. In some development boards it is hardwired to programmable input/output ports of the FPGA and can thus be connected to FPGA logic resources using the serial AC'97 Rev. 2.1 communication protocol. Even though AC'97 protocol-compliant codecs are not state of the art anymore, as they have been replaced by newer protocols like Intel's HD Audio, they are still used in several newer development boards in the lower price segment.

The signals connected to a hardware model consist of:

- SDATA_IN : Signal from the microphone/line input of the converter
- BIT_CLK : Serial data clock
- SYNC : Synchronization bit to synchronize data frames
- SDATA_OUT : Output data sent to converter
- RESET : System-wide reset signal for cold and soft reset.

All signals are processed at a clock speed of 12 288 MHz by the master clock signal BIT_CLOCK. The SYNC signal subdivides the transmission protocol line into frames of 20.8 µs (20.8 µs = 1 /48 000 Hz) length with one tag slot of 16 bit length and 12 data slots of 20 bit length each. A decomposed serial data frame is depicted in Fig. 20.13.

Depending on the direction of the SDATA signal, slot numbers 1 and 2 contain command and status signals respectively. The following two slots contain PCM-coded data. Depending on the transmission mode and the implementation of the AC'97 protocol, the slots named RSV, which stands for reserved, contain PCM, status or control data.

As one can see, the AC'97 protocol makes use of a time-division-multiplex data transfer. The serialization of the parallel data is pipelined in the core module of the AC'97 interface design. The BIT_CLOCK signal is the master clock for the parallelization timing as well as the serialization of the data.

20.7.2 I2S Interface

Another interface protocol for communication between different hardware chips used in many signal processing appliances is the I2S interface protocol. Compared to the AC'97, the I2S protocol is more simplistic having a much smaller protocol overhead and thus is less demanding on hardware resources. Especially in newer audio chips or in in DIY minicomputers like the Raspberry Pi 2, it is used to transmit/receive pulse code modulation (PCM)-coded audio data. The I2S is a protocol specified by Philips semiconductors as depicted in Fig. 20.14.

Compared to the other serial protocols, I2S communication is performed in full duplex using a master/slave architecture. The master is either the data transmitter or the receiver. A schematic overview of a transmitter and receiver interaction is depicted in Fig. 20.14 in the upper left corner with the signal timing at the bottom.

At each device, a transceiver circuit is implemented to route and process the datastream from one device to the other device.

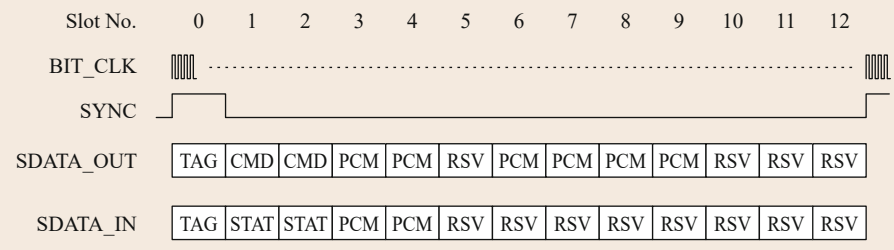


Fig. 20.13 AC'97 protocol signal overview

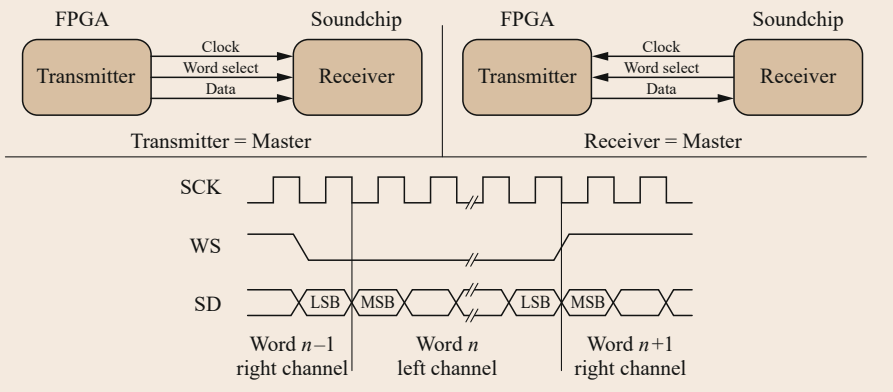


Fig. 20.14 I2S system configuration consisting of one transmitter and one receiver. The timing diagram illustrates the serial data transmission of a stereo signal

20.7.3 PCIe Interface

A commonly used high-speed input/output protocol, which is implemented in many modern FPGA boards in the high-performance range, is a peripheral component interface express (peripheral component interface express) interface. In a DSP-centric application design, the PCIe port can be used to transfer data between a hardware implementation of a DSP algorithm on a FPGA board (device) and a graphical user interface (GUI) running on a personal computer (host). In this section, an introduction to the basic functionalities of the PCIe protocol and a short overview of a example implementation, including a communication protocol, is given.

20.7.4 PCIe Fundamentals

In 2002, the PCI interest group, known as the PCI-Sig consortium, which consists of 900 hardware and software companies, published the first specifications of the PCIe protocol, as an extension to the already-established PCI and PCI-X protocols [20.33]. The basic protocol has undergone several revisions, at the point of writing it has version number 3.1 with 4.0 and 5.0 architectures already published [20.34]. Today, the PCIe interface is a de facto standard interface for high data throughput communication of peripheral devices, exchanging data with the central processing unit in personal computers [20.33]. One of the most notable differences of the PCIe interface compared to the older PCI and PCI-X protocols is the serial structure of the data transfer lanes, instead of the prior parallel structure. Another feature that discerns these protocols is the point-to-point communication of PCIe, enabling the bus to handle higher clock rates without protocol overhead of bus arbitration, found for instance in the original PCI protocol specification [20.35, pp. 134]. The maximum data transfer rates that can be achieved with a PCIe 3.0 interface are $\approx 16\,\mathrm{GB/s}$ [20.33]. In the presented models, the utilized PCIe interface is a version 2.1 revision with a 4x lane interface configuration. The maximal data rate including the 10 b/8 b protocol overhead is approximately $2\,\mathrm{Gb/s}$. Communication with the GUI running on the personal computer is achieved by implementing a Windows RAM driver, writing configuration data to the FPGA board and receiving sound data from the FPGA board.

20.7.5 PCIe Protocol Timing

The timing of the PCIe protocol depends on the type of PCIe connection type and the version. The bandwidth for the PCIe 2.0 implementation used in many development boards has a link speed of approximately $16\,\mathrm{GB/s}$ for a 16x connection. The Xilinx ML605 board for instance has a PCIe 2.0 8x link with a maximal transmission speed of $4\,\mathrm{GB/s}$. Because any real system has additional overhead added by the operating system driver and PCIe protocol overhead, the raw bit rate is mostly smaller. The communication protocol overhead and resulting effective bit rate has to be assessed upfront to decide if the data-rate is sufficient for the implemented hardware design.

20.7.6 Further FPGA Design Considerations

When implementing algorithms in hardware there are several guidelines that can be applied to help decide if a speedup is to be expected by using a programmable hardware platform like a FPGA. A number of practical principles developed for computational biology are published in [20.36]. They are aimed at describing general design patterns so they can be applied to DSP algorithms as well.

Two central questions regarding a hardware implementation of an algorithm are:

(*a*) Can the computations be expressed in a manner that is appropriate for FPGA hardware?
(*b*) Does the algorithm allow for the use of a suitable data type that leads to an optimized form of an algorithm?

Regarding point (a) there are several traits of a computational method that can facilitate speedups on FPGA hardware but the most important questions are: *Can my algorithm be parallelized*?, and *To what extend can it be parallelized*? If your answers are *Yes*! and *A lot*!, only then can you harness the full potential given by the massively parallel hardware facilities of FPGAs. Another question regarding point (a) is whether the algorithm can be expressed without too many conditional branches (for instance *if-else-case* instructions). Nested code cannot easily be transferred to FPGA hardware and it is important to assess if an algorithm can be simplified in this regard.

Regarding point (b), large speedups can be expected if the data can be represented with a data type that enables bit width reductions without losing accuracy. If for instance a fixed-point data type with a small bit width suffices for a certain operation there are data-type specific simplifications that can be used to speed up algebraic operations.

The flexible structure of FPGAs allows for arbitrarily arranged signals that are not bound to the 16/32/64 bit of half, single or double precision of most modern digital hardware that uses a floating-point representation. Another central advantage of FPGA hard-

ware over more closed hardware solutions is the complete control over every part of an implementation and the resulting possibilities of optimization at the core level of an algorithm. This can be of concern when a sample-accurate algorithm is sought after, and the exact predictability of the sample throughput of a system is important. Using timing simulations, a hardware implementation can be tracked with clock cycle accuracy and implemented on hardware, guaranteeing the same accuracy.

20.8 Real-Time DSP Applications

As mentioned in the introduction and the preceding sections, FPGAs have been shown to perform quite well for different kinds of DSP applications. A comprehensive overview of DSP methods on FPGAs in general can be found in [20.37] and in [20.38]. Both publications stress the advantages of FPGAs due to their flexible nature compared to other computing devices but also show the pitfalls when designing DSP applications for hardware. A rule of thumb for hardware implementations of algorithms is: The more an algorithm can be condensed to its core functionality, the more efficient a hardware model is and the faster and more effective it can perform its tasks in hardware implementations [20.39, pp. 20 ff].

When using FPGAs to accelerate numerical calculations, one form of optimizing the calculation towards speed can be achieved by parallelization of the underlying algorithm. This has been proven successful for a wide range of different applications, like real-time noise source identification [20.40], high speed direction-of-arrival algorithms [20.41] or delay-sum beam forming [20.42]. Other works using FPGAs for digital signal processing (DSP) applications are published by [20.43], where two-dimensional/three-dimensional (2-D/3-D) plane wave filters are realized by IIR/FIR filters, or the work of [20.44], who focuses on converting analog controllers to digital controllers using filter-design techniques. Similar to the mentioned work, there are several papers proposing methods of implementing DSP filter designs (IIR/FIR) on a FPGA chip [20.45, 46].

The parallel processing capabilities predestine the FPGA to be used in real-time applications [20.28], as shown for example for particle track recognition [20.47], high-speed cross-correlation [20.48] digital beam forming [20.49] among other publications. Real-time filter applications are shown in [20.50] or [20.51]. Other real-time applications are presented in [20.40, 52] or [20.53]. As was shown in the mentioned works, a considerable speedup could be achieved for numerical methods, some of which could be calculated in real time for the first time by fully utilizing the parallel processing capabilities of modern FPGA chips.

In addition to the mentioned papers, focusing on highly specialized topics of signal processing, there are several works using FPGAs to calculate various acoustical phenomena employing finite difference schemes. Among the earliest publications using a FPGA to solve a two-dimensional wave equation with a finite difference time domain (FDTD) method on a FPGA is the work by [20.54]. A physical model of a string implemented on a FPGA was proposed by [20.55]. Other notable publications regarding numerical calculation of the wave equation using finite difference methods are the works of *Erden Motuk*, for example [20.56] or [20.57]. Here, as well as in his thesis [20.58], Motuk utilizes a FDTD algorithm to solve the two-dimensional wave equation for membranes or plates. Other approaches focusing on synthesizing sound on a FPGA are published in [20.59] or [20.55]. A real-time implementation of finite difference models of acoustic instruments is presented in [20.60] and in more detail in [20.61].

20.9 Real-Time Filtering Applications

A central operation of digital signal processing is the application of some type of filter to some kind of signal. Being closely linked to a linear convolution, filtering techniques are applied in a variety of systems using differing design strategies. Applications that make use of a convolution operation can be found in image processing like blurring or enhancing a digital photo, high-frequency applications like signal reconstruction of noisy wireless data transmission, and filtering frequency content of audio signals in a recording studio. In systematic musicology, filtering operations can range from plain applications like amplifying a certain frequency band in a signal using a narrow-band notch filter, to more complex tasks like modeling the mammalian cochlea as a filter bank.

A vast amount of literature has been published regarding basic topics of filtering methods as well as on specialized topics focusing on hardware implementa-

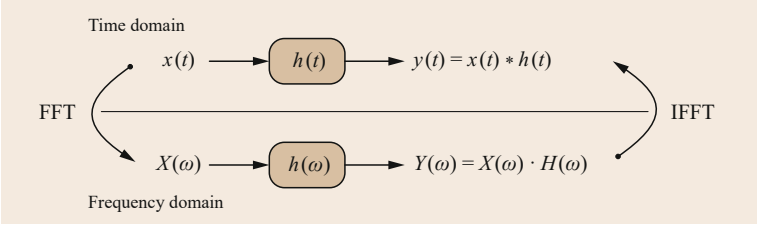


Fig. 20.15 The correspondence between time domain convolution and frequency domain multiplication

tion and acceleration of convolution algorithms. A landmark work, which is often regarded as the hour of birth of modern signal processing, was the rediscovery of the fast Fourier transform (FFT) published by Cooley and Tukey in 1965. The published work revolutionized many digital signal processing methods because of an equivalence, given by the convolution theorem [20.62], which states that a time-domain convolution of two signals can be performed as a multiplication in the frequency domain.

A basic assumption of digital signal processing is that it is possible to describe an analog system as a linear, time-invariant system (LTI system). If we have such an LTI system, the output ($y(t)$) of this system reacting to a certain input ($x(t)$) can be described completely by its impulse response ($h(t)$) in the time domain. In the frequency domain the system is characterized by its transfer function ($H(\omega)$).

Because of this, filtering (convolution) methods can be classified into two main groups, time-domain methods and frequency-domain methods. Modern FPGAs have been shown to perform very well for commonly applied filter designs in the time domain as well as in the frequency domain. In addition to that, the FPGAs' flexible structures and additional on-board DSP resources enable them to perform quite well even for more uncommon filtering approaches like wave digital filters (WDFs) [20.50, 63, 64] or number theoretic transforms (NTTs) [20.37, pp. 401 ff.]. This section tries to elucidate how a modern FPGA can be applied to filtering operations.

20.9.1 Filtering in the Time Domain

The system shown in the upper half of Fig. 20.15 depicts a time-domain convolution operation, where the output signal $y(t)$ is the input $x(t)$ weighted by the systems impulse response $h(t)$. In mathematical terms this is formally defined as

$$y(t) = x(t) * h(t) = \int_{-\infty}^{\infty} h(\tau)x(t-\tau)\delta\tau \; . \qquad (20.2)$$

In the digital domain we are concerned with discrete signals, thus the integral becomes a finite sum. Assumed we have a LTI system, the overall structure remains and can be written as

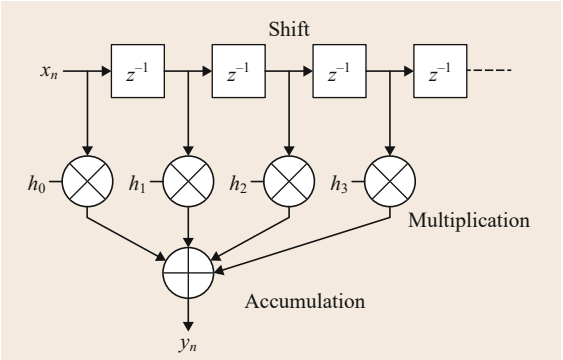


Fig. 20.16 FIR filter design using multipliers, one adder and discrete shifts (z^{-1})

$$y(n) = x(n) * h(n) = \sum_{k=0}^{L-1} h(k)\, x(n-k) \; . \qquad (20.3)$$

Equation 20.3 represents a convolution with a finite impulse response FIR filter of length L. Obviously, the only arithmetic functions needed to perform a convolution are a multiplier and an accumulator, summing the products, see Fig. 20.16 for a schematic block diagram. This procedure is known as multiply-and-accumulate (MAC), a term associated with filtering operations in general.

In addition to this class of filters there exists a second class of filters commonly applied in DSP: infinite impulse response (infinite impulse response) filters. For hardware implementations, both filter classes have certain advantages and disadvantages, which are summarized in Table 20.7.

A literature review on hardware implementations of filter designs shows that most designers prefer FIR filters over IIR filters because of their smaller susceptibility to rounding errors and numerical noise and their easily achievable linear phase. Both DSP blocks presented in Sect. 20.3 are optimized towards high-speed FIR filter designs and are highly adaptable to different filter sizes.

Table 20.7 Comparison of FIR and IIR filter design considerations

Property	FIR filters	IIR filters
Phase	Linear phase is realizable	Normally nonlinear phase
Stability	Nonrecursive FIRs are always stable	Unstable limit cycles are common
Rounding error	High resilience	Very sensitive to rounding errors and numerical noise
Filter size	Large–very large	Small
Coefficient calculation	Complex–very complex	Easy
High-speed designs	Easily realizable	Can be intricate
Pipelined design	Yes	Hard–impossible

FIR Filter Implementation on FPGAs

FIR filters can be implemented on FPGAs using some of the special logic blocks presented in Sect. 20.5. For filter designs, DSP slices can be used to perform additions and multiplications of a filter. To store the coefficients, an on-board RAM should be used. Another recommendation for high-speed filter design on FPGAs is the use of a systolic, i. e., pipelined, filter design, taking symmetry properties of FIR coefficients into account.

Figure 20.17 shows the symmetric characteristic of linear-phase FIR filters. This can be used to minimize

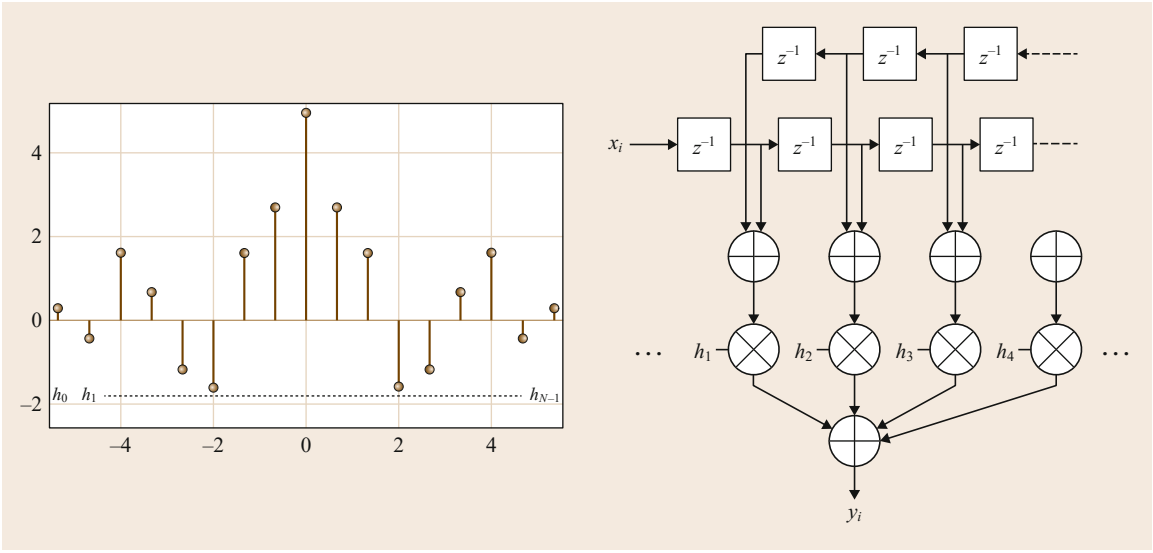


Fig. 20.17 FIR filter design with reduced multipliers making use of the symmetric coefficient structure

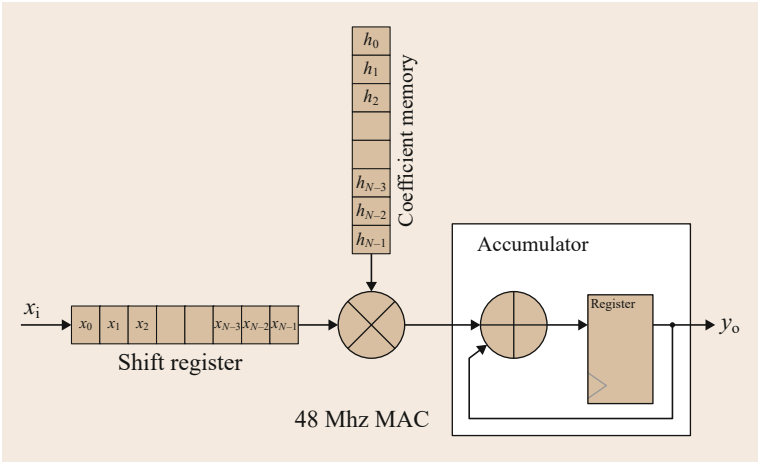


Fig. 20.18 FIR filter that operates at a multiple of the sample rate

the multiplication circuit usage. From the left side of Fig. 20.17, it should be obvious that the filter coefficients are odd symmetric, meaning that $h_0 = h_{N-1}$, $h_1 = h_{N-2}$ and so on. In a Virtex-6 FPGA, this can be realized by using the pre-adder of the DSP48e1 slice (Fig. 20.6). This minimizes the usage of multipliers from $N+1$ to $N/2+1$, with N being the filter order.

Another technique that can be applied to maximize data throughput, especially applicable for filters operating in the audio frequency range, is using a higher clock frequency for the arithmetic operations than the audio sample rate. This is known as oversampling. Take for example an audio signal that is digitized with a sample rate of 48 kHz, then we have a multiplication and accumulation circuit that operates at 48 MHz, which is achievable with most modern FPGAs, and we can process samples with a factor 1000. Figure 20.18 shows a schematic overview of such a system.

20.10 Real-Time Physical Modeling of Large-Scale Geometries

Physical modeling is an established method of simulating acoustic vibrations of musical instruments for sound synthesis purposes [20.65]. One of the large drawbacks of physical modeling methods is the correlation between model accuracy and computational cost. Over the last 30 years a multitude of advances in modeling techniques and in numerical methods in general have paved the way for accurate large-scale models of musical instruments.

Nonetheless, larger models of coupled geometries of musical instruments were not computable in real-time using standard hardware devices and even accelerators like GPGPUs seem to be unpractical for robust real-time synthesis. FPGAs on the other have been shown to be able to compute whole geometry real-time physical models of musical instruments, see for instance [20.60, 66–68] for models of coupled geometries, or [20.57] for a real-time plate physical model on a FPGA.

An introduction to physical modeling for sound synthesis is described in Chap. 19 of this handbook. Other comprehensive considerations regarding general methodologies and techniques can be found in [20.65, 69]. As shown in those works, physical modeling sound synthesis is a robust method for auralizing the vibrations of physically plausible models of musical instruments with high fidelity sound quality. In this section, a real-time implementation of a physical model of an acoustic instrument on a FPGA board is presented. Finite difference (FD) methods are used to discretize the model in the spatial and temporal domain. A focus is put on implementation specifics for FPGA hardware.

20.10.1 Finite Difference Equations of Vibrating Systems

The physical mechanisms that are involved in sound production and propagation in different parts of most musical instruments can be described by partial differential equations (PDEs), particularly the wave equation. Hence, musical instruments can be regarded as systems of coupled PDEs that obey the rules of Newtonian mechanics. As a consequence, all acoustic radiations can be explained by time-varying characteristics of the equations of motion and the interaction of the coupled PDEs. The continuous PDEs that govern Newton's equations of motion for the respective instrument part are discretized using finite difference (FD) operators in the spatial as well as the temporal domain. FD approximations can be derived by using the fundamental theorem of calculus, which states that the derivative of a variable function $u(x)$ along dimension x is defined by taking the limit of a *finite* difference Δx of the dependent variable Δu like

$$u_x = \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} \tag{20.4}$$

with u_x indicating a first derivative by x. For nonzero but small Δx this expression can be utilized to approximate a differential as a difference

$$u_x \approx \delta_x u \,, \tag{20.5}$$

with δ_x indicating a centered first-order difference operator by x.

This generalized finite difference operator notation is applied throughout the remainder of this work. It is based on the notation used in works like [20.65, 70, 71].

A discrete shift operator acting on a one-dimensional (1-D) function $\mathbf{u}$ at position x is indicated by τ with

$$\begin{aligned} \tau_{x+}(u(t,x)) &= u(t, x+\Delta x) \\ \tau_{x-}(u(t,x)) &= u(t, x-\Delta x) \,. \end{aligned} \tag{20.6}$$

A difference approximation in the forward $(+)$ and backward $(-)$ direction at position x can be written as

$$\begin{aligned} \delta_{x+}\mathbf{u}|_x &= \frac{1}{\Delta x}(u(x+\Delta x) - u(x)) \\ &= \frac{1}{\Delta x}(\tau_{x+} - 1)\mathbf{u} \\ \delta_{x-}\mathbf{u}|_x &= \frac{1}{\Delta x}(u(x) - u(x-\Delta x)) \\ &= \frac{1}{\Delta x}(1 - \tau_{x-})\mathbf{u} \,. \end{aligned} \tag{20.7}$$

The same can be done in the temporal domain by defining the forward (+) and backward (−) approximation at time instant t as

$$\delta_{t+}\mathbf{u}|_t = \frac{1}{\Delta t}(u(t+\Delta t)-u(t)) = \frac{1}{\Delta t}(\tau_{t+}-1)\mathbf{u}$$
$$\delta_{t-}\mathbf{u}|_t = \frac{1}{\Delta t}(u(t)-u(t-\Delta t)) = \frac{1}{\Delta t}(1-\tau_{t-})\mathbf{u}\,. \tag{20.8}$$

An interesting feature of this operator notation is that higher-order approximations can be achieved by a convolution of lower-order operators. Using (20.7), a second-order centered finite difference operator can be computed by

$$\begin{aligned}\delta_{xx} &= \delta_{x-} * \delta_{x+}\\ &= \left[\frac{1}{\Delta x}(1-\tau_{-1})\right] * \left[\frac{1}{\Delta x}(\tau_{+}-1)\right]\\ &= \frac{1}{\Delta x^2}(\tau_{+}-1-1+\tau_{-})\\ &= \frac{1}{\Delta x^2}(\tau_{-}-2+\tau_{+})\end{aligned} \tag{20.9}$$

with the equivalence $\tau_+\,\tau_- = 1$.

Higher-order operators can be calculated similarly by

$$\delta_{4x} = \delta_{xx} * \delta_{xx}\,. \tag{20.10}$$

Time Integration

The temporal integration of the FD scheme for the presented FPGA implementation is realized with the Newton–Störmer–Verlet integration scheme [20.72]. This scheme was successfully utilized for nonreal-time FD models of a classical guitar [20.73], a violin [20.74], and real-time models like the model of a simplified banjo [20.67] and a whole geometry banjo [20.61]. The dependent variables of this scheme are the displacement u and the velocity v in solid bodies like strings, membranes and soundboards. When calculating air volumes the pressure p, and its first temporal derivative p_t are used. The PDEs used in this treatise conform to the Stäckel condition, meaning that they are separable, hence reducible to systems of coupled DEs of lower order. This can be achieved by introducing the intermediate variable $\mathbf{v}$; the global velocities. Using this it is possible to write the second-order wave equation as

$$\begin{aligned}\dot{\mathbf{v}} &= c^2\,\mathbf{u}_{xx}\\ \dot{\mathbf{u}} &= \mathbf{v}\end{aligned} \tag{20.11}$$

with the wave velocity in the medium c; a material-dependent constant in the linear case. Applying first-order FD operators to (20.12), the discrete scheme for the wave equation can be written as

$$\begin{aligned}\delta_{t+}\mathbf{v} &= c^2\,\delta_{xx}\mathbf{u}\\ \delta_{t+}\mathbf{u} &= \mathbf{v}\,.\end{aligned} \tag{20.12}$$

As is shown in [20.72, 75–77] this form of a simple Newton integrator is symplectic, meaning that the numerical flow is true to the physical flow of the solution. A necessary stability conditions of scheme (20.12) is the Courant–Friedrichs–Levy (CFL) condition [20.78]. For general dimension D it can be written as

$$\Delta t \sum_{i=1}^{D} \frac{c_i}{\Delta x_i} \le \lambda \tag{20.13}$$

with Δt, Δx the discrete temporal and spatial steps, c the velocity of propagation in the medium and λ the CFL number depending on the respective scheme. The basic FD scheme (20.12) for solving the wave equation without dispersion or dissipation has a CFL number $\lambda = 1$. This stability condition changes if damping or higher-order terms, which model stiffness, are included in the formulation. When this stability condition is satisfied, the basic Newton–Störmer–Verlet [20.79] scheme needs no further correction algorithm, unlike the Newton–Raphson alpha method [20.80] for instance, which adds artificial damping to the system to ensure stability.

20.10.2 Discrete FD Operators

An extension to the well-established FD operator notation, used at least since the early 20th century, is presented in this section. The operator notation allows us to abstract several mathematical operations into a simpler notation. In the following, this concept is extended to an even lower abstraction level by resolving the underlying mathematical operations to the specific operations depending on the data type and underlying hardware structure. Assuming a fixed-point data type, a centered finite difference operator can be expressed using

$$\hat{\delta}_x = T_\Delta[\hat{\epsilon}_{\Delta x+}, -\hat{\epsilon}_{\Delta x-}] \tag{20.14}$$

with $\hat{\epsilon}_{\Delta x+/\Delta x-}$ = a read operation from a register, a finite difference cell right (+) or left (-) of the actual cell and T_Δ a multiplicand that depends on the stride of the discrete grid in the spatial domain. A second-order centered FD operator in vector notation can be written as

$$\hat{\delta}_{xx} = T_\Delta[(\hat{\epsilon}_{\Delta x-}), (<1), (\hat{\epsilon}_{\Delta x+})] \tag{20.15}$$

with $T_\Delta = 1/\Delta x$ and (< 1) indicating a shift operation. This shift operation can be used to replace a multiplication by two in fixed-point arithmetic. A higher-order digital FD operator used for the fourth-order differential equation of the beam can be constructed by a convolution of two second-order digital FD operators

$$\hat{\delta}_{4x} = \hat{\delta}_{xx} * \hat{\delta}_{xx} \,. \tag{20.16}$$

This can be extended to higher-spatial-order difference operators leading to a specific numbers of digital operations for the respective operator given in Table 20.8.

Using the values given in Table 20.2, the used resources of a model must be approximated before implementation in hardware.

20.10.3 Finite Difference Physical Modeling on FPGAs

To gain benefit from the structure of FPGA hardware, the FD models introduced here are formulated in a maximal parallel way. As is shown in the derivation of FD methods, central parts of the core computations can be processed in parallel. This is achieved by dividing the spatial domain into parallel node points on a regular grid. The specific configuration of the discrete points of the respective domain enables a concurrent update of every node for every sampling time step. In addition to this parallel structure, the locality of the memory-access patterns are optimized by using spatially local block-RAM (BRAM) registers for every node point. This minimizes memory write/read overhead.

Scheme (20.12) can be split into sequential parts as well as parallel parts. The sequential parts are the calculation of the velocities and the deflections, which are interdependent and thus must be computed serially. On a given geometry, the update of the velocity and the deflection is computed for every discrete point at concurrent time instances. This means that the serial calculations of (20.12) can be done in parallel for every discrete node on the discrete geometry. Figure 20.19 shows an example of a parallel computation of three adjacent points on a string, at positions $k-1$, k and $k+1$ respectively.

Table 20.8 Digital operations for FD operators used in this work

Operator	Reg. Op.	Shift Op.	Mult.	Add./Sub.
$\hat{\delta}_x$	2	0	1	2
$\hat{\delta}_{xx}$	3	1	1	2
$\hat{\delta}_{4x}$	5	4	1	5
$\hat{\delta}_{2x2y}$	5	1	1	4
$\hat{\delta}_{\nabla^4}$	13	10	1	12
$\hat{\delta}_{2x2y2z}$	7	2	1	6

20.10.4 Serial-Parallel Implementation

When implementing more than one geometry model on a FPGA, or when a geometry is so large that it can't be implemented on one chip in parallel, it is possible to calculate parallel kernels shown in Fig. 20.19 several times. The maximum number of computable kernels depends on the sampling rate and on the clock rate of the FPGA design. The design in Fig. 20.19 is a segment of a 1-D string consisting of 10 parallel kernels. Figure 20.20 depicts an exemplary implementation of 10 parallel kernels getting evaluated eight successive times. In the *LOAD* block, all parallel kernels of slice 1 load the computed values of time step t, in the middle block the values for time step $t+1$ are being computed, in the *WRITE* block all computed values are written to the slice 1 memory position. This leads to the model of a string that consists of $8 \times 10 = 80$ grid points.

20.10.5 Model Routing on FPGA

For larger instrument designs consisting of several geometries it is a good design practice to subdivide the FD computations into several self-contained parts.

Using this strategy, it is possible to structure a model on a FPGA comparable to a design of several coupled IP cores. For a proto-lute made of one string coupled to a wooden box, one part could consist of the string calculations, the other could be the wooden box. Each of the two cores could be subdivided as well into smaller units, but their input and output signals would be defined signal ports using a specified data protocol.

By this it is possible to use and reuse parts of a design, which run correctly in a later design without a complete redesign.

As presented in [20.61] this structured layer approach can be used to realize full geometries of musical instruments. An exemplary part of the layer system is shown in Fig. 20.21. Therein several geometry parts are connected via defined output and input signals and can be controlled and modified by changing certain control parameters.

Interfacing With a Model

The interface layer (IL) is the topmost layer in the model. Here, all external data input and output is managed, decoded and routed. It is the interaction layer with all input and output devices, such as the PCIe interface,

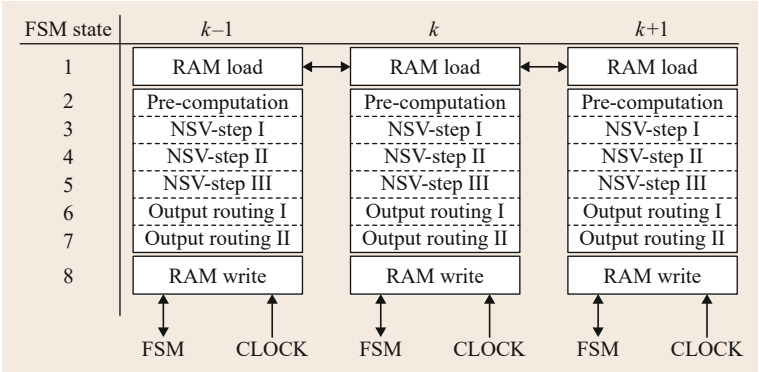


Fig. 20.19 Parallel implementation of sequential NSV-scheme parts

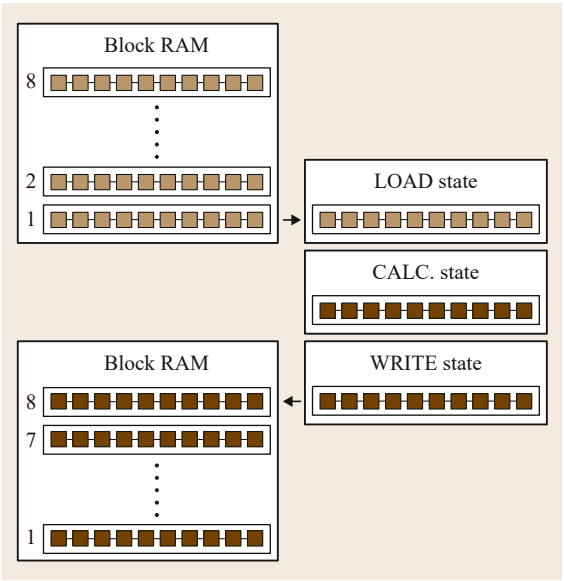


Fig. 20.20 Serial–parallel implementation of sequential NSV-scheme parts. Slices 1-8 consist of 10 parallel computation kernels each writing-to and reading-from their own memory address

a AC97 interface or a I2S interface for data transport. A block diagram of the IL is shown in Fig. 20.22.

Depending on the needs and given ports, either a PCIe, AC97 or the I2S interface can be used for communication and data transfer. In the IL, data from the respective transmitter is decoded and then routed to the instrument model, the model routing layer (MoR). In the output stage of the IL, the sound and status data is encoded depending on the respective receiver protocol.

20.10.6 Real-Time Physical Model of a Violin

In this section, an implementation of a violin on FPGA hardware is developed. It gives an overview of the mathematical models used for the singular parts as well as a detailed insight into characteristic implementation details. The real-time physical model of the violin consists of the frontplate and backplate of the violin. The enclosed air with two orifices, the wooden bridge, four strings and a model for the bow-string interaction complete the proposed model. The model can be controlled from a PC via a PCIe interface and interacted with by using a Leap motion controller.

20.10.7 Violin Research History

The *classical* violin is an instrument with a history in musicological research regarding the acoustic properties of the instrument as well as its organologic development. An introduction to the historic evolution of the modern violin is given for instance in [20.81]. A publication concerned with the early history of the violin can be found in [20.82]. A history starting from the time the modern violin was discernible as a distinct instrument and the development of the violin up to modern times is published in [20.83]. Early experimental studies on the acoustic properties of the violin were performed by Felix Savart [20.84] who focused his research on the vibrations of the front- and backplate and their different acoustic behavior, proposing that *good* violins have a special relation between their *eigenmodes*. He suggested that the sound post inside the violin stiffens the treble side of the violin, transforming the rocking motion of the bridge, which excites a dipole mode of the top plate into a monopole mode by this unilateral stiffening of the geometry [20.85, pp. 165–166]. His research led him to design a trapezoid-shaped violin. Following his publications there were several important publications regarding violin acoustics during the 19th century. One important work was published by Hermann von Helmholtz, who is credited with the first description of the stick-slip mechanism of the bow-string interaction. The theories of Helmholtz were further extended by Lord Rayleigh in [20.84]; findings

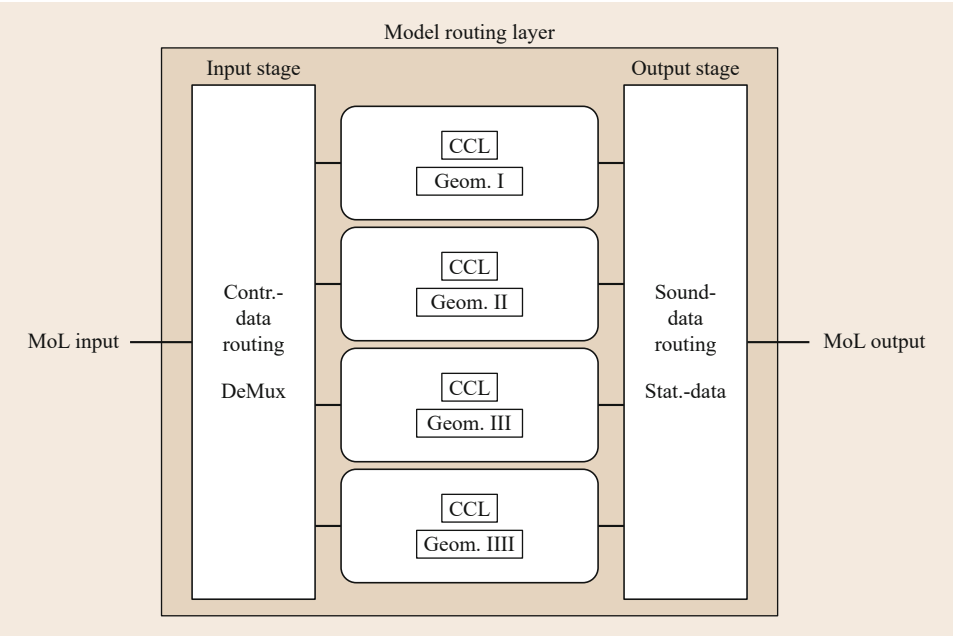


Fig. 20.21 Model routing layer overview

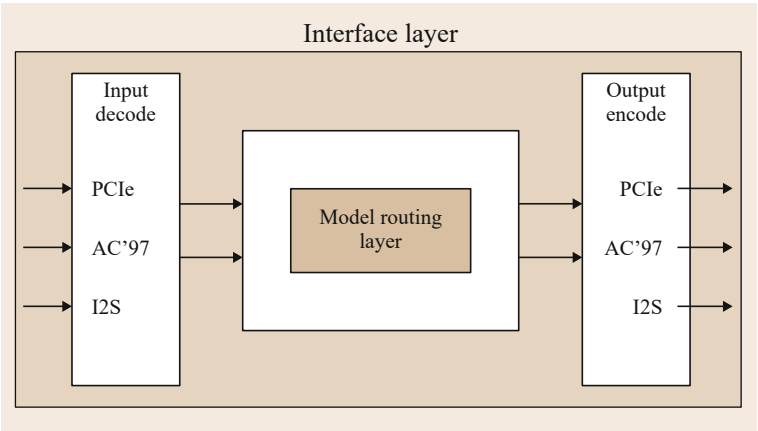


Fig. 20.22 Block diagram of the interface layer

were published by Lord Rayleigh and *Hermann von Helmholtz* [20.86]. At the beginning of the 20th century, Raman published several works concerned with violin acoustics and especially the bow-string interaction. An encompassing discussions about the structural mechanics and acoustic interactions of the violin can be found in [20.87].

Typical radiations of the f-holes and the interaction between lower body modes and the air volume are discussed in a technical manner for the whole class of the Hutchins–Schelleng violin octet in [20.88, 89], for a list of instruments from the Hutchins–Schelleng octet see Table 20.2 in [20.90, p. 325]. *C. Hutchins* followed the work of Savart and suggested that if the top plate *eigen*frequencies were slightly higher than the ones of the back plate, the violin sounds brighter; otherwise it sounds duller [20.91].

J.-L. Florens [20.92] presents a hardware design for interacting with a virtual model of a violin bow showing realistic simulation results. A virtual interaction model with digitized gesture data of bowing is presented in the work of *Matthias Demoucron* [20.93]. Based on this work *Esteban Maestre* [20.94] shows an implementation of a similar methodology with a high-level front end.

20.10.8 Violin String Model

The primary part of the violin's sound production mechanism are its four strings. Most modern violin strings

are made of a synthetic core wrapped with metal or some kind of alloy, like aluminum or titanium. The exact physical properties of those materials vary between string manufacturer and according to the desired sound characteristics.

The internal damping of modern violin strings is designed to facilitate the production of a stable Helmholtz motion under different playing conditions. Thus, most violin strings have a considerably higher internal damping compared to other string instrument strings. In comparison to a guitar string, this results in a less pronounced frequency content in the higher range of the spectrum. A thorough review on violin string research is published in [20.95].

In this treatise the strings are modeled as stiff strings with losses at the boundaries as well as internal damping. The motion is modeled for one transverse polarization leaving longitudinal and torsional effects out of the consideration, as shown in [20.96] longitudinal and torsional vibrations are measurable in cello and contrabass string vibration.

The equation of motion for the strings in one vertical polarization for $x \in 0, \ldots, L$ can be written as

$$\begin{aligned} \mathbf{u}_{tt} &= c^2\mathbf{u}_{xx} - B\mathbf{u}_{4x} - \beta\mathbf{u}_t + \alpha\mathbf{u}_{txx} \\ u(t,0) &= 0; \quad u(t,0)_x = 0 \\ u(t,L) &= 0; \quad u(t,L)_{xx} = 0 \end{aligned} \tag{20.17}$$

with $c = \sqrt{T/\mu}$, $T =$ tension and $\mu =$ linear mass density. The term $\beta\mathbf{u}_t$ is a viscous damping term, and the mixed derivative $\alpha\mathbf{u}_{xxt}$ is a phenomenologically justified damping term modeling the internal damping. A similar term was first proposed in [20.97] and extended to the form used here in [20.98]. The damping constants α and β are two heuristically approximated damping parameters. Due to the fact that all strings of the violin are metal strings with a considerable thickness they have a finite elasticity. This effect, the stiffness, is modeled by including a bar-like term $B\mathbf{u}_{4x}$. The factor B is defined as $B = ESK^2/\mu$, with Young's modulus E, cross-sectional area S and K the radius of gyration.

This adds a bar-like characteristic to the string and results in a stretching of partials, which is small in the case of thin strings [20.99]. Because we have a fourth-order term in the equation, we need four boundary conditions (BC): two at each side. On the side of the bridge ($x = 0$) we use clamped BC and fixed BC at the nut ($x = L$).

Separating (20.17) by introducing the velocity v yields

$$\begin{aligned} \mathbf{v}_t &= c^2\mathbf{u}_{xx} - \kappa^2\mathbf{u}_{4x} - \beta\mathbf{u}_t + \alpha\mathbf{u}_{txx} \\ \mathbf{u}_t &= \mathbf{v}\,. \end{aligned} \tag{20.18}$$

Discretized Equation

The separated and discretized scheme (20.18) can be written as

$$\begin{aligned} \hat{\delta}_{t+}\mathbf{v} &= (c^2\hat{\delta}_{xx} - \kappa^2\hat{\delta}_{4x} - \beta\hat{\delta}_{t-} + \alpha\hat{\delta}_{xxt-})\mathbf{u} \\ \hat{\delta}_{t+}\mathbf{u} &= \mathbf{v}\,. \end{aligned} \tag{20.19}$$

The hardware operations of the discrete difference operators can be added up to approximate the hardware utilization of one string. The multiplicands in front of the discrete difference operators can be combined with the multiplicand of the operator shown in (20.20). The requirements for one discrete point of the string are given in Table 20.3.

VHDL Model of the Strings

One string of the violin model is discretized with 80 points on the FPGA. Using the approach presented in Sect. 20.10.4, each string can be computed using 10 parallel computation kernels, where each of these kernels is computed eight times.

Summarizing the requirements for the strings, the values in Table 20.9 are multiplied by 40 (four strings with 10 kernels), which yields the values given in Table 20.10.

20.10.9 String–Bow Interaction

The excitation of the violin is implemented as a bow–string interaction model. The physical properties of the string–bow interaction are modeled based on a stick–slip interaction as presented in [20.100]. The system can be formulated using an iterative approach, modeling the forces acting at the contact point of bow and string. The fundamental idea behind the implemented

Table 20.9 Operator count for a damped one-dimensional string

Operation	Reg. Ops.	Shift	Mult.	Ad./Su.
$c^2\hat{\delta}_{xx}u$	3	1	1	2
$\kappa^2\hat{\delta}_{4x}u$	5	4	1	5
$\beta\hat{\delta}_{t-}u$	2	5[a]	1	8
$\alpha\hat{\delta}_{xx(t-)}u$	2	5[a]	1	8
Sum	12	15	4	23

[a] The damping parameters are approximated as shift operations

Table 20.10 Hardware requirements for four strings consisting of 10 computation kernels

Operation	Reg. Ops.	Shift	Mult.	Ad./Su.
Sum:	480	600	120	920

stick–slip model, and most other friction models as published in [20.95], is the assumption that the bow sticks to the string, tearing the string in the direction of the bow motion. Hence, the contact-point velocity of the string equals the velocity of the bow during contact. If the net force of the string acting in the opposite direction is stronger than the force exerted by the bow, the string starts to slip off the bow. The force of the string can either be triggered by the deflection-dependent net force or by the pulses of the *Helmholtz* motion traveling on the string, catapulting the string off the bow. A schematic state diagram is given in Fig. 20.23.

The input signals starting the iteration are loaded at the beginning of the loop and the state changes depending on the truth-value of the respective conditions. This leads to an interactive playability of the model enabling a musician to expressively play the virtual violin. Several videos of this interaction model can be found at http://www.systematicmusicology.de. A flowchart of the excitation model is shown in Fig. 20.24.

String–Bow Model on a FPGA

What makes the string–bow model on a FPGA especially interesting is that the throughput of the systems has only a latency of several FPGA clock cycles. Therefore, a change in pressure or velocity of the bow is directly transmitted to the respective string that is selected. The selection of a string can be toggled by setting a *bow-contact* bit to **1**.

Figure 20.25 shows a schematic timing diagram of a bow contact on a selected string.

20.10.10 Bridge

The violin bridge has two feet, which transmit the vibrational energy of the string to the frontplate of the instrument. The geometry and physical properties of the violin bridge leads to distinct *eigen*vibrations, which adds a resonance peak in the radiated spectrum known as the *bridge hill*. In most violins it can be found around a frequency of ≈ 3 kHz [20.101].

Under realistic playing conditions, the bridge of the violin acts as a bidirectional transmitter of the acoustic vibrations between the strings and the frontplate. Its two feet exert a time-varying force on the frontplate and add a time-varying boundary condition for the strings. It is modeled as a two-dimensional plate with forces acting in the in-plane direction denoted as u and w, with u being perpendicular to the frontplate of the violin.

The differential equation for the respective directions x and y can be written as

$$\begin{aligned}\mathbf{u}_{tt} &= c^2(\mathbf{u}_{xx} + \nu_{1L}\mathbf{u}_{yy} + \nu_{2R}\mathbf{w}_{xy}) - \beta\mathbf{u}_t \\ \mathbf{w}_{tt} &= c^2(\mathbf{w}_{xx} + \nu_{1R}\mathbf{w}_{yy} + \nu_{2L}\mathbf{u}_{xy}) - \beta\mathbf{w}_t \end{aligned} \quad (20.20)$$

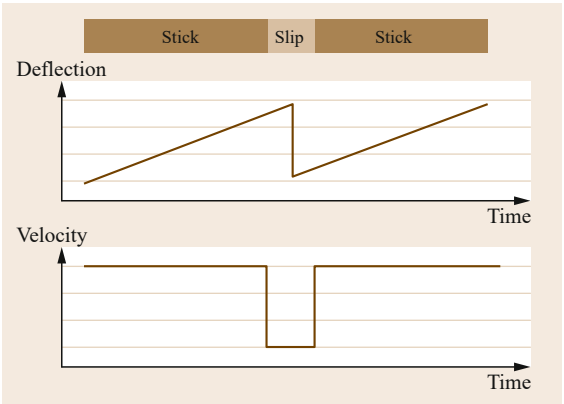


Fig. 20.23 A state representation of the bow model exciting the violin string

with material-dependent parameters ν_*, c^2 given below, and damping parameter β.

Discretized Equation

Using the discrete difference operator notation and separating them, Equations (20.20) can be rewritten as

$$\begin{aligned}\hat{\delta}_{t+}\mathbf{v}_x &= [c^2(\hat{\delta}_{xx} + \nu_{1L}\hat{\delta}_{yy}) - \beta\hat{\delta}_{t-}]\mathbf{u} \\ &\quad + \nu_{2R}\, c^2\, \hat{\delta}_{xy}\mathbf{w} \\ \hat{\delta}_{t+}\mathbf{u} &= \mathbf{v}_x \\ \hat{\delta}_{t+}\mathbf{v}_y &= [c^2(\hat{\delta}_{yy} + \nu_{1R}\hat{\delta}_{xx}) - \beta\hat{\delta}_{t-}]\mathbf{w} \\ &\quad + \nu_{2L}c^2\hat{\delta}_{xy}\mathbf{u} \\ \hat{\delta}_{t+}\mathbf{w} &= \mathbf{v}_y \end{aligned} \quad (20.21)$$

with v_y, v_x the velocities in the x and y directions respectively.

Coupling of the Strings to the Bridge

The coupling of the string vibration to the bridge can be formulated by taking the acting forces at the interaction point *ip* into consideration

$$F_{ip} = B\, u_{3x}^{\mathrm{str}} + T\, u_x^{\mathrm{str}} \quad (20.22)$$

with B the bending stiffness as defined before and T the tension of the string.

20.10.11 Top Plate/Back Plate

The body of the violin is made of wood. In most instruments, the frontplate of the violin is made of spruce whereas the backplate is made of maple. The frontplate and the backplate are held together by ribs around the rim of the instrument. In addition, the sound post connects both plates. The characteristic f-holes on the

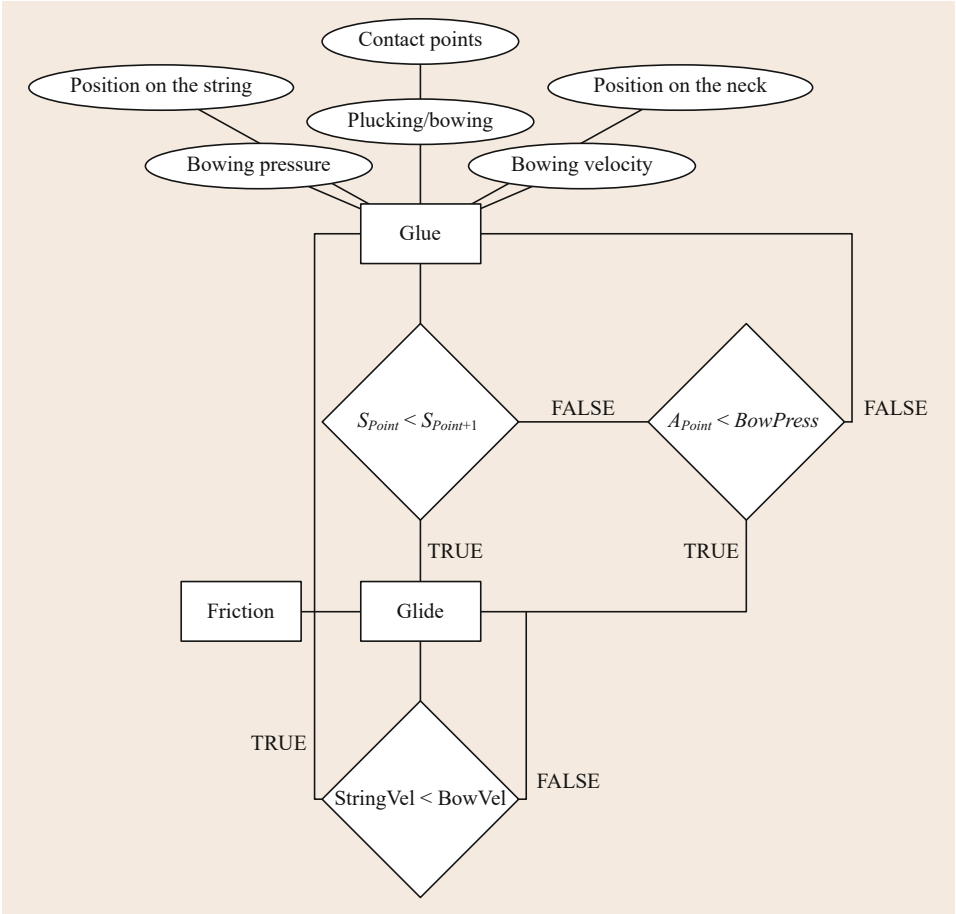


Fig. 20.24 A flow chart representation of the bow model exciting the violin string

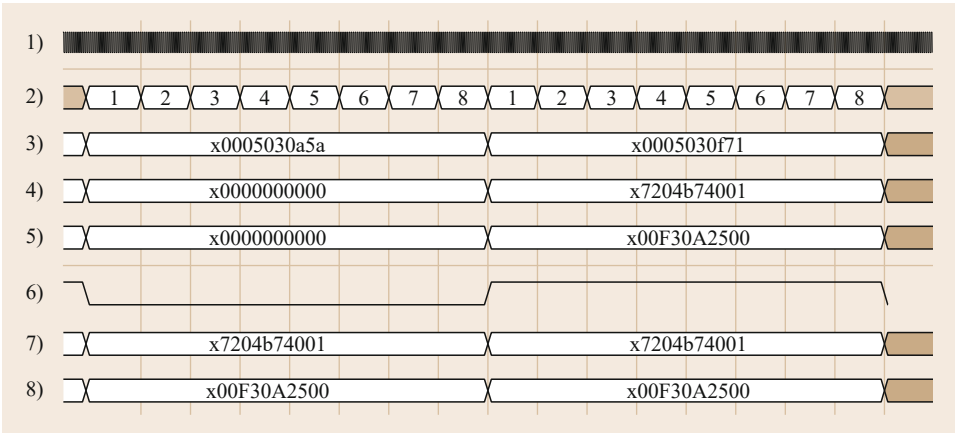


Fig. 20.25 A timing diagram of selected bow–string signals. When the *bow-contact* bit is set, the pressure and the velocity values of the bow interact with the string at the selected bowing point. 1) FPGA clock. 2) Serial/parallel string calculation state. 3) String deflection. 4) String–bow model contact pressure. 5) String–bow model contact velocity. 6) Bow contact. 7) Bow pressure. 8) Bow velocity

frontplate of the instrument give rise to a resonating enclosed air volume. Both plates are arched, which is commonly believed to increase the stiffness, which in turn changes the frequency of the plates without having to change the mass or the thickness of the plates. The arching of the plates can be modeled by using a PDE for a buckled plate.

The top plate and the backplate of the violin are modeled as orthotropic wood plates using the discrete version of a fourth-order differential equation.

This leads to the following PDE of a wooden plate including stiffness, damping and buckling

$$\mathbf{u}_{tt} = \tilde{c}^2[\mathbf{u}_{xx} + \mathbf{u}_{yy}] - \kappa \mathbf{u}_{\nabla^4} - \beta \mathbf{u}_t + \alpha \mathbf{u}_{xxt} - \mathbf{F}_{\text{ext}} \tag{20.23}$$

with $\tilde{c} = \sqrt{T(x,y)/\sigma}, T(x,y)$ the variable membrane tension, σ the area density, $\mathbf{u}$ the membrane displacement normal to the membrane plane and external forces $\mathbf{F}_{\text{ext}} = \mathbf{F}_{\text{air}} + \mathbf{F}_{\text{br}}$, where $\mathbf{F}_{\text{air}}$ and $\mathbf{F}_{\text{br}}$ are the forces exerted by the air volume and by the bridge respectively. Here, $\kappa = E h^2/12\,\rho(1-\nu^2)$ with Young's modulus E, h is the thickness, ρ is the density and ν is Poisson's ratio.

The biharmonic operator is defined as

$$\mathbf{u}_{\nabla^4} = (\mathbf{u}_{4x} + 2\mathbf{u}_{2x2y} + \mathbf{u}_{4y})\,. \tag{20.24}$$

Discretized Equation

The calculation of the plate deflection is partitioned into two substeps. First the bending moments M are calculated, with M the acceleration on the surface of the plate. The separated and discretized counterpart of (20.23) using a digital operator notation can be written as

$$\begin{aligned} \hat{\delta}_{t+}\mathbf{v} &= \left(c^{\hat{2}}\delta_{xxyy}c + \alpha\hat{\delta}_{xxyyt-} - \kappa\hat{\delta}_{\nabla^4} - \beta\hat{\delta}_{t-}\right)\mathbf{u} \\ &\quad + \mathbf{F}_{\text{ext}} \\ \hat{\delta}_{t+}\mathbf{u} &= \mathbf{v} \end{aligned} \tag{20.25}$$

with $\hat{\delta}_{\nabla^4}$ the discrete, digital biharmonic operator.

Coupling of the Bridge Feet and the Frontplate

The coupling between the bridge feet and the frontplate is modeled by constituting a spring with constant $k_{\text{br/fp}}$ between the bridge and the membrane leading to the coupled equation at the respective points.

$$u^p_{tt} = \frac{1}{m_{\text{fp}}}\mathbf{F}(u^b) - k_{\text{br/fp}}u^{\text{fp}} \tag{20.26}$$

$$u^m_{tt} = \frac{1}{m_{m_{\text{fp}}}}\mathbf{F}(u^m) - k_{\text{br/fp}}u^b\,, \tag{20.27}$$

with $\mathbf{F}$ the force function of the uncoupled part of the respective geometry and m_* the mass of the bridge (br) and the frontplate (fp) at the interaction area.

20.10.12 Model of the Air Cavity

An essential part of the violin sound is the presence of air inside the instrument's body. There are several important air modes of the instrument that can be found in the radiated spectrum of the instrument. The f-holes of the violin are modeled by the using Sommerfeld radiation conditions at the respective points [20.70, p. 50]. Using Helmholtz' equation for pressure $\mathbf{p}$ in air, the PDE can be written as

$$\mathbf{p}_{tt} = \frac{1}{K\rho}(\mathbf{p}_{xx} + \mathbf{p}_{yy} + \mathbf{p}_{zz}) - \beta\mathbf{p}_t \tag{20.28}$$

with pressure $\mathbf{p}$, air compressibility K, and density ρ, which are assumed constant inside the domain. Again, β is a heuristically approximated damping coefficient modeling the viscosity of air.

Coupling Between Air and the Body

The coupling between the air volume and the vibrating violin body is implemented by considering the continuity of normal velocity on the inside of the violin body leading to the conditions

$$\mathbf{p}_z(x,y,0,t) = \rho\mathbf{v}_t^{\text{frpl}}(x,y,t) \tag{20.29}$$

for the zero plane of the air volume in z direction and

$$\mathbf{p}_z(x,y,H,t) = -\rho\mathbf{v}_t^{\text{bpl}}(x,y,t) \tag{20.30}$$

for the air layer in contact with the backplate.

Discretized Equation

By setting $\mathbf{v} = \hat{\delta}_{t+}\mathbf{p}$ we can write the discretized and separated equation as

$$\begin{aligned} \hat{\delta}_{t+}\mathbf{v} &= \left(\frac{1}{K\rho}(\hat{\delta}_{xx} + \hat{\delta}_{yy} + \hat{\delta}_{zz}) - \beta\hat{\delta}_{t-}\right)\mathbf{p} \\ \hat{\delta}_{t+}\mathbf{p} &= \mathbf{v}\,. \end{aligned} \tag{20.31}$$

Note that $\mathbf{v}$ must not be confused with the particle velocity formulated in (20.11).

20.10.13 Application Example

In the following application example, the violin model presented in this section is running on a ML605 FPGA development board connected to a standard PC via the

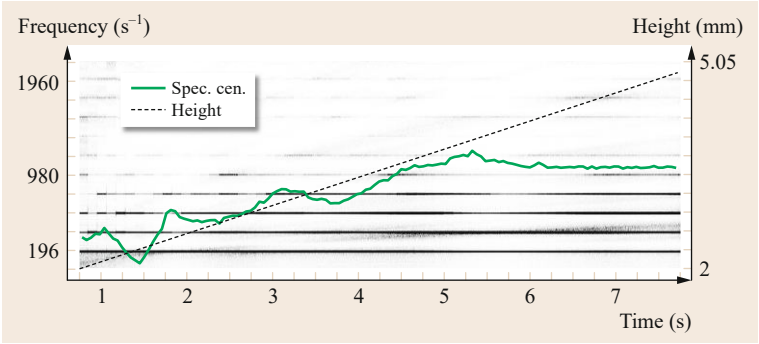


Fig. 20.26 Simulation output of increasing frontplate height in successive steps of 0.05 mm of a violin excited by a constant bowing force and velocity. In *black* and *white* is the spectrogram of the sound, in *green* the spectral centroid

PCIe bus. It is controlled by an application running in C# and C++. The model is excited with steady bow force and velocity, interacting with one single string. While a stable bowing sound of the violin is produced, the height of the frontplate is increased in 0.05 mm steps. The resulting sound is picked up at several positions in the air volume above the frontplate of the violin and transfered to the PC-host application via PCIe.

Figure 20.26 shows the influence of the changing frontplate thickness on the spectral centroid of the radiated sound. From psychoacoustic considerations it is known that the spectral centroid is a robust indicator for the perceived brightness of a sound.

As one can see from Fig. 20.26, the perceived brightness does not change linearly by increasing the thickness of the violin frontplate linearly but depends on the salient interaction between the string, bridge and frontplate.

This is one example of how a real-time FD model can be used by researchers or instrument makers who want to research the influence of a certain physical parameter on the resulting sound. This could enable users to tune an instrument to a desired sound characteristic and research the influence of fine structures on the resulting sound. Sounds and further application examples of the real-time violin model can be found at www.systematicmusicology.de.

20.11 Summary and Outlook

As shown in this chapter, FPGAs offer a vast amount of possible applications in the field of digital signal processing for research as well as for hardware prototyping. The raw processing power and flexible architecture enables them to perform highly specialized tasks using the inherently bit-parallel structure as well as the high clock speed of specialized logic blocks for sequential processing. But, as the saying goes, *with great power comes great responsibility*: handling large FPGA designs can become exceedingly difficult for a designer and places substantial demands on the design process and the appertaining tool chain. Nonetheless, as the design tools get more elaborate, automating and optimizing various steps of the development cycle, FPGA programming gets more and more accessible.

This is one of the future routes of FPGA development systems that two of the largest vendors are working on at the moment. In 2013, Altera has published a library making their FPGAs programmable using OpenCL code in C or C++. Xilinx devices can also be programmed from C or C++ using the high-level synthesis (HLS) tools introduced around 2013.

Simplifying the overall design process and thereby reducing the initial learning curve for hardware programming will surely be one of the central topics in future FPGA developments. On the hardware side there are several exciting developments in the field of microprocessor-FPGA interaction. The line of Xilinx and Altera chips that include an ARM processor is growing constantly and the purchase of Altera by Intel in 2015, as well as the cooperation between Xilinx and IBM announced in the same year, points to even more elaborate microprocessor-FPGA combinations, including server CPUs of the x86 family connected directly to FPGAs.

Altogether the recent developments indicate that FPGAs will play an important role in high-performance computing over the next decade and will also become more accessible to a larger audience due to simplifications in the design process.

The implementations and examples of signal processing applications in FPGAs presented in this chapter highlight how the flexibility of FPGAs can be used to simulate and model timing-critical problems in music

DSP applications. The main purpose of this chapter was the introduction of a platform capable of *real* real-time implementations of algorithms as well as the possibility to utilize special hardware traits to speed up timing-critical computations in acoustic research. In musical acoustics and especially in physical modeling applications there is an ever rising need for high(er)-performance computations to enable researchers as well as musicians to interact with highly accurate models in real time. Hence, extending the toolbox of a researcher to more specialized hardware platforms can open up new areas of research by facilitating the use of methods that were impractical previously due to throughput limitations of *standard* hardware.

References

20.1 Cisco: *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015–2020 White Paper* (Cisco Systems Inc., San Jose 2016)

20.2 A. Brandt: *Noise and Vibration Analysis: Signal Analysis and Experimental Procedures* (Wiley, Chichester 2011)

20.3 E.S. Gopi: *Digital Signal Processing for Medical Imaging Using MATLAB* (Springer, New York 2013)

20.4 U. Zölzer, X. Amatriain, D. Arfib, J. Bonada, G. De Poli, P. Dutilleux, G. Evangelista, F. Keiler, A. Loscos, D. Rocchesso, M. Sandler, X. Serra, T. Todoroff: *DAFX: Digital Audio Effects* (Wiley, Chichester 2002)

20.5 M. Holters, U. Zölzer: Physical modelling of a wah-wah effect pedal as a case study for application of of the nodal DK method to circuits with variable parts. In: *Proc. 14th Int. Conf. Digit. Audio Eff. DAFx-11* (2011) pp. 31–35

20.6 D. Arfib: Different ways to write digital audio effects programs. In: *Proc. Digit. Audio Eff. (DAFx-98), Barcelona* (1998) pp. 188–191

20.7 K. Zuse: *Der Computer – Mein Lebenswerk* (Springer, Berlin, Heidelberg 2010)

20.8 J. Fowers, G. Brown, P. Cooke, G. Stitt: A performance and energy comparison of FPGAs, GPUs, and multicores for sliding-window applications. In: *Proc. ACM/SIGDA Int. Symp. Field Prog. Gate Arrays* (ACM, New York 2012) pp. 47–56

20.9 G.W. Leibnitz: Explication de lárithmétique binaire, Mem. Acad. R. Sci. **3**, 85–93 (1703)

20.10 F. Perkins: *Leibniz and China: A Commerce of Light* (Cambridge Univ. Press, Cambridge 2009)

20.11 G. Boole: *The Mathematical Analysis of Logic* (Philosophical Library, New York 1847)

20.12 G. Boole: *An Investigation of the Laws of Thought: On Which are Founded the Mathematical Theories of Logic and Probabilities* (Dover, Mineola 1854)

20.13 T. Hailperin: Boole's algebra isn't Boolean algebra, Math. Mag. **54**(4), 173 (1981)

20.14 J. Corcoran: Aristotle's prior analytics and Boole's laws of thought, Hist. Phil. Log. **24**(4), 261–288 (2003)

20.15 C.E. Shannon: *A Symbolic Analysis of Relay and Switching Circuits*, Ph.D. Thesis (Massachusetts Insitute of Technology, Cambridge 1940)

20.16 B.J. Copeland (Ed.): *Colossus: The Secrets of Bletchley Park's Codebreaking Computers* (Oxford Univ. Press, Oxford 2006)

20.17 P.E. Ceruzzi: *Computing: A Concise History*, The MIT Press Essential Knowledge Series (MIT Press, Cambridge 2012)

20.18 C. Maxfield: *The Design Warriors Guide to FPGAs* (Elsevier, Oxford 2004)

20.19 L. Wirbel: Remembering Ross Freeman, http://www.edn.com/electronics-blogs/fpga-gurus/4306558/Remembering-Ross-Freeman (EDN Network 2009)

20.20 L. Pantaleone, E. Todorivich: Accelerating embedded software processing in a FPGA with PowerPC and Microblaze. In: *Natl. Conf. Inform. Eng. Inf. Syst., San Louis* (2013)

20.21 N. Hemsoth: FPGAs Glimmer on the HPC Horizon, Glint in Hyperscale Sun, https://www.nextplatform.com/2015/11/17/fpgas-glimmer-on-the-hpc-horizon-glint-in-hyperscale-sun/ (The Next Platform 2015)

20.22 T.P. Morgan: Why Hyperscalers And Clouds Are Pushing Intel Into FPGAs, https://www.nextplatform.com/2015/07/29/why-hyperscalers-and-clouds-are-pushing-intel-into-fpgas/ (The Next Platform 2015)

20.23 Xilinx: *Configurable Logic Block User Guide* (Xilinx, San Jose 2010)

20.24 Xilinx: *Virtex-6 FPGA Configurable Logic Block User Guide* (Xilinx, San Jose 2012)

20.25 Altera: *Stratix V Device Handbook* (Altera, San Jose 2015)

20.26 D.S. Brown, G.Z. Vranesic: *Fundamentals of Digital Logic with VHDL Design*, McGraw-Hill Series in Electrical and Computer Engineering, 3rd edn. (McGraw-Hill, New York 2009)

20.27 A.V. Pedroni: *Circuit Design with VHDL* (MIT Press, Cambridge 2004)

20.28 J. Reichardt, B. Schwarz: *VHDL-Synthese: Entwurf digitaler Schaltungen und Systeme*, 4th edn. (Oldenbourg, München 2009)

20.29 J.P. Ashenden: *The Designer's Guide to VHDL*, 3rd edn. (Morgan Kaufmann, San Francisco 2010)

20.30 P. Pong: *FPGA Prototyping by VHDL Examples* (Wiley-Interscience, Hoboken 2008)

20.31 M. Zwoliński: *Digital System Design with VHDL*, 2nd edn. (Prentice Hall, Harlow 2004)

20.32 J.P. Ashenden: *The Designer's Guide to VHDL*, 2nd edn. (Morgan Kaufmann, San Francisco 2002)

20.33 A.H. Wilen, J.P. Schade, R. Thornburg: *Introduction to PCI Express: A Hardware and Software Developer's Guide*, Engineer to Engineer Series

(Intel, Santa Clara 2003)
20.34 C. McGinnis: PCI-SIG® Fast Tracks Evolution to 32GT/s with PCI Express 5.0 Architecture, http://www.businesswire.com/news/home/20170607005351/en/PCI-SIG%C2%AE-Fast-Tracks-Evolution-32GTs-PCI-Express (PCI-SIG 2017)
20.35 H. Liebig, T. Flik, M. Menge: *Mikroprozessortechnik und Rechnerstrukturen* (Springer, London 2005)
20.36 C.M. Herbordt, T. VanCourt, Y. Gu, B. Sukhwani, A. Conti, J. Model, D. DiSabello: Achieving high performance with FPGA-based computing, Computer **40**, 50–57 (2007)
20.37 U.M. Baese: *Digital Signal Processing with Field Programmable Gate Arrays*, 2nd edn. (Springer, Berlin, Heidelberg 2007)
20.38 R. Woods (Ed.): *FPGA-Based Implementation of Signal Processing Systems* (Wiley, Chichester 2008)
20.39 B.A. Kahng, J. Lienig, L.I. Markov, J. Hu: *VLSI Physical Design: From Graph Partitioning to Timing Closure* (Springer, Dordrecht 2011)
20.40 K. Veggeberg, A. Zheng: Real-time noise source identification using programmable gate array FPGA technology, Proc. Meet. Acoust. **5**, 2582 (2009)
20.41 C. Hai, W. Ping: The high speed implementation of direction-of-arrival estimation algorithm, IEEE Int. Conf. Commun. Circuits Syst. West Sino Expo. **2**, 922–925 (2002)
20.42 P. Chen, X. Tian, Y. Chen, X. Yang: Delay-sum beamforming on FPGA. In: *ICSP 2008 Proc.* (2008) pp. 2542–2545
20.43 A. Madanayake, L. Bruton, F. Comis, C. Comis: FPGA architectures for real-time 2D/3D FIR/IIR plane wave filters. In: *Proc. Int. Symp. Circuits Syst. ISCAS*, Vol. 3 (2004)
20.44 S. Kai, Y. Xu, S. Jiang, H. Zhu: Converting analog controllers to digital controllers with FPGA. In: *9th Int. Conf. Signal Process.* (2008)
20.45 O. Maslennikow, A. Sergiyenko: Mapping DSP algorithms into FPGA. In: *Int. Symp. Parallel Comput. Electr. Eng. (PARELEC'06), Bialystok* (2006) pp. 208–213, https://doi.org/10.1109/PARELEC.2006.51
20.46 T. Brich, K. Novacek, A. Khateb: The digital signal processing using FPGA. In: *ISSE 2006, 29th Int. Spring Semin. Electron. Technol.* (2006) pp. 322–324
20.47 M. Liu, W. Kuehn, Z. Lu, A. Jantsch: System-on-an-FPGA design for real-time particle track recognition in physics experiments. In: *11th Euromicro Conf. Digit. Syst. Design Archit., Methods Tools* (2008)
20.48 B. von Herzen: Signal processing at 250 mHz using high-performance FPGA's, IEEE Trans. Very Large Scale Int. (VLSI) Syst. **6**(2), 238–246 (1998)
20.49 Z. Wang, R. Jin, J. Geng, Y. Fan: FPGA implementation of downlink DBF calibration. In: *Antennas Propag. Soc. Int. Symp.* (2005)
20.50 K.H. Moustafa, M.F. Ahmed, M. Romeh, A. Fahmy: Real time processing of a lattice wave digital matched filter. In: *IEEE Int. Symp. Signal Process. Inf. Techol. ISSPIT* (2011) pp. 404–408
20.51 S. Thilagam, P. Karthigaikumar: Implementation of adaptive noise canceller using FPGA for real-time applications. In: *2015 2nd Int. Conf. Electr. Commun. Syst. (ICECS), Coimbatore* (2015) pp. 1711–1714, https://doi.org/10.1109/ECS.2015.7124878
20.52 H.P. Afshar, P. Ienne: Highly versatile DSP blocks for improved FPGA arithmetic performance. In: *Proc. IEEE Symp. Field-Prog. Custom Comput. Mach. FCCM* (2010) pp. 29–236
20.53 L. Struyf, S. De Beugher, D.H. Van Uytsel, F. Kanters, T. Goedemé: The battle of the giants: A case study of GPU vs FPGA optimisation for real-time image processing, Proc. PECCS **1**, 112–119 (2014)
20.54 W. Chen, P. Kosmas, M. Leeser, C. Rappaport: An FPGA implementation of the two-dimensional finite-difference time-domain (FDTD) algorithm. In: *Proc. 2004 (ACM/SIGDA) 12th Int. Symp. Field Prog. Gate Arrays, New York* (2004) pp. 213–222
20.55 J.A. Gibbons, D.M. Howard, A.M. Tyrrell: FPGA implementation of 1d wave equation for real-time audio synthesis, IEEE Proc. Comput. Digit. Tech. **152**(5), 619–631 (2005)
20.56 E. Motuk, R. Woods, S. Bilbao: Implementation of finite-differece schemes for the wave equation on FPGA. In: *IEEE Int. Acoust. Speech Signal Process. ICASSP* (2005) p. 3
20.57 E. Motuk, R. Woods, S. Bilbao, J. McAllistere: Design methodology for real-time FPGA-based sound synthesis, IEEE Trans. Signal Process. **55**(12), 5833–5845 (2007)
20.58 H.E. Motuk: *System-On-Chip Implementation of Real-Time Finite Difference Based Sound Synthesis*, Ph.D. Thesis (Queen's Univ., Belfast 2006)
20.59 G. Martins, M. Barata, L. Gomes: Low cost method to reproduce sound with FPGA. In: *IEEE Int. Symp. Ind. Electron. ISIE* (2008)
20.60 F. Pfeifle, R. Bader: Real-time finite difference physical models of musical instruments on a field programmable gate array (fpga). In: *Proc. Int. Conf. Digit. Audio Eff. (DAFx-12), New York* (2012) pp. 63–70
20.61 F. Pfeifle, R. Bader: Real-time finite difference method physical modeling of musical instruments using field-programmable gate array hardware, J. Audio Eng. Soc. **63**(12), 1001–1016 (2015)
20.62 K. Kroschel, K.-D. Kammeyer: *Digitale Signalverarbeitung – Filterung und Spektralanalyse mit MATLAB-Übungen*, 6th edn. (Vieweg+Teubner, Wiesbaden 2006)
20.63 N.T. Rajapaksha, C. Wijenayake, A. Madanayake, L.T. Bruton: Raster-scanned wave-digital filter architectures for multi-beam 2d IIR broadband beamforming. In: *Proc. Int. Conf. Microelectron. ICM* (2010) pp. 112–115
20.64 H. Li, A. Kummert, S. Schauland, J. Velten: 3D wave digital filter implementation on a virtex2 FPGA board with external SDRAM. In: *2009 Int. Workshop Multidimens. (nD) Syst., Thessaloniki* (2009) pp. 1–5, https://doi.org/10.1109/NDS.2009.5191463

20.65 S. Bilbao: *Numerical Sound Synthesis: Finite Difference Schemes and Simulation in Musical Acoustics* (Wiley, Chichester 2009)

20.66 F. Pfeifle, R. Bader: *Musical Acoustics, Neurocognition and Psychology of Music, Chapter Real-Time Physical Modelling of a Real Banjo Geometry Using FPGA Hardware Technology* (Rolf Bader, Frankfurt am Main 2009) pp. 71–86

20.67 F. Pfeifle, R. Bader: Real-time virtual banjo model and measurements using a microphone array, J. Acoust. Soc. Am. **125**(4), 2515–2515 (2009)

20.68 F. Pfeifle, R. Bader: Membrane modes and air resonances of the banjo using physical modeling and microphone array measurements, J. Acoust. Soc. Am. **127**(3), 1870–1870 (2010)

20.69 S. Bilbao: Robust physical modeling sound synthesis for nonlinear systems, IEEE Signal Process. Mag. **24**(2), 32–41 (2007)

20.70 C. Jordan: *Calculus of Finite Differences* (Chelsea, New York 1950)

20.71 J. Strikwerda: *Finite Difference Schemes and Partial Differential Equations*, 2nd edn. (SIAM, Philadelphia 2005)

20.72 E. Hairer, C. Lubich, G. Wanner: *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, Springer Series in Computational Mathematics, Vol. 31 (Springer, Berlin, Heidelberg 2002)

20.73 R. Bader: *Computational Mechanics of the Classical Guitar* (Springer, Berlin, Heidelberg 2005)

20.74 R. Bader: Nonlinearities in the sound production of the classical guitar. In: *Proc. Forum Acust. 2005* (2005) pp. 685–689

20.75 B.E. Moore: Conformal multi-symplectic integration methods for forced-damped semi-linear wave equations, Math. Comput. Simul. **80**(1), 20–28 (2009)

20.76 D.W. Markiewicz: *Survey on Symplectic Integrators* (Univ. California, Berkeley 1999) preprint

20.77 R. Mclachlan: Symplectic integration of hamiltonian wave equations, Numer. Math. **66**, 465–492 (1994)

20.78 R. Courant, K. Friedrichs, H. Lewy: Über die partiellen Differenzengleichungen der mathematischen Physik, Math. Ann. **100**(1), 32–74 (1928), https://doi.org/10.1007/BF01448839

20.79 L. Verlet: Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules, Phys. Rev. **159**(1), 98–103 (1967)

20.80 K.J. Bathe: *Finite-Element Methoden* (Springer, Berlin, Heidelberg 2002)

20.81 M. Campbell, P. Campbell: *The Science of String Instruments* (Springer, New York 2010) pp. 301–315

20.82 B. Geiser: Studien zur Frühgeschichte der Violine. In: *Publikationen der Schweizerischen Musikforschenden Gesellschaft, Serie 2* (Gemeinsamer Bibliotheksverbund (GBV)/Verbundzentrale des GBV (VZG), Bern 1974)

20.83 D.D. Boyden: *The History of Violin Playing from Its Origins to 1761, and Its Relationship to the Violin and Violin Music* (Oxford Univ. Press, Oxford 1967)

20.84 K.C. Wali: *Cremona Violins: A Physicist's Quest for the Secrets of Stradivari* (World Scientific, Hackensack 2010)

20.85 D. Ullmann: *Chladni und die Entwicklung der Akustik von 1750–1860* (Birkhäuser, Basel 1996)

20.86 H. von Helmholtz: *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (Friedrich Vieweg und Sohn, Braunschweig 1896)

20.87 L. Cremer: *Physik der Geige* (Hirzel, Stuttgart 1981)

20.88 G. Bissinger, E.G. Williams, N. Valdivia: Violin f-hole contribution to far-field radiation via patch near-field acoustical holography, J. Acoust. Soc. Am. **121**(6), 3899–3906 (2007)

20.89 G. Bissinger: *The Science of String Instruments* (Springer, New York 2010) pp. 317–345

20.90 H.N. Fletcher, T.D. Rossing: *Physics of Musical Instruments*, 2nd edn. (Springer, New York 2000)

20.91 C.M. Hutchins: Klang und Akustik der Geige, Spektrum Wiss. **2**, 112–122 (1981)

20.92 J.-L. Florens: Expressive bowing on a virtual string instrument, Lect. Notes Comput. Sci. **2915**, 487–496 (2004)

20.93 M. Demoucron: *On the Control of Virtual Violins – Physical Modelling and Control of Bowed String Instruments*, PhD Thesis (Université Pierre et Marie Curie – Paris VI, Royal Institute of Technology, Stockholm 2008)

20.94 E. Maestre: Analysis/synthesis of bowing control applied to violin sound rendering via physical models, Proc. Meet. Acoust. **19**(1), 3271 (2013)

20.95 J. Woodhouse, P.M. Galluzo: The bowed string as we know it today, Acta Acust. united Acust. **90**, 579–589 (2004)

20.96 E. Bavu, J. Smith, J. Wolfe: Torsional waves in a bowed string, Acta Acust. united Acust. **91**, 241–246 (2005)

20.97 A. Chaigne, A. Askenfelt: Numerical simulations of piano strings. I. A physical model for a struck string using finite difference methods, J. Acoust. Soc. Am. **95**(2), 1112–1118 (1994)

20.98 J. Bensa, S. Bilbao, R. Kronland-Martinet, J.O. Smith III: The simulation of piano string vibration: From physical models to finite difference schemes and digital waveguides, J. Acoust. Soc. Am. **114**(2), 1095–1107 (2003)

20.99 H.N. Fletcher, T.D. Rossing: *The Physics of Musical Instruments* (Springer, New York 1998)

20.100 F. Pfeifle, R. Bader: Real-time finite-difference string-bow interaction field programmable gate array (fpga) model coupled to a violin body, J. Acoust. Soc. Am. **130**(4), 2507–2507 (2011)

20.101 J. Woodhouse: On the "bridge hill" of the violin, Acta Acust. united Acust. **91**, 155–165 (2005)

Part C

Music Psy

Part C Music Psychology – Physiology

Ed. by Stefan Koelsch

There is no music without a brain. The brain is the biological substrate of music perception, creative processes, and of planning, executing, and monitoring the movements necessary to play music. For the investigation of the neural correlates of music, neurophysiological methods are used such as functional magnetic resonance imaging (fMRI) or electroencephalography (EEG). While listening to a musical piece, or while playing music, the brain dynamics of an individual is recorded, both in time and space. While the EEG mainly records neural activity in the neocortex, fMRI is able to measure neural activity both in the cortex and in subcortical structures located deep inside the brain, e.g., the thalamus, the amygdala, or the basal ganglia. EEG has a high temporal resolution (in the submillisecond range), enabling researchers to record brain activity with sampling rates necessary for the investigation of very fast perceptual or executive processes. fMRI, on the other hand, has a high spatial resolution (in the millimeter- or even submillimeter range), enabling researchers to differentiate brain regions, and thus the *where* of music processing in the brain.

EEG-recorded brain-electric potentials, also referred to as *event-related potentials* (ERPs), have been employed to investigate auditory-perceptual processes, auditory attention, auditory memory, musical syntax, musical meaning, and music-evoked emotions, as well as action-related processes during music production. For example, the so-called mismatch negativity (MMN) has been used to investigate auditory sensory memory, auditory grouping, and musical Gestalt. The so-called early right anterior negativity (ERAN), on the other hand, has been used to investigate the processing of musical syntax. It was shown that the ERAN is mainly generated in Broca's area, a region of the brain known to analyze the syntax of language – thus showing that both musical and linguistic syntax is processed in overlapping regions of the brain. On the other hand, studies with fMRI have provided detailed knowledge about brain structures that are involved in music processing. For example, fMRI studies have shown that beat perception involves not only the cerebellum, but also the basal ganglia, the premotor cortex, and the supplementary motor area.

However, understanding the functional neuroanatomy of music processing is challenging, e.g., because brain regions are usually active within networks, interacting strongly with each other, or even playing different roles during different tasks within different networks. How the interaction of many neurons leads to a sorting and organizing of musical content (which is then used to identify and recognize rhythms, melodies, or timbres) is not yet understood.

This section gives an overview of some different aspects and musical features investigated with neuroscientific methods (mainly EEG and fMRI), and modeling of neural networks in relation to perceptual tests. It gives insight into the state of the art in the field and discusses recent findings and trends.

In Chap. 21, *Simon Grondin*, *Emi Hasuo*, *Tsuyoshi Kuroda* and *Yoshitaka Nakajima* discuss auditory time perception, where perception based on the interval between two events is different from that of beats and a rhythm differentiating musical time. Also there is a difference in the perception of time in music and in language. The paper discusses the influence of pitch, intensity or musical markers in time perception. Furthermore, distinguishing the perception of musicians and nonmusicians gives further insight into the problem.

Elvira Brattico, *Chiara Olcese* and *Mari Tervaniemi* review the perception of musical sound in the brain in Chap. 22. They discuss the role of attention in the problem. Discussing fMRI and EEG investigations, different brain regions are identified for their contributiond to processing. In particular, experiments using the MMN and the ERAN are found to be helpful in recognizing ERPs that can give deeper insights into the attention problem of automatically or consciously perceived musical sounds or syntax.

Lola L. Cuddy reviews long-term memory in music in Chap. 23. This is discussed in terms of semantics and meaning, and episodic memory or procedural memory. Discussing neurophysiological evidence, Cuddy reports studies showing the preservation of musical memory in Alzheimer patients, and discusses potential use of these findings for therapy.

In Chap. 24, *Katrin Schulze*, *Victoria Williamson* and *Stefan Koelsch* discuss the relation between music and language in working memory. Based on fMRI and EEG studies, the paper discusses the role of Broca's region, the premotor cortex as well as other regions for working memory, and discusses which of these regions might differ between verbal and tonal working memory. They suggest an overlap between neural resources underlying working memory for tones and words, and put forward the hypothesis that sensorimotor codes (i. e., action-related processes) are fundamental for auditory working memory.

Marcus Pearce and *Martin Rohrmeier* discuss musical syntax in Chap. 25 in terms of its theoretical foundations. Based on Noam Chomsky's generative grammar, musical syntax is described as a set of hierarchically organized rules underlying the generation of a musical piece. The difference between finite-state and

context-free grammars is discussed, where the former assumes local dependencies among (adjacent) musical events, while the latter allows nonlocal dependencies between (nonadjacent) musical events.

Following these ideas, in Chap. 26 *Marcus Pearce* and *Martin Rohrmeier* continue to discuss musical syntax, now in terms of neural and perceptual models. The implementation of finite-context models as *n*-gram systems, compared to hidden-Markov models as finite-state models is discussed, next to neural networks. These findings are discussed with respect to perceptual as well as neurophysiological findings, methods and ideas.

In Chap. 27, *Tram Nguyen*, *Aaron Gibbings* and *Jessica Grahn* discuss musical rhythm and beat perception. Sensorimotor synchronization is discussed as the basic mechanism listeners use to perceive a beat or rhythm. The models and perception tests find differences between beat as a steady pulse, and rhythm as patterns of musical texture. This is also supported by fMRI, positron emission tomography (PET) or EEG experiments, but still research is ongoing and many questions remain open.

Chapter 28 by *Giacomo Novembre* and *Peter Keller* is devoted to music and action. It relates music perception to performance and reviews the perceptual data as well as the neuroimaging evidence. The relation between performance and social bonding is discussed. The role of beat perception on action is shown. The paper finds a coupling between the brain regions responsible for perception and performance.

Finally, Chap. 29 by *Tuomas Eerola* discusses music and emotion. He distinguishes between automatic emotions in response to music-making as a mapping of musical features and emotions an expectancy of musical events, which might be violated and cause emotions; and an entrainment between music and the subject, which leads to an emotional movement in the listener. He discusses measurement techniques for emotions and discusses future problems in this field.

21. Auditory Time Perception

Simon Grondin, Emi Hasuo, Tsuyoshi Kuroda, Yoshitaka Nakajima

In this chapter, we propose to review studies on the capability of making explicit judgments about the duration of auditory time intervals. After a brief look at the main methods used to study time perception, we then focus on factors affecting sensitivity to time (e.g., discrimination levels) including repetition of intervals and the duration range under investigation, as well as whether the interval is embedded in a music or in a speech sequence. Factors affecting the perceived duration of sounds and time intervals are then described using for example markers' length, space, pitch, and intensity. The last section of this chapter reviews the main theoretical perspectives in the field of temporal information processing, with an emphasis on the distinction between the traditional beat-based versus interval-based mechanisms.

One critical aspect of music perception and production is the capability of dealing with time. For instance, variations in time are minimal when several notes are played together, or almost, to form a chord or an arpeggio chord, whereas several notes are also sometimes delivered in very rapid succession to form an appoggiatura, or played successively but slightly more slowly in order to contribute to tempo.

In this chapter, we place emphasis on a description of empirical studies and theoretical issues related to the ability to estimate short time intervals. In other words, the content is dedicated to explicit judgments concerning duration. Phenomena like temporal displacement [21.1] will not be reviewed herein. Temporal displacement means there is no perfect correspondence between the physical arrival of stimuli and the perception of their simultaneity or successiveness. In addition, we will also not review the capacity of judging the order of arrival (temporal order judgments: TOJ) of stimuli. Note however that in their classical work, *Hirsh* and *Sherrick* [21.2] reported that the delay between sensory events for TOJ is approximately 20 ms. This delay is much longer than the delay (temporal separation) necessary to distinguish between simultaneity and successiveness. In the latter case, the delay depends on the sensory modality employed; typically 2–3 ms is needed in the auditory modality, and more in the visual and the tactile modality [21.3]. In the former case, TOJ, this delay is independent of the sensory modality. For the specific case of audition, the reader interested in the study of dichotic TOJ may consult a summary of threshold results in [21.4, Table 1].

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_21

21.1 Methods for Studying Interval Processing

There is a long history of research on psychological time and interval processing [21.5]. In order to understand the meaning of the results extracted from this literature, it is relevant to become familiar with some basic methods and terminology [21.6].

One fundamental distinction is related to the type of performance that is of interest. If someone is asked to tap twice with a finger and we have 3 s between the taps, it is likely that the main performance index will be how close to the target the finger production is. Maybe there will be an overestimation or an underestimation of time; it is the degree of deviation, also referred to as perceived duration in the next section, which is of interest. Suppose now that the 3 s tapping task is repeated several times. It may occur that the mean production is close to 3 s, but this does not mean that the timing system is excellent. The timekeeping system of a participant where each production lasts between 2.986 and 3.014 s is judged as better than the timekeeping system of someone whose productions are in the range 2.5 to 3.5 s. The mean productions (perceived duration) of the participants could be the same, but it is obvious that there is much more variability (less sensitivity) in the system of the second participant. In the next section, the dependent variable of interest is related to variability.

Another fundamental notion to learn is that the basic structure of an interval can vary. Essentially, one should distinguish a filled interval, where a continuous signal marks the beginning and end of the time period to be processed, from an empty interval, which is an empty period of time marked by two brief signals of or below 10 ms typically. Note that any of these signals can be issued from one sensory modality or another.

The methods described herein provide the possibility to obtain performance indexes related to both variability and perceived duration. These methods will not be described in detail, the aim of the presentation being to provide the reader with the main tools to approach the study of internal time. One basic method for investigating capabilities and mechanisms related to time perception is to present a participant with a signal (a stimulus) and to ask this person to estimate its duration verbally in chronometric units (seconds or minutes). This method is called verbal estimation. If intervals are not too long, an experimenter may rather choose to ask the participant to reproduce, usually with finger taps, the duration of these intervals. This method is referred to as time reproduction. There are different reproduction methods: two brief successive presses to mark the beginning and end of an interval, one continuous press to mark the entire interval, or a sensory signal that marks the beginning of an interval and a key press by the participant that marks the end. These variations may lead to slightly different results [21.7]. It is also possible to adopt a method where the experimenter verbally reports the target duration, in chronometric units, and the participants should reproduce this duration (time production), usually with finger taps. Note that the term production is sometimes used in the literature for indicating that, after the presentation of an interval or a series of intervals of equal duration, a participant has to produce a series of intervals of the same length with a series of successive finger taps. This may or may not involve a synchronization phase, i. e., a series of taps synchronized with the signals marking time, before tapping without these signals [21.8]. This interval production method proved to be very useful for distinguishing the respective proportion of variance belonging to the motor process involved in the tapping activity and the variance issuing specifically from the mechanism responsible for keeping track of time [21.9].

In addition to these methods (time estimation, reproduction, and production), there is a fourth one, interval comparison. There are several ways of presenting intervals to enable comparison. Many of the methods used for investigating human timing derive from classical psychophysics, and some from animal timing studies. Basically, a forced-choice procedure can be used where two successive intervals are presented and a participant is asked whether the second one is shorter or longer than the first one. In experimental designs where there are one standard and multiple comparison intervals, the standard is usually presented first. When the comparison interval is presented first, the discrimination level is lower [21.10]. There could be one presentation of a standard and of a comparison interval on each trial, or there could be multiple presentations of the same intervals. As we will see in Sect. 21.2.1, the number of intervals changes the discrimination level.

Amongst the other main comparison methods, there is one called single stimulus. With such a procedure, a response is required from a participant after each presentation of an interval. In one version of this method, the *bisection* task, the shortest and the longest (standards) of a series of intervals are presented several times, and in subsequent experimental trials, after each presentation of one of this series, a participant is asked to say if the presented interval is closer to the shortest or to the longest standard interval. Instead of presenting these standards, the experimenter could present several times the mid interval from a series, and then, in subsequent trials, ask if the presented interval from that

series is identical or different: this is a *generalization* task. Also, other studies on the sensitivity to time used a task where anisochrony is to be detected in a sequence of isochronous brief stimuli.

With many of these methods, it is possible to measure a difference threshold (or just noticeable difference, JND), which is the least two intervals need to be different to distinguish between them. The value of this threshold depends on the method and on the operational definition adopted for measuring it. Moreover, this value (threshold) is sometimes divided by the value of the duration under investigation: this ratio is the Weber fraction.

Finally, note that for measuring the perceived duration of a time interval, there is another method, called the method of adjustment, where a standard interval and a comparison interval are presented successively, and the participants are asked to adjust the duration of the comparison to make it subjectively equal to that of the standard [21.11].

21.2 Processing Time Intervals: Variability

In this section of the chapter, we will focus on different factors affecting the sensitivity to time. After a description of the impact of the number of intervals presented and of the duration range on the levels of duration discrimination, we will report how sensitive to time we are in the context of music and speech. The section ends with a comparison of audition with other sensory modalities for processing time intervals.

21.2.1 Number of Intervals and Duration Range

There are several factors that determine the level of sensitivity to temporal intervals. One of these factors is of particular interest for the present book: Increasing the number of intervals presented, i. e., inducing a pace [21.10, 12, 13]. Another factor of interest is related to the range of duration (tempo) under investigation.

When single intervals are presented, it is well established that the Weber fraction depends on the range of duration. In the auditory modality, this fraction is higher for very brief intervals (< 0.1 s) than for intervals from 0.1 to 2 s [21.14]. Such results are true for filled as well as for empty intervals [21.15, 16].

In a series of experiments where the threshold was estimated using an adaptive procedure with a convergence toward a 70.7% probability of correct detections, *Drake* and *Botte* [21.17] tested the influence of the number of intervals on sensitivity to time. In one experiment, they used two or more consecutive tones to mark one, two, four, or six successive intervals. They presented the standard interval(s), and then the comparison interval(s) (same number of intervals) which were shorter or longer. Participants had to indicate whether the standard or the comparison sequence was faster. Fors tandard intervals lasting 0.1 to 1 s, discrimination was better with two than with one interval, and with four than with two intervals; but it was not better with six than with four intervals. The tempo sensitivity (Weber fraction) in Drake and Botte was about 5% with a 100 ms standard, and slightly above 2% at 200 ms when a series of six intervals were presented. This means that listeners could detect a change in duration when this duration was changed by more than 5 ms for the 100 ms standard, and by more than 4 ms for the 200 ms standard. Note, however, that this value is partly related to the operational definition adopted for establishing the threshold.

Weber fractions for the discrimination of brief intervals marked by auditory stimuli presented in sequences ranged in other studies from 3 to 13% [21.18], and are often reported to be about 5−10% [21.19, 20]. In an interval discrimination study by *Hirsh* et al. [21.21] where participants were asked to detect anisochrony (the deviation of a tone) in sequences of tones, the Weber fractions were about 6−8%, 11−12%, and 20%, respectively, for standard intervals (interonset intervals) equal to 200, 100, and 50 ms. Higher Weber fractions (lower performances) with briefer intervals could be due to the fact that participants might have been unable to perceive some of the presented intervals correctly. Indeed, when participants are asked to report the number of sounds they perceive successively, these sounds must be separated by interonset intervals longer than about 100 ms for estimating the number correctly [21.22] (see also [21.23]). Nevertheless, the impairments of discrimination with shorter intervals, reported by Hirsh et al., are compatible with what we found for the discrimination of single intervals.

While it is well established that varying the number of presented intervals influences discrimination levels, it was only recently that researchers varied the number of intervals in each sequence, the first or the second, independently [21.10, 12, 13]. For instance, *Miller* and *McAuley* [21.13] showed that increasing the number of intervals in the second (comparison) sequence, but not the number of intervals in the first (standard)

sequence, improves time sensitivity. This finding has been shown for standard intervals ranging from 300 to 700 ms [21.12, 13].

One of the most complete studies on the influence of the number of intervals, standard or comparison, on temporal discrimination at different duration ranges was recently reported by *ten Hoopen* et al. [21.24]. In their first experiment, which included base durations of 100, 200, 400, and 800 ms, two sets of sequential sound patterns were presented: one sequence of 9 empty time intervals (standard) of equal duration (marked by 10 brief sounds). These standard intervals were followed in continuity by 1, 2, 3, 5, 7, or 9 comparison intervals and participants had to judge whether the tempo in the (total) sequence became faster or slower. The results revealed that sensitivity increased with an increase of the number of comparison intervals. In another part of this experiment, there was one last comparison interval immediately preceded by 1, 2, 3, 5, 7, or 9 standard intervals and participants had to judge whether the last sound of the total sequence came too early or too late. Sensitivity was not improved by an increasing number of preceding standards.

In ten Hoopen et al.'s second experiment, there were four sequential conditions: 9 leading standard intervals followed by 1, 2, ..., or 9 comparison intervals; 9 leading comparison intervals followed by 1, 2, ..., or 9 standard; 9 comparison intervals preceded by 1, 2, ..., or 9 standard, and 9 standard intervals preceded by 1, 2, ..., or 9 comparison intervals. The threshold was lower (higher sensitivity) when the increase in the number of intervals occurred before the tempo change rather than after. In their following experiment [21.24], the number of intervals was varied from 1, 2, 3, 4, to 5, yielding 25 standard/comparison pairings. The effect of the number of intervals before the tempo change was larger than that after the tempo change. The authors also noted that the thresholds were largest when the tempo changes occurred early, i. e., at the beginning of the sequence, and second largest when the change occurred at the very end of the sequence; the threshold was smaller when tempo changes occurred during the sequence. Finally, in their investigation, ten Hoopen et al. showed that adding one or two intervals improves performance but additional interval presentations provide negligible additional benefits.

The effect of the number of intervals is most often tested on brief intervals (< 1 s). Recently, it has been shown that the benefit of multiple interval presentations applies from 1 to 1.9 s, and is not limited to duration discrimination [21.25]. The effect applies with a duration reproduction task and a categorization task. What is more specifically interesting in this report is the fact that variability does not increase proportionally as a function of time (higher variability-to-time ratio with longer intervals), but this effect does occur similarly with single and multiple presentation conditions. These results are also consistent with others showing the violation of Weber's law [21.26]; for reviews, see [21.27] or [21.28].

Working with long intervals invariably raises the question of using counting or any other segmenting strategy to keep track of time. It is known that intervals do not have to be very long to gain benefit from a strategy where subintervals are used [21.29, 30]. For the reproductions of very long intervals (6–24 s), it is shown that both singing and counting reduce variability considerably, and keep the reproductions close to the target interval [21.31], see also [21.32]. Moreover, *Grondin* and *Killeen* [21.31] showed that with both strategies, singing and counting, musician participants remained much closer to the target intervals than nonmusicians. Also, the variability of the reproduced intervals is not only much lower for musicians, but the variability-to-duration ratio tends to keep decreasing as intervals get longer with musicians, but not with nonmusicians.

We have reviewed the literature where discrimination is improved when the number of presented time intervals is increased. However, it should be noted that discrimination between different temporal patterns can be both improved and impaired by the very fact that two or more intervals neighbor each other. *Sasaki* et al. [21.33] investigated how temporal patterns of two adjacent time intervals were discriminated from each other. For some patterns, the discrimination was more difficult than would have been predicted from a common duration discrimination experiment in which each time interval was isolated. This happened when the first time interval was varied from 60 to 100 ms, whereas the second time interval was varied from 120 to 80 ms, keeping the total duration of the adjacent intervals at 180 ms. This impairment can be related to the fact that two neighboring intervals were perceived as having very similar durations in these conditions (i. e., temporal assimilation occurred). We will return to this issue in Sect. 21.3.2, but the result indicates that different modes of processing may underlie the discrimination of single and of multiple empty intervals.

Indeed, multiple presentations of sounds could make listeners perceive a particular type of rhythm [21.35, 36]. Moreover, successive sounds could be sequentially grouped by their similarity or proximity [21.37]. It is beyond the purpose of this chapter to discuss the rhythmic and the sequential grouping, but duration discrimination is indeed modulated according to what type of grouping listeners perceive from the succession of sounds. *Trainor* and *Adams* [21.38]

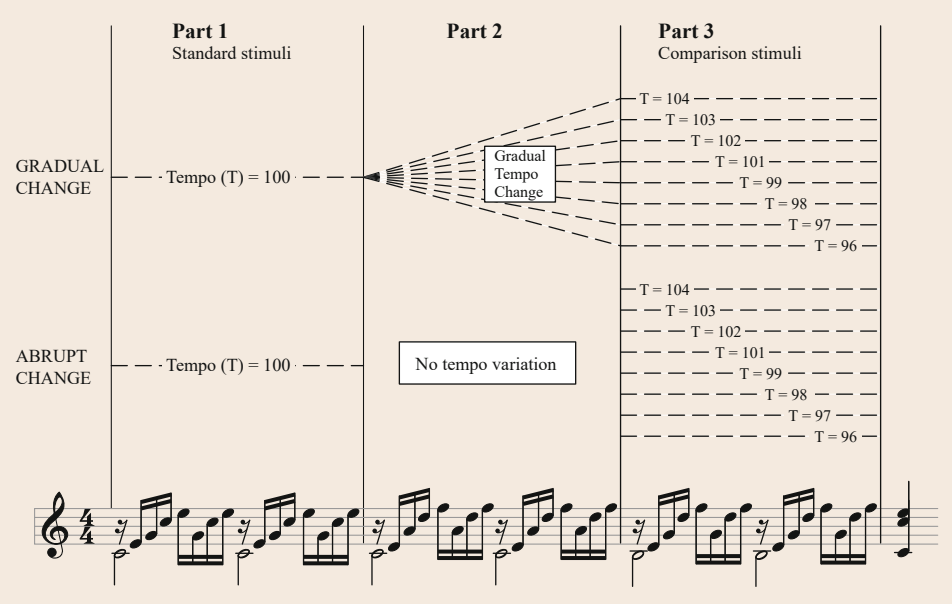


Fig. 21.1 Diagram of experimental conditions. Each excerpt consisted of three parts, with *Parts 1* and *3* being identical for the gradual- and abrupt-variation conditions. *Part 2* involved either a gradual change of musical tempo or an interruption. The excerpt was taken from Bach Prelude No. 1 (Vol. 1) in C Major for well-tempered clavier (after [21.34])

presented the repetition of sound triplets consisting of a 200 ms sound (short–S), a 200 ms sound (S), and a 600 ms sound (long–L), like SSLSSLSSL..., these sounds being separated by interstimulus intervals of 200 ms. The authors found that adult and infant participants were better at detecting an increase in the duration of intervals between SS or between SL compared to those between LS. According to previous studies involved in rhythm [21.39] (see also [21.40]), the pattern SSLSSLSSL... could be perceived as segmented into rhythmic groups of SSL, like [SSL][SSL][SSL]...; in other words, longer sounds are likely to mark the end of rhythmic groups. The SS and SL intervals were within the perceived groups of SSL, while the LS intervals did not belong to any perceived groups. A similar effect was reported by *Geiser* and *Gabrieli* [21.41] with different rhythmic patterns. The sequential grouping by frequency proximity [21.42] and by timbre or loudness similarity [21.43] could also result in a better discrimination of empty intervals within perceived groups than intervals between groups (but see [21.44–46] which failed to find any difference between these types of intervals).

21.2.2 Interval Discrimination in Music and Speech

In most psychophysical research on time-interval discrimination capacity, intervals are presented in well-controlled, but ecologically restricted conditions. In everyday life, however, changes in duration between events are embedded in much more complex contexts.

One of these complex contexts is music. Although temporal changes are essential for establishing the esthetic dimension of music, not much is known about precise sensitivity to slow changes in the context of musical detection. The capacity to detect tempo changes with clicks is about equal to a just noticeable difference of 4% [21.47], and in music, it seems to vary between 5 to 13%, depending on the base tempo under investigation [21.48].

However, in their experiment addressing this tempo change issue, *Grondin* and *Laforest* [21.34] reported a much higher sensitivity for the musical context. They used the method of constant stimuli, and two types of interstimuli intervals (ISI) were compared. As illustrated in Fig. 21.1, a standard musical excerpt was first presented, followed by a comparison excerpt. Compared to the standard, only the tempo (time between notes) was modified (accelerated or decelerated). The ISI conditions were an abrupt tempo change (empty ISI) or a gradual tempo change, i. e., a slow but continuous change of the tempo of the excerpt during the ISI. Sensitivity to tempo changes may indeed depend on the ability acquired through musical training to detect such changes. Therefore, this experiment also included in its design a comparison of performances by musicians and nonmusicians.

This experiment by Grondin and Laforest reveals the great sensitivity of all participants, even the nonmusicians, for discriminating tempi when interval discrimination is embedded within music. For both musician and nonmusician groups, the only difference between Parts 1 and 3 (Fig. 21.1) was the time between notes. There were about 150 ms between notes in the standard excerpt. The Weber fraction (variability/standard tempo) varied from 2.3 to 3.3% with the control group, and was below 1% with the musicians (Fig. 21.2). Within this duration range, sensitivity for discriminating intervals is usually reported to be lower than that. The difference between musicians and nonmusicians is not surprising considering that there seems to be fundamental differences between these groups for processing intervals [21.49]. What is remarkable here is the acute sensitivity of musicians, even when changes are gradual.

Note that the experiment revealed no bias toward easier perception for acceleration or for deceleration (no difference of the bisection point in the different conditions). This result was inconsistent with a portion of the literature on tempo discrimination [21.50, 51]. Greater sensitivity to acceleration or deceleration depends on base tempo [21.52], and the tempo adopted for the experiment by Grondin and Laforest [21.34] was probably too fast.

As for the discrimination of intervals in the context of speech, there are not that many reports describing this fundamental capability. *Quené* [21.53] systematically investigated this question in experiments where participants were presented with a standard sentence and a sentence in which the tempo was accelerated or decelerated. Quené reported just noticeable differences of about 5%, with sensitivity for accelerations being 1 to 2% higher (lower JND) than sensitivity for decelerations. This author also showed that participants' judgments were really based on tempo rather than on other acoustic properties. In Quené's investigation though, the difference between the standard and comparison intervals varied from 5 to 20%. In other words, the 5% reported value may be an overestimation of the real value, i. e., an underestimation of the real capacity to discriminate tempo in speech.

Kato et al. [21.54] also addressed the question of tempo sensitivity in speech but instead of using sentences, they used a series of four-syllable (mora) words, with each syllable consisting of a consonant and a vowel. In their study, the mean word duration was 713 ms, with modifications varying from 2.8 to 11.2%. When entire words were modified, Kato et al. reported a Weber fraction of 3.5%; this value was just above 5.1% when the vowel onsets were specifically modified. Note that Kato et al. also observed better discrimination for accelerations than for decelerations.

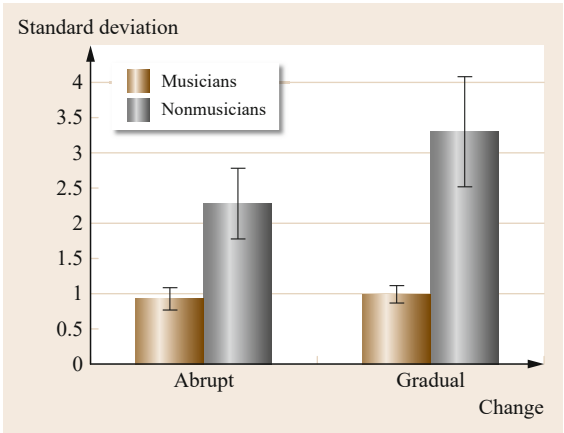


Fig. 21.2 Mean standard deviation for musicians and nonmusicians in abrupt- and gradual-change conditions (error bars correspond to the standard error) (after [21.34])

A recent experiment provided direct comparisons, with the same participants, of performance levels for sentences and tone conditions. *Grondin* et al. [21.55] conducted an experiment with French-speaking participants. In addition to delivering sentences in French, Slovenian (a foreign language participants did not understand) was also used because in French, semantic meaning may have attracted attentional resources. Moreover, during the experiment, certain grouping strategies were imposed on participants, the same in both languages. This temporal patterning of syllables was expected to provide rhythm, and sentences were selected accordingly. In each language, one sentence was divided rhythmically into 4 three-syllable fragments (3–3–3–3: *Il avait demandé à François de l'aider* and *Predite, popoldne, do nase, bicise*). The other sentence was divided into 3 four-syllable fragments (4–4–4: *Emmanuelle est repartie lundi matin* and *Solnce sije, kladvo bije, zgodnje ure*). There were two tone conditions. In one, there was a series of 11 identical empty intervals marked by 12 sounds; and in the other, there were 12 identical filled intervals, with 20 ms between intervals. There were also conditions where empty and filled intervals marked by tones were segmented into the 3–3–3–3 or the 4–4–4 structure.

As illustrated in Fig. 21.3, discrimination was excellent in all conditions, but remained much better in the tone conditions than in the speech conditions: Weber fractions below 2% with tones and around 4.5% in the speech conditions. Observing better discrimination with tones than with sentences reveals that extensive speech training does not improve sensitivity to time interval variations. The results rather indicate that the acoustical variations in speech reduce the discrimination of time interval variations. Keeping nontemporal

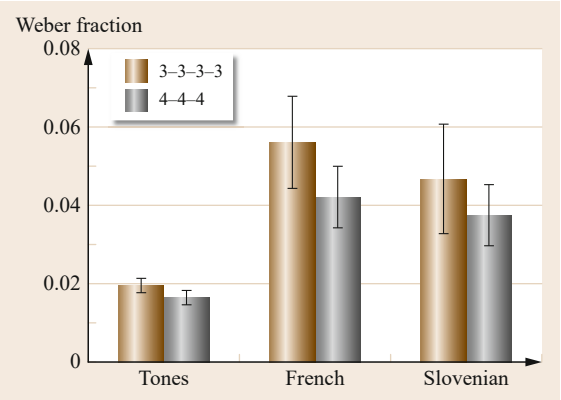


Fig. 21.3 Mean Weber fraction in each rhythm (3–3–3–3; 4–4–4) condition as a function of stimulus condition (tones, French, Slovenian) (after [21.55])

features simple and constant, as is the case with tones but not with speech stimuli, facilitates temporal discrimination. However, knowing or not the semantic meaning of a language does not seem to significantly affect tempo discrimination in a speech context.

This study also revealed that inducing internal subsequences within tone conditions, as opposed to using a series of equal intervals, does not lead to better discrimination. If anything, including rhythmic patterns in sequences of tones led to slightly lower discrimination levels. It is as if it is not possible to access a second level of rhythm for improving discrimination levels. Moreover, discrimination with filled intervals was at least as good as the one with empty intervals. This probably means that the high performance levels were linked to a repetition effect rather than to some rhythmic benefit. Finally, note that there was no significant difference between detecting acceleration versus deceleration.

21.2.3 Comparison of Audition with Other Sensory Modalities

The fact that audition basically constitutes the processing of successive sounds, and that this processing is continuously solicited by the need to understand speech and enjoy music, leads to the prediction that auditory signals must be more efficient for marking time, in a temporal processing task, than signals delivered from other sensory modalities.

There are many pieces of information in the literature for confirming that auditory temporal processing is better than visual or tactile temporal processing (for a review, see [21.56]). It is not the purpose of the present chapter to summarize this literature but the following information should draw a global picture about this issue. It is known for instance that participants are much better at discriminating intervals marked by auditory signals than by visual signals, and this finding applies to filled and empty intervals, and to intervals lasting 125 ms to 4 s [21.15]. In addition, the discrimination of empty intervals is easier with auditory signals than with tactile signals [21.57], and for brief empty intervals, discrimination is easier when signals are delivered from the same modality than from different (intermodal) modalities [21.57–59].

Recently, *Kuroda* et al. [21.60] presented two neighboring empty intervals delimited by three successive signals, each one being a sound (auditory–A) or a flash (visual–V). Participants judged whether the second interval, whose duration was systematically varied, was shorter or longer than the 500 ms first interval. Compared with VVV (three visual stimuli) and AAA (three auditory stimuli), discrimination was impaired for VAV (auditory stimulus between two visual stimuli) but not much for AVA (visual stimulus between two auditory stimuli), a finding consistent with what *Fraisse* [21.61] found with similar stimuli. The study by Kuroda et al. is also consistent with the idea of the superiority of audition (AAA) over vision (VVV) for processing temporal intervals.

Finally, just like there are benefits to expect from multiple presentations of intervals marked by auditory signals, increasing the number of intervals marked by flashes serves to improve performance [21.10], but this benefit may not apply with very brief intervals and may depend on the method adopted for presenting sequences of flashes [21.62].

21.3 Processing Time Intervals: Perceived Duration

Music is made up of multiple sounds of certain durations which are played successively. Perceiving the durations of these sounds and the time intervals created by these sounds is essential for enjoying music. It is said that perceived rhythm is closely related to the relationship between the onsets of successive sounds (e.g., [21.36, 63–65]). This is understandable considering that when an instrument that can produce long sounds, e.g., a violin, and an instrument that can only produce short sounds, e.g., a drum, play the same rhythm together, the sounds coincide somewhere around their onsets (e.g., [21.66], see also [21.67] for perceptual attack time

of musical tones). Accordingly, many of the rhythm perception studies have been conducted focusing on the timing of sound onsets (e.g., [21.68, 69]).

Interestingly, the perceived durations of sounds and time intervals are not always veridical to their physical durations. In this section, perceptual phenomena related to the perceived duration of sounds and time intervals will be explained.

21.3.1 Markers' Length

Sounds used in music, such as sounds of musical instruments and voice, have certain durations. The duration of musical sounds could vary depending on the instrument and the way the sound is played. Some sounds can be about tens or hundreds of milliseconds (e.g., [21.35, 36, 70]), while longer sounds can be more than a second (e.g., [21.36, 71]).

The duration of a sound can be considered as the duration of a filled interval. The perceived duration of a filled interval is said to be longer than that of an empty interval of the same physical duration [21.72–74]). In other words, the duration of a filled interval is overestimated compared to an empty interval. This phenomenon is sometimes called the filled duration illusion [21.73] or the sustained sound illusion [21.75], but note that in some cases, the filled interval which causes the filled duration illusion could refer to a time interval which is marked by two brief sounds like an empty interval but has one or more brief sounds inserted within this interval (e.g., [21.76, 77]). This illusion is a good example demonstrating that the perceived duration of a time interval is influenced by the structure of the interval.

Despite the simplicity of the stimuli, the filled duration illusion shows some complex aspects of time perception. For example, the amount of the illusion, i. e., the amount of overestimation of the filled interval compared to the empty interval, differed between studies. *Zwicker* [21.74] reported that the filled interval was perceived to be as long as an empty interval in which the physical duration was doubled (see also [21.78]). *Craig* [21.72] reported that a silent interval between two sustained sounds needs to be 657 ms longer than the first sound (i. e., a filled interval) to be perceived as having the same duration. *Wearden* et al. [21.73] reported that the perceived duration of an empty interval is only 55–65% of the perceived duration of a filled interval when their physical durations were the same. Moreover, there seemed to be individual differences in the occurrence of the filled duration illusion. In *Hasuo* et al. [21.79], where the subjective durations of empty and filled intervals of 20–180 ms were measured with the method of adjustment, the illusion occurred clearly for some participants but not for many others. Such individual differences were replicated in *Hasuo* et al. [21.80] with time intervals up to 520 ms. These authors also reported that the occurrence of the illusion is influenced by the experimental method.

One of the explanations for the overestimation of a filled interval compared to an empty interval assumes that there is more delay in perceiving the offset of a sound compared to the onset (e.g., [21.72, 81, 82]). In this case, the perceived duration of a filled interval will be longer than that of an empty interval because it will take longer to catch the end of the interval to be judged in the case of filled intervals (Fig. 21.4). Similar arguments were provided by *Fastl* and *Zwicker* [21.78], who schematized temporal changes of auditory-nerve excitation. In the internal-marker hypothesis of *Grondin* [21.15], it is not excluded that a part of the explanation is also located at the beginning of the timekeeping activity. Another explanation is based on the assumption that our internal pacemaker runs faster when there is sound compared to when there is silence (e.g., *Wearden* et al. [21.73] – see Sect. 21.4 for details about the internal-clock model). The same time interval will seem longer when the pacemaker runs faster. Similar explanation is sometimes applied to explain the difference in perceived duration between sensory modalities: when there is an auditory stimulus and a visual stimulus of the same physical duration, the perceived duration of the auditory stimulus will be longer than that of the visual stimulus (e.g., [21.83, 84]), and this phenomenon can be explained by assuming that the sound makes the pacemaker run faster than the light [21.84] (see also [21.56]). These explanations account for the overestimation of a filled interval relative to an empty interval, but given the complexity in the occurrence of the filled duration illusion, it seems difficult to come to a clear conclusion on which of these explanations is more plausible. Rather, it is more likely that the timekeeping process for these time intervals can be modulated depending on the listener and the experimental method [21.80].

Sound duration is not only important for the perception of the sound itself, but also for the perception of the time interval between sounds. *Woodrow* [21.85] showed that when there are two sounds marking a silent time interval of 500 ms between the offset of the first sound and onset of the second sound (offset-onset interval), lengthening either of the two sounds increases the perceived duration of this interval even though the physical duration of this interval was fixed. Similarly, *Grondin* et al. [21.86] showed that an intermodal offset-onset interval (250, 500, and 750 ms) is perceived to be longer when the first or the second marker is 100 ms rather than 5 ms. *Kuroda* et al. [21.60] showed a similar effect of

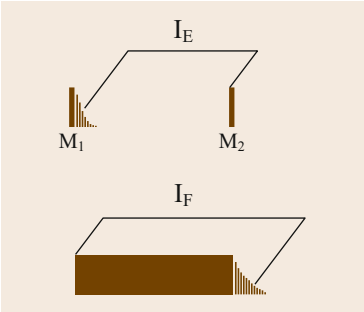


Fig. 21.4 From an internal-marker hypothesis' viewpoint, differences in perceived duration between a filled interval (I_F) and an empty interval (I_E) may depend on differences at the beginning and end of the marking activity. With an empty interval, an observer may wait for the end of the first marker (M_1) before starting to time, but would start timing a filled interval as soon as a signal is detected; and may stop timing as soon as the second marker (M_2) of an empty interval is detected, but would have to wait for the disappearance of the signal marking the time to stop the timekeeping of a filled interval

marker duration by changing the duration of the middle marker in a stimulus sequence of three successive inter- or intramodal markers. *Hasuo* et al. [21.87] observed the influences of sound marker duration, focusing on the time interval between the onsets of two successive sounds (onset-onset interval). Also in this case, the perceived duration of time intervals (120, 240, 360 ms) increased as the second marker lengthened from 20 to 100 ms [21.88]. Lengthening the first marker also increased the perceived duration of the time interval when the time interval was 240 ms or longer (in their Experiment 2), but this effect did not appear in their other experiments involving the duration of the first marker.

Interestingly, sound durations can influence the perceived duration of time intervals even in contexts related to music. *Hasuo* et al. [21.89] observed the effects of sound marker duration using a very simple rhythm pattern consisting of three successive sounds marking two neighboring time intervals by their onsets. In this case, similarly to *Hasuo* et al.'s [21.87] results, lengthening the second sound increased the perceived duration of the first time interval, and lengthening the third sound increased the perceived duration of the second time interval. In other words, a sound whose onset marked the end of the time interval to be judged increased the perceived duration of this interval. *Schubert* and *Fabian* [21.90] found similar effects of sound durations using a part of an actual music piece. They used the dotted rhythm in the opening two bars of Variation 7 of J. S. Bach's Goldberg Variation (BWV988). The rhythmic cell consisted of a dotted quaver (dotted eighth), a semiquaver (sixteenth), and a quaver (eighth), in this order, and the actual dottedness of the rhythm as well as the durations of the tones were varied. They found that shortening the last sound of the rhythmic cell (i. e., changing the quaver to a semiquaver and a semiquaver rest) caused the rhythm to be perceived as more dotted even when the actual dottedness was unaltered. This could be related to *Hasuo* et al.'s [21.89] results in the sense that shortening the sound which terminates the time interval makes the interval seem shorter; shortening the third sound of the rhythmic cell may have caused the second time interval between the second and the third sound to be perceived as shorter, which must have made the contrast between the first and the second time interval greater – making the perceived dottedness clearer.

21.3.2 Interactions Between Intervals

Music is made up of multiple sounds that are played successively. Thus, time intervals often appear in a context of many neighboring time intervals rather than in isolation. When there is more than one time interval, the perceived durations of these intervals can be influenced by one another; research indicates that the subjective duration of time intervals can be affected by preceding or following intervals (e.g., [21.91–94]).

The ratio between neighboring time intervals is important in music. In western music, simple integer ratios such as 1 : 1 or 2 : 1 often appear [21.35, 95], and even when the actual durations of neighboring time intervals deviate from such simple ratios, our perception can be shifted towards these simple ratios [21.68, 92, 96, 97]. One of the phenomena that could be related to the formation of the 1 : 1 perception is an illusion in time perception called *time shrinking* [21.98]. Time shrinking typically takes place in a temporal pattern consisting of three successive sounds marking two neighboring time intervals, T1 and T2, in this order (/T1/T2/; slashes denote short sound markers delimiting T1 and T2). When T1 is slightly shorter than T2, the perceived duration of T2 will be considerably shorter than when it is presented in isolation. A typical temporal condition in which time shrinking takes place is when T1 $\leq$ 200 ms, and maximum time shrinking occurs when T2–T1 is about 80 ms (see [21.77] for a review). Studies showed that time shrinking also appears with visual [21.99] and tactile [21.100, 101] stimuli, but in these cases, the time condition in which time shrinking occurs seems to be wider than in the auditory modality [21.99, 101]. When time shrinking occurs, the physically longer T2 will be shortened perceptually, thus being perceived as having a duration similar to T1. In other words, time shrinking can be considered to be a type of temporal as-

similation between two neighboring durations [21.102]. When T2 is lengthened further and exceeds the range in which time shrinking occurs, the underestimation of T2 disappears suddenly and an overestimation starts to appear [21.103]. In this case, the physically longer T2 will be even more different from T1 – this is probably a contrast between the two neighboring time intervals.

Time shrinking shows the influence of the preceding time interval (T1) on the perception of the following time interval (T2). A question may arise whether the perception of the preceding interval (T1) is influenced by the following interval (T2). Accordingly, *Miyauchi* and *Nakajima* [21.104] measured the subjective durations of both T1 and T2, and compared them to their subjective durations when the time intervals were presented in isolation. When T1 was slightly shorter than T2, T1 was overestimated, and T2 was underestimated; the perceived durations of T1 and T2 approached each other, indicating bilateral assimilation. Such assimilation took place when $-80 \leq$ T1–T2 ≤ 40 ms. The assimilation effect on T1 was much smaller, however, compared to time shrinking on T2. When the difference between T1 and T2 exceeded this range, the assimilation changed to contrast.

The assimilation and contrast between neighboring time intervals are important for the constitution of the 1 : 1 rhythmic category [21.77]. Because of the large effects of time shrinking, the temporal range of the 1 : 1 category is asymmetrical, especially for short durations (e.g., [21.89, 97]). In addition to the behavioral studies, there are electrophysiological studies by *Mitsudo* et al. [21.105, 106], who showed that the auditory temporal assimilation could be associated with brain activities that appeared after the presentation of T2 in the right prefrontal area.

When there are four or more successive sounds marking three or more neighboring time intervals, the perceived duration of the intervals can influence each other in complex ways [21.93].

Another duration illusion in a similar context occurs with filled time intervals. In this illusion, called the *time-stretching illusion*, a sine tone is perceived as longer when it is preceded by a noise than when presented in isolation [21.107]. This illusion also occurs with a band noise preceded by a wider band noise of the same spectral density [21.108]. According to Sasaki et al., the overestimation of the tone duration increases (1) as the noise intensity relative to the tone intensity is increased, (2) as the noise is closer to the tone in frequency, (3) as the tone duration is increased, and (4) as the noise duration is increased up to about 300 ms. However, Sasaki et al., as well as *Kuroda* and *Grondin* [21.109], also reported that some amount of overestimation remained even when the noise was weaker than the tone (see also [21.110]). Sasaki et al. and Carlyon et al. reported that the illusion disappeared when a gap of more than 20 ms was inserted between the two sounds. Finally, Kuroda and Grondin reported that the illusion did not occur when a tone was followed, instead of being preceded, by a noise.

21.3.3 Space, Pitch, and Intensity Effect

Duration processing is sometimes discussed in terms of the relation with space. This issue has attracted attention probably because space and time have to be integrated to perceive motion that is a fundamental feature of perceptual events. How space and time are integrated has been investigated since the early 20th century, and the pioneering works examined how time modulates spatial perception, instead of how space modulates duration perception, mainly with the visual and the tactile modality (see [21.111] for a review). These works reported the *tau effect* where time affects the perception of spatial distance. This phenomenon takes place, for example, with three successive brief flashes, A, B, and C, which are aligned at equal spatial intervals. Two spatial distances, A–B and B–C, are perceived as identical if these flashes are presented at equal time intervals. However, the A–B spatial distance is perceived as shorter than the B–C distance, despite the physical equality of these distances, if the A–B time interval is shorter than the B–C time interval; as well, the A–B spatial distance is perceived as longer than the B–C distance if the A–B time interval is longer than the B–C time interval. In other words, the ratio of two neighboring spatial distances is perceived as similar to that of the corresponding time intervals.

Since then, several studies demonstrated that the opposite effect, i. e., the effect of space on duration perception, could be observed with three visual flashes (e.g., [21.112, 113]). This phenomenon, called the *kappa* effect, was also tested with the auditory modality, whereas most of the studies adopted frequency (pitch) separation instead of spatial distance as an independent variable that would modulate duration perception. For example, *Shigeno* [21.114] used three successive sinusoidal sounds each lasting 200 ms. The initial and the last sounds were fixed at 1000 and 2500 Hz, respectively, while the middle sound's frequency was systematically varied. The relative durations of the first and the second interstimulus time intervals were systematically varied though the total duration was kept at 1000 ms. When the middle sound's frequency was closer to the last sound's frequency than to the initial sound's frequency, the first time interval was likely to be perceived as longer than the second one. Indeed, in this frequency condition, the middle

sound had to be presented before the physical bisection point so that participants perceived the two time intervals as equivalent. Using synthesized vowels /i/ and /e/ as the initial and the last, or the last and the initial sounds, respectively, Shigeno also found that the first time interval was likely to be perceived as longer than the second one when the formant frequencies (vowel quality) of the middle sound were closer to those of the last sound than to those of the initial sound (see also [21.115] using voiced stop consonants).

Note that *Cohen* et al. [21.116] were the ones who first tested the kappa effect with three successive sounds of different frequencies but failed to find clear evidence of the effect. However, they found another effect in the subsequent experiment where two continuous sounds (filled time intervals), instead of empty intervals, were consecutively presented. The first and the second sounds were at 1000 and 3000 Hz, 2000 and 4000 Hz, or 2500 and 2973 Hz, respectively. These sounds were also presented in the opposite order. The total duration of the two sounds was fixed at 1.5 s, while participants adjusted the relative durations of these sounds so that they perceived these durations as equivalent. Participants tended to adjust the first sound to be longer than the second one when the former sound was of lower frequency than the latter. In other words, when two sounds of different frequencies are consecutively presented, the sound of lower frequency is perceived as shorter. The effect was replicated by *Yoblick* and *Salvendy* [21.117].

Grondin and *Plourde* [21.118] examined whether the kappa effect would occur by spatial distance instead of frequency separation (see also [21.119]). Four 20 ms 1 kHz sinusoidal sounds were successively presented, delimiting three neighboring time intervals. The first three sounds were delivered from the same spatial location while the last one was delivered from a different location on a vertical plane. Both the first and the second time intervals were 75, 150, or 225 ms, while the last time interval, which was compared with the preceding ones by participants, was systematically varied. Increasing spatial distance between the two locations (from 1.1 to 2.2 m or from 2.2 to 4.4 m) resulted in an overestimation of the last time interval, except when the preceding intervals lasted 75 ms. This overestimation could be regarded as the occurrence of the kappa effect, but disappeared when participants knew where the sounds would be presented before each trial (one spatial condition per session) instead of when they were uncertain of the locations (all spatial conditions were mixed within one session).

The kappa effect is explained by the imputed-velocity model, positing the constancy of motion speed or velocity [21.77, 111]. According to this model, the kappa-effect (three-stimuli) pattern is perceived as consisting of a single object appearing three times, instead of three discrete objects appearing successively. This single object is perceived as passing through space with constant speed. Speed constancy is physically kept when three stimuli are presented at equal temporal intervals and at equal spatial distances, but not kept if the middle stimulus is made close to the initial or to the last one in space, when time intervals are kept equal. Because this speed inconstancy is re-adjusted by the perceptual system, the first time interval is perceived as shorter than the second one when the middle stimulus is close to the initial one; as well, the first time interval is perceived as longer than the second one when the middle stimulus is close to the last one. The single object, therefore, is perceived as moving between three spatial locations with constant speed.

Henry and *McAuley* [21.120] examined whether this model could apply to three successive sounds of different fundamental frequencies, ranging from 329.63 to 523.25 Hz. The key point of their experiments was the fact that the interonset interval between the initial and the last sounds were varied, this interval being 728, 1000, or 1600 ms. When the whole duration was shorter, pitch was perceived as changing faster during the stimulus pattern, thus strengthening the impression of motion. Indeed, the acceleration of the pitch velocity magnified the kappa effect. This is consistent with what was expected on the basis of the idea that the kappa effect is caused by the perception of motion implied by three successive signals. However, *Alards-Tomalin* et al. [21.121] demonstrated the occurrence of the kappa effect by the intensity separation (i. e., with three successive sounds of different intensities) but reported that the imputed-velocity model could not apply to this kappa effect. In their experiment, three additional sounds were presented before the kappa-effect (three-sound) pattern in order to extend the ascending or descending intensity change of the kappa-effect pattern. These additional sounds were expected to strengthen the impression of motion but indeed reduced, instead of magnifying, the kappa effect.

Note that the imputed-velocity model does not help to predict any spatial effects on the perception of single time intervals delimited by two signals. Indeed, the presentation of only two signals does not imply whether speed is constant or not. There are not many studies examining spatial effects on the perception of single time intervals, even with the visual [21.122–125] and the tactile modality [21.126–128]. However, *Roy* et al. [21.129] reported that single intervals lasting around 125 or 250 ms were perceived as shorter when they were marked by two sounds 3.3 m apart than when marked by two sounds 1.1 m apart from each other. In

other words, with single time intervals, longer spatial distance resulted in shorter perceived duration.

Since the frequency of two consecutive sounds affects their perceived duration as mentioned earlier (see also [21.130] with single intervals), it might be possible that sound intensity (loudness) also affects the perceived duration of filled time intervals. Several studies have investigated how intensity affects the perceived duration of single filled intervals but the results seemed inconsistent [21.131, 132]. However, *Matthews* et al. [21.133] indicated that perceived duration is affected by the relative intensity of signal and background sounds, instead of the absolute intensity of signals. In their experiment, a standard sound of 600 ms and a comparison sound of variable duration were successively presented. A background sound of about 42 (quiet) or 72 dBA (loud) was presented before the standard, between the standard and the comparison, and after the comparison. The standard and the comparison were at 46 and 68 dBA, or 68 and 46 dBA, respectively. In the quiet background, the probability that participants judged the comparison as longer than the standard was increased when the comparison intensity (68 dB) was higher than the standard one (46 dB) compared with when the comparison intensity (46 dB) was lower than the standard (68 dB). In the loud background, the probability of judging the comparison as longer was decreased when the comparison intensity was higher than the standard one compared with when the comparison intensity was lower than the standard. In other words, louder comparisons were perceived as longer when the standard and the comparison intensity was above the (quiet) background intensity but were perceived as shorter when the standard and the comparison intensity was below the (loud) background intensity.

21.4 Theoretical Perspectives

In the present portion of the chapter, the focus is given to a theoretical issue that is important to consider in the perspective of a handbook on music, and neuroscientific explanations are kept to a minimum. For a more complete view on the main theoretical proposals based on neuroscience, the reader is invited to consult some recent reviews [21.134–138].

One classical way of approaching the question of the mechanism(s) involved in temporal processing is to distinguish duration-based and beat-based timing [21.62, 139]. It is actually an interval-based mechanism that is proposed in the traditional internal (single, central) clock perspective originally proposed by *Creelman* [21.140] and *Treisman* [21.141]. This clock is described as a pacemaker emitting pulses whose accumulation in a counter determines the impression about time, more pulses accumulated resulting in longer perceived duration. Errors when estimating time intervals could be attributed to the properties of the pacemaker's rate of emission [21.27], and it is reasonable to posit that the counter is not error free [21.142]. Another source of errors in such a timing process is related to attention [21.143, 144]. A time interval is marked by one or more sensory signals, and fluctuations in attentional processes may change the beginning and end of a timekeeping period marked by the signals [21.145]. This fact is often referred to as a switch process. As well, if another task is to be performed in parallel during the timekeeping operation, the experience of time will be transformed. The attention devoted to a nontemporal task is not available for keeping track of time. When less attention is allocated to time, fewer pulses are accumulated and, consequently, time is perceived as being shorter [21.146, 147]. To explain this fact, *Zakay* and *Block* [21.148] posited that attention is under the control of a gate mechanism. Moreover, depending on the type of temporal tasks to be completed, memory and decisional processes are processing levels that are likely to be involved, and therefore, potential sources of errors. That is what is argued in the information processing version of the scalar expectancy theory [21.149].

Instead of a process based on a single interval presentation, sequences of sensory signals lead to a different level of discrimination, as we have seen earlier, and therefore, probably provide a different source of information on which temporal judgments can be made. This possibility is known as the beat-based hypothesis. Within such a view, the efficiency of timing is based on the expectation generated by the predictability of the arrival of the next signal (or stimulus). Indeed, this efficiency is based, according to *Jones* [21.150, 151] (see also [21.64]) on the synchronization (attunement) of some internal oscillatory activity with environmental stimuli having some coherence or predictability. In terms of Jones' *Dynamic attending theory*, this process is called a future-oriented attending mode.

The multiple presentations of intervals not only generate a beat but also provide an opportunity to stabilize the representation of an interval in memory, which should reduce variance in a duration discrimination process for instance. This perspective is known as the multiple-look model [21.17]. This model assumes that

increasing the number of standard intervals (models) presented improves the representation of this standard in memory. This model is supported by the data showing that increasing the number of standard presentations actually improves performance (up to a certain number of presentations), but not by the fact that the number of standard and the number of comparison intervals both influence performance, with more presentations in each case improving sensitivity to time [21.10, 12, 13].

Returning to the duration-based versus beat-based distinction, neuroscientific investigations now provide evidences that the role of the cerebellum, at least for the processing of subsecond intervals, differs according to the type of temporal processing required. Indeed, both mechanisms are likely available, but rely on distinct neural substrates [21.152, 153]. Moreover, there are probably important individual differences in the way these mechanisms are used. *Grahn* and *McAuley* [21.154] make a distinction between people on the basis of their way of extracting an implicit periodic beat. Indeed, their functional magnetic resonance imaging study helped to distinguish between *strong beat perceivers* and *weak beat perceivers*. When the two groups are compared, the former one has greater neural activity in the supplementary motor area (SMA), left premotor cortex, and left insula, but the neural activity is greater in the participants for the latter group in the right premotor cortex, the left posterior superior and left middle temporal gyri.

21.5 Conclusion

Experimental psychologists have been interested in testing a variety of auditory temporal experiences. The capacity to make an explicit judgment about the duration of sounds, or of an interval between sounds, has received a great deal of research attention. This capacity is crucial because the experience of audition, in speech or music for instance, is essentially temporal.

It is known that sensitivity to time is better in the auditory modality than in other modalities, and is increased when multiple, instead of single, intervals are presented. However, under specific conditions, the repetition of intervals may generate distortions like the impression that the last interval of a sequence is briefer (a time-shrinking effect). Having space between sound sources, or changing frequencies between successive sounds, also has an impact on perceived duration. It is also generally recognized that filled durations are perceived as longer than empty durations, but this effect depends on the method used and on individual differences.

Implicit judgments about time are traditionally reported to depend on an internal clock. This clock is usually described as a pacemaker-counter device, the experience of time relying on the accumulation by the second component of pulses emitted by the first component. Some recent findings in neuroscience reveal that there could be different systems for judging time, some people relying more spontaneously than others on the possibility of using a system based on the processing of beat.

References

21.1 G.B. Vicario: Temporal displacement. In: *The Nature of Time: Geometry, Physics and Perception*, ed. by R. Buccheri, M. Saniga, M. Stuckey (Kluwer, Dordretch 2003) pp. 53–66

21.2 I.J. Hirsh, C.E. Sherrick: Perceived order in different sense modalities, J. Exp. Psychol. **62**, 423–432 (1961)

21.3 M. Wittmann: Moments in time, Front. Integr. Neurosci. **5**, 66 (2011)

21.4 H. Babkoff, L. Fostick: The role of tone duration in dichotic temporal order judgment, Atten. Percept. Psychophys. **75**, 654–660 (2013)

21.5 J.E. Roeckelein: History of conceptions and accounts of time and early time perception research. In: *Psychology of Time*, ed. by S. Grondin (Emerald, Bingley 2008) pp. 1–50

21.6 S. Grondin: Methods for studying psychological time. In: *Psychology of Time*, ed. by S. Grondin (Emerald, Bingley 2008) pp. 51–74

21.7 G. Mioni, F. Stablum, S.M. McClintock, S. Grondin: Different methods for reproducing time, different results, Atten. Percept. Psychophys. **76**(3), 675–681 (2014)

21.8 R. Bartolo, L. Prado, H. Merchant: Information processing in the primate basal ganglia during sensory-guided and internally driven rhythmic tapping, J. Neurosci. **34**, 3910–3923 (2014)

21.9 A.M. Wing, A.B. Kristofferson: Response delay and the timing of discrete motor responses, Percept. Psychophys. **14**, 5–12 (1973)

21.10 S. Grondin, J.D. McAuley: Duration discrimination in crossmodal sequences, Perception **38**, 1542–

1559 (2009)
21.11 T. Kuroda, E. Hasuo: The very first step to start psychophysical experiments, Acoust. Sci. Technol. **35**, 1–9 (2014)
21.12 J.D. McAuley, N.S. Miller: Picking up the pace: Effects of global temporal context on sensitivity to the tempo of auditory sequences, Percept. Psychophys. **69**, 709–718 (2007)
21.13 N.S. Miller, J.D. McAuley: Tempo sensitivity in isochronous tone sequences: The multiple-look model revisited, Percept. Psychophys. **67**, 1150–1160 (2005)
21.14 D. Getty: Discrimination of short temporal intervals: A comparison of two models, Percept. Psychophys. **18**, 1–8 (1975)
21.15 S. Grondin: Duration discrimination of empty and filled intervals marked by auditory and visual signals, Percept. Psychophys. **54**, 383–394 (1993)
21.16 T.H. Rammsayer: Differences in duration discrimination of filled and empty auditory intervals as a function of base duration, Atten. Percept. Psychophys. **72**, 1591–1600 (2010)
21.17 C. Drake, M.-C. Botte: Tempo sensitivity in auditory sequences: Evidence for a multiple-look model, Percept. Psychophys. **54**, 277–286 (1993)
21.18 A. Friberg, J. Sundberg: Time discrimination in a monotonic, isochronic sequence, J. Acoust. Soc. Am. **98**, 2524–2531 (1995), https://doi.org/10.1121/1.413218
21.19 N. Ehrlé, S. Samson: Auditory discrimination of anisochrony: Influence of the tempo and musical backgrounds of listeners, Brain Cogn. **58**, 133–147 (2005), https://doi.org/10.1016/j.bandc.2004.09.014
21.20 A. Halpern, C. Darwin: Duration discrimination in a series of rhythmic events, Percept. Psychophys. **31**, 86–89 (1982)
21.21 I.J. Hirsh, C.B. Monahan, K.W. Grant, P.G. Singh: Studies in auditory timing: 1. Simple patterns, Percept. Psychophys. **47**, 215–226 (1990)
21.22 P.G. Cheatham, C.T. White: Temporal numerosity: III. Auditory perception of number, J. Exper. Psychol. **47**, 425–428 (1954)
21.23 Y. Nakajima: A model of empty duration perception, Perception **16**, 485–520 (1987)
21.24 G. ten Hoopen, S. van den Berg, J. Memelink, B. Bocanegra, R. Boon: Multiple-look effects on temporal discrimination within sound sequences, Attent. Percept. Psychophys. **73**, 2249–2269 (2011)
21.25 S. Grondin: Violation of the scalar property for time perception between 1 and 2 seconds: Evidence from interval discrimination, reproduction, and categorization, J. Exp. Psychol. Hum. Percept. Perform. **38**, 880–890 (2012)
21.26 L.A. Bizo, J.Y.M. Chu, F. Sanabria, P.R. Killeen: The failure of Weber's Law in time perception and production, Behav. Process. **71**, 201–210 (2006)
21.27 S. Grondin: From physical time to the first and second moments of psychological time, Psychol. Bull. **127**, 22–44 (2001), https://doi.org/10.1037/0033-2909.127.1.22
21.28 S. Grondin: About the (non)scalar property for time perception. In: *Neurobiology of Interval Timing*, ed. by H. Merchant, V. de Lafuente (Springer, New York 2014) pp. 17–32
21.29 S. Grondin, G. Meilleur-Wells, R. Lachance: When to start explicit counting in time-intervals discrimination task: A critical point in the timing process of humans, J. Exp. Psychol. Hum. Percept. Perform. **25**, 993–1004 (1999)
21.30 S. Grondin, B. Ouellet, M.-È. Roussel: Benefits and limits of explicit counting for discriminating temporal intervals, Can. J. Exp. Psychol. **58**, 1–12 (2004)
21.31 S. Grondin, P.R. Killeen: Tracking time with song and count: Different Weber functions for musicians and non-musicians, Atten. Percept. Psychophys. **71**, 1649–1654 (2009)
21.32 S. Grondin, P.R. Killeen: Effects of singing and counting during successive interval productions, NeuroQuantology **7**, 77–84 (2009)
21.33 T. Sasaki, Y. Nakajima, G. ten Hoopen: Categorical rhythm perception as a result of unilateral assimilation in time-shrinking, Music Percept. **16**, 201–222 (1998)
21.34 S. Grondin, M. Laforest: Discriminating slow tempo variations in a musical context, Acoust. Sci. Technol. **25**, 159–162 (2004)
21.35 S. Handel: *Listening: An Introduction to the Perception of Auditory Events* (MIT Press, Cambridge 1989)
21.36 A.D. Patel: *Music, Language, and the Brain* (Oxford Univ. Press, New York 2008)
21.37 A.S. Bregman: *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge 1990)
21.38 L.J. Trainor, B. Adams: Infants' and adults' use of duration and intensity cues in the segmentation of tone patterns, Percept. Psychophys. **62**, 333–340 (2000)
21.39 T.L. Bolton: Rhythm, Am. J. Psych. **6**, 145–238 (1894)
21.40 J.R. Iversen, A.D. Patel, K. Ohgushi: Perception of rhythmic grouping depends on auditory experience, J. Acoust. Soc. Am. **124**, 2263–2271 (2008)
21.41 E. Geiser, J.D.E. Gabrieli: Influence of rhythmic grouping on duration perception: A novel auditory illusion, PloS ONE **8**, e54273 (2013), https://doi.org/10.1371/journal.pone.0054273
21.42 P.J. Fitzgibbons, A. Pollatsek, I.B. Thomas: Detection of temporal gaps within and between perceptual tonal groups, Percept. Psychophys. **16**, 522–528 (1974)
21.43 L.A. Thorpe, S.E. Trehub: Duration illusion and auditory grouping in infancy, Dev. Psychol. **25**, 122–127 (1989)
21.44 T. Kuroda, E. Hasuo, S. Grondin: Discrimination of brief gaps marked by two stimuli: Effects of sound length, repetition, and rhythmic grouping, Perception **42**, 82–94 (2013)
21.45 D.L. Neff, W. Jesteadt, E.L. Brown: The relation between gap discrimination and auditory stream segregation, Percept. Psychophys. **31**, 493–

501 (1982)
21.46 T. Kuroda, E. Tomimatsu, S. Grondin, M. Miyazaki: Perceived empty duration between sounds of different lengths: Possible relation with repetitionand rhythmic grouping, Atten. Percept. Psychophys. **78**, 2678–2689 (2016)
21.47 G. Madison: Detection of linear temporal drift in sound sequences: Empirical data and modelling principles, Acta Psychol. **117**, 95–118 (2004)
21.48 M.C. Ellis: Thresholds for detecting tempo change, Psychol. Music **19**, 164–169 (1991), https://doi.org/10.1177/0305735691192007
21.49 B.H. Repp, H.B. Mendlowitz, M.J. Hove: Does rapid auditory stimulation accelerate an internal pacemaker? Don't bet on it, Timing Time Percept. **1**, 65–76 (2013)
21.50 J.D. McAuley, G.R. Kidd: Effect of deviations from temporal expectations on tempo discrimination of isochronous tone sequences, J. Exp. Psychol. Hum. Percept. Perform. **24**, 1786–1800 (1998)
21.51 D. Sheldon: Effects of tempo, musical experience, and listening modes on tempo modulation perception, J. Res. Music Educ. **42**, 190–202 (1994)
21.52 P.G. Vos, M. van Assen, M. Franek: Perceived tempo change is dependent on base tempo and direction of change: Evidence for a generalized version of Schulze's (1978) internal beat model, Psychol. Res. **59**, 240–247 (1997)
21.53 H. Quené: On the just noticeable difference for tempo in speech, J. Phonetics **35**, 353–362 (2007)
21.54 H. Kato, M. Tsuzaki, Y. Sagisaka: Functional differences between vowel onsets and offsets in temporal perception of speech: Local-change detection and speaking-rate discrimination, J. Acoust. Soc. Am. **113**, 3379–3389 (2003)
21.55 S. Grondin, N. Bisson, C. Gagnon: Sensitivity to time interval changes in speech and tone conditions, Atten. Percept. Psychophys. **73**, 720–728 (2011)
21.56 S. Grondin: Sensory modalities and temporal processing. In: *Time and Mind II: Information Processing Perspectives*, ed. by H. Helfrich (Hogrefe Huber, Goettingen 2003) pp. 75–92
21.57 S. Grondin, R. Rousseau: Judging the relative duration of multimodal short empty time intervals, Percept. Psychophys. **49**, 245–256 (1991)
21.58 E. Gontier, E. Hasuo, T. Mitsudo, S. Grondin: EEG investigations of duration discrimination: The intermodal effect is induced by an attentional bias, PLoS ONE **8**(8), e74073 (2013)
21.59 S. Grondin, M.-E. Roussel, P.-L. Gamache, M. Roy, B. Ouellet: The structure of sensory events and the accuracy of time judgments, Perception **34**, 45–58 (2005)
21.60 T. Kuroda, E. Hasuo, K. Labonté, V. Laflamme, S. Grondin: Discrimination of two neighboring intra- and intermodal empty time intervals marked by three successive stimuli, Acta Psychol. **149**, 134–141 (2014)
21.61 P. Fraisse: La perception de la durée comme organisation du successif. Mise en évidence expérimentale, Annee Psychol. **52**, 39–46 (1952)
21.62 S. Grondin: Discriminating time intervals presented in sequences marked by visual signals, Percept. Psychophys. **63**, 1214–1228 (2001)
21.63 S. Handel: The effect of tempo and tone duration on rhythm discrimination, Percept. Psychophys. **54**, 370–382 (1993)
21.64 E. Large: Resonating to musical rhythm: Theory and experiment. In: *Psychology of time*, ed. by S. Grondin (Emerald, Bingley 2008) pp. 189–232
21.65 S. McAdams, C. Drake: Auditory perception and cognition. In: *Stevens' Handbook of Experimental Psychology: Sensation and Perception*, Vol. 1, ed. by S. Yantis (Wiley, New York 2002) pp. 424–432
21.66 R.A. Rasch: Synchronization in performed ensemble music, Acustica **43**, 121–131 (1979)
21.67 J.W. Gordon: The perceptual attack time of musical tones, J. Acoust. Soc. Am. **82**, 88–105 (1987)
21.68 P. Desain, H. Honing: The formation of rhythmic categories and metric priming, Perception **32**, 341–365 (2003)
21.69 B.H. Repp: Perception of timing is more context sensitive than sensorimotor synchronization, Percept. Psychophys. **64**, 703–716 (2002)
21.70 R.A. Rasch: *Aspects of the Perception and Performance of Polyphonic Music*, Dissertation (University of Groningen, Groningen 1981)
21.71 R.M. Warren: *Auditory Perception: An Analysis and Synthesis*, 3rd edn. (Cambridge Univ. Press, Cambridge 2008)
21.72 J.C. Craig: A constant error in the perception of brief temporal intervals, Percept. Psychophys. **13**, 99–104 (1973)
21.73 J.H. Wearden, R. Norton, S. Martin, O. Montford-Bebb: Internal clock processes and the filled-duration illusion, J. Exp. Psychol. Hum. Percept. Perform. **33**, 716–729 (2007)
21.74 E. Zwicker: Subjektive und objektive Dauer von Schallimpulsen und Schallpausen (Subjective and objective duration of sound impulses and sound pauses), Acustica **22**(70), 214–218 (1969)
21.75 B.H. Repp, R.J. Marcus: No sustained sound illusion in rhythmic sequences, Music Percept. **28**, 121–133 (2010)
21.76 E.A.C. Thomas, I. Brown Jr.: Time perception and the filled-duration illusion, Percept. Psychophys. **16**, 449–458 (1974)
21.77 G. ten Hoopen, R. Miyauchi, Y. Nakajima: Time-based illusions in the auditory mode. In: *Psychology of Time*, ed. by S. Grondin (Emerald, Bingley 2008) pp. 139–187
21.78 H. Fastl, E. Zwicker: *Psychoacoustics: Facts and Models* (Springer, Berlin 2007)
21.79 E. Hasuo, Y. Nakajima, K. Ueda: Does filled duration illusion occur for very short time intervals?, Acoust. Sci. Technol. **32**, 82–85 (2011)
21.80 E. Hasuo, Y. Nakajima, E. Tomimatsu, S. Grondin, K. Ueda: The occurrence of the filled duration illusion: A comparison of the method of adjustment with the method of magnitude estimation, Acta Psychol. **147**, 111–121 (2014)
21.81 R. Efron: The relationship between the duration of a stimulus and the duration of a perception,

Neuropsychologia **8**, 37–55 (1970)
21.82 R. Efron: The minimum duration of a perception, Neuropsychologia **8**, 57–63 (1970)
21.83 J.T. Walker, K.J. Scott: Auditor – Visual conflicts in the perceived duration of lights, tones and gaps, J. Exp. Psychol. Hum. Percept. Perform. **7**, 1327–1339 (1981)
21.84 J.H. Wearden, H. Edwards, M. Fakhri, A. Percival: Why "sounds are judged longer than lights": Application of a model of the internal clock in humans, Q. J. Exp. Psychol. **51B**, 97–120 (1998)
21.85 H. Woodrow: Behavior with respect to short temporal stimulus forms, J. Exp. Psychol. **11**, 167–193 (1928)
21.86 S. Grondin, R.B. Ivry, E. Franz, L. Perreault, L. Methé: Markers' influence on the duration discrimination of intermodal intervals, Percept. Psychophys. **58**, 424–433 (1996)
21.87 E. Hasuo, Y. Nakajima, S. Osawa, H. Fujishima: Effects of temporal shapes of sound markers on the perception of interonset time intervals, Atten. Percept. Psychophys. **74**, 430–445 (2012)
21.88 E. Hasuo, Y. Nakajima, M. Wakasugi, T. Fujioka: Effects of sound marker durations on the perception of inter-onset time intervals: A study with instrumental sounds, Jap. J. Psychon. Sci. **34**, 2–16 (2015)
21.89 E. Hasuo, Y. Nakajima, Y. Hirose: Effects of sound-marker durations on rhythm perception, Perception **40**, 220–242 (2011)
21.90 E. Schubert, D. Fabian: Perception and preference of dotted 6/8 patterns by experienced and less experienced baroque music listeners, J. Music Percept. Cogn. **7**, 113–132 (2001)
21.91 M. Franěk, J. Mates, T. Radil, K. Beck, E. Pöppel: Sensorimotor synchronization: Motor responses to pseudo-regular auditory patterns, Percept. Psychophys. **55**, 204–217 (1994)
21.92 D.J. Povel: Internal representation of simple temporal patterns, J. Exp. Psychol. Hum. Percept. Perform. **7**, 3–18 (1981)
21.93 T. Sasaki, D. Suetomi, Y. Nakajima, G. ten Hoopen: Time-shrinking, its propagation, and Gestalt principles, Percept. Psychophys. **64**, 919–931 (2002)
21.94 H.H. Schulze: The perception of temporal deviations in isochronic patterns, Percept. Psychophys. **45**, 291–296 (1989)
21.95 P. Fraisse: Rhythm and tempo. In: *The Psychology of Music*, ed. by D. Deutsch (Academic, New York 1982) pp. 149–180
21.96 P.J. Essens, D.J. Povel: Metrical and nonmetrical representations of temporal patterns, Percept. Psychophys. **37**, 1–7 (1985)
21.97 R. Miyauchi, Y. Nakajima: The category of 1:1 ratio caused by assimilation of two neighboring empty time intervals, Hum. Mov. Sci. **26**, 717–727 (2007)
21.98 Y. Nakajima, G. ten Hoopen, G. Hilkhuysen, T. Sasaki: Time-shrinking: A discontinuity in the perception of auditory temporal patterns, Percept. Psychophys. **51**, 504–507 (1992)
21.99 H. Arao, D. Suetomi, Y. Nakajima: Does time-shrinking take place in visual temporal patterns?, Perception **29**, 819–830 (2000)
21.100 E. Hasuo, T. Kuroda, S. Grondin: About the time-shrinking illusion in the tactile modality, Acta Psychol. **147**, 122–126 (2014)
21.101 J.B.F. Van Erp, M.M.A. Spapé: Time-shrinking and the design of tactons. In: *Haptics: Perception, Devices and Scenarios*, Vol. 5024, ed. by M. Ferre (Springer, Berlin, Heidelberg 2008) pp. 289–294
21.102 Y. Nakajima, G. ten Hoopen, T. Sasaki, K. Yamamoto, M. Kadota, M. Simons, D. Suetomi: Time-shrinking: The process of unilateral temporal assimilation, Perception **33**, 1061–1079 (2004)
21.103 Y. Nakajima, E. Hasuo, M. Yamashita, Y. Haraguchi: Overestimation of the second time interval replaces time-shrinking when the difference between two adjacent time intervals increases, Front. Hum. Neurosci. **8**, 281 (2014)
21.104 R. Miyauchi, Y. Nakajima: Bilateral assimilation of two neighboring time intervals, Music Percept. **22**, 411–424 (2005)
21.105 T. Mitsudo, Y. Nakajima, G.B. Remijn, H. Takeichi, Y. Goto, S. Tobimatsu: Electrophysiological evidence of auditory temporal perception related to the assimilation between two neighboring time intervals, NeuroQuantology **7**, 114–127 (2009)
21.106 T. Mitsudo, Y. Nakajima, H. Takeichi, S. Tobimatsu: Perceptual inequality between two neighboring time intervals defined by sound markers: Correspondence between neurophysiological and psychological data, Front. Psychol. **5**, 937 (2014), https://doi.org/10.3389/fpsyg.2014.00937
21.107 T. Sasaki, Y. Nakajima, G. ten Hoopen, E. van Buuringen, B. Massier, T. Kojo, T. Kuroda, K. Ueda: Time stretching: Illusory lengthening of filled auditory durations, Atten. Percept. Psychophys. **72**, 1404–1421 (2010)
21.108 R.P. Carlyon, J.M. Deeks, Y. Shtyrov, J. Grahn, H.E. Gockel, O. Hauk, F. Pulvermüller: Changes in the perceived duration of a narrowband sound induced by a preceding stimulus, J. Exp. Psychol. Hum. Percept. Perform. **35**, 1898–1912 (2009)
21.109 T. Kuroda, S. Grondin: No time-stretching illusion when a tone is followed by a noise, Atten. Percept. Psychophys. **75**, 1811–1816 (2013)
21.110 K. Ueda, M. Ohtsuki: The effect of sound pressure level difference on filled duration extension, J. Acoust. Soc. Jpn. (E) **17**, 159–161 (1996)
21.111 B. Jones, Y.L. Huang: Space-time dependencies in psychophysical judgment of extent and duration: Algebraic models of the tau and kappa effect, Psychol. Bull. **91**, 128–142 (1982)
21.112 S. Abe: Experimental study on the co-relation between time and space, Tohoku Psychologica Folia **3**, 53–68 (1935)
21.113 J. Cohen, C.E.M. Hansel, J.D. Sylvester: A new phenomenon in time judgment, Nature **172**, 901 (1953)
21.114 S. Shigeno: The auditory tau and kappa effects for speech and nonspeech stimuli, Percept. Psychophys. **40**, 9–19 (1986)

21.115 S. Shigeno: The auditory tau and kappa effects for voiced stop consonants, J. Psychol. Res. **29**, 71–80 (1987)

21.116 J. Cohen, C.E.M. Hansel, J.D. Sylvester: Interdependence of temporal and auditory judgments, Nature **174**, 642–644 (1954)

21.117 D.A. Yoblick, G. Salvendy: Influence of frequency on the estimation of time for auditory, visual, and tactile modalities: The kappa effect, J. Exp. Psychol. **86**, 157–164 (1970)

21.118 S. Grondin, M. Plourde: Discrimination of time intervals presented in sequences: Spatial effects with multiple auditory sources, Hum. Mov. Sci. **26**, 702–716 (2007)

21.119 J.-C. Sarrazin, M.-D. Giraudo, J.B. Pittenger: Tau and kappa effects in physical space: The case of audition, Psychol. Res. **71**, 201–218 (2007)

21.120 M.J. Henry, J.D. McAuley: Evaluation of an imputed pitch velocity model of the auditory kappa effect, J. Exp. Psychol. Hum. Percept. Perform. **35**, 551–564 (2009)

21.121 D. Alards-Tomalin, L.C. Leboe-McGowan, T.A. Mondor: Examining auditory kappa effects through manipulating intensity differences between sequential tones, Psychol. Res. **77**, 480–491 (2013)

21.122 S. Grondin: Judgments of the duration of visually marked empty time intervals: Linking perceived duration and sensitivity, Percept. Psychophys. **60**, 319–330 (1998)

21.123 D.R. Price-Williams: The kappa effect, Nature **173**, 363–364 (1954)

21.124 M.-E. Roussel, S. Grondin, P. Killeen: Spatial effects on temporal categorization, Perception **38**, 748–762 (2009)

21.125 T. Kuroda, S. Grondin, M. Miyazaki, K. Ogata, S. Tobimatsu: The kappa effect with only two visual markers, Multisens. Res. **29**, 703–725 (2016)

21.126 S. Grondin, T. Kuroda, T. Mitsudo: Spatial effects on tactile duration categorization, Canad. J. Exp. Psychol. **65**, 163–167 (2011)

21.127 T. Kuroda, S. Grondin: Discrimination is not impaired when more cortical space between two electro-tactile markers increases perceived duration, Exp. Brain Res. **224**, 303–312 (2013)

21.128 T. Kuroda, M. Miyazaki: Perceptual versus motor spatiotemporal interactions in duration reproduction across two hands, Sci. Rep. **6**, 23365 (2016)

21.129 M. Roy, T. Kuroda, S. Grondin: Effect of space on auditory temporal processing with a single-stimulus method. In: *Advances in Sound Localization*, ed. by P. Strumillo (InTech, Rijeka 2011) pp. 95–104

21.130 H. Burghardt: Die subjektive Dauer schmalbandiger Schalle bei verschiedenen Frequenzlagen (The subjective duration of narrow-band sound at different frequencies), Acustica **28**, 278–284 (1973)

21.131 B. Berglund, U. Berglund, G. Ekman, M. Frankehaeuser: The influence of auditory stimulus intensity on apparent duration, Scand. J. Psychol. **10**, 21–26 (1969)

21.132 H. Eisler, A.D. Eisler: Time perception: Effects of sex and sound intensity on scales of subjective duration, Scand. J. Psychol. **33**, 339–358 (1992)

21.133 W.J. Matthews, N. Stewart, J.H. Wearden: Stimulus intensity and the perception of duration, J. Exp. Psychol. Hum. Percept. Perform. **37**, 303–313 (2011)

21.134 M.J. Allman, W.H. Meck: Pathophysiological distortions in time perception and timed performance, Brain **135**, 656–677 (2012)

21.135 A. Gorea: Ticks per thought or thoughts per tick? A selective review of time perception with hints on future research, J. Physiol. **105**, 153–163 (2011)

21.136 S. Grondin: Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions, Att. Percept. Psychophys. **72**, 561–582 (2010)

21.137 F. Macar, F. Vidal: Timing processes: An outline of behavioural and neural indices not systematically considered in timing models, Can. J. Exp. Psychol. **63**, 227–239 (2009)

21.138 H. Merchant, D. Harrington, W.H. Meck: Neural basis of the perception and estimation of time, Annu. Rev. Neurosci. **36**, 313–336 (2013)

21.139 S.W. Keele, R. Nicoletti, R. Ivry, R.A. Pokorny: Mechanisms of perceptual timing: Beat-based or interval-based judgements?, Psychol. Res. **50**, 251–256 (1989)

21.140 C.D. Creelman: Human discrimination of auditory duration, J. Acoust. Soc. Am. **34**, 582–593 (1962)

21.141 M. Treisman: Temporal discrimination and the indifference interval: Implications for a model of the "internal clock", Psychol. Monogr. **77**(13), 1–31 (1963)

21.142 P.R. Killeen, T. Taylor: How the propagation of error through stochastic counters affects time discrimination and other psychophysical judgments, Psych. Rev. **107**, 430–459 (2000)

21.143 S. Grondin, F. Macar: Dividing attention between temporal and nontemporal tasks: A performance operating characteristic – POC – analysis. In: *Time, Action, Cognition: Towards Bridging the Gap*, ed. by F. Macar, V. Pouthas, W. Friedman (Kluwer, Dordrecht 1992) pp. 119–128

21.144 F. Macar, S. Grondin, L. Casini: Controlled attention sharing influences time estimation, Memory Cogn. **22**, 673–686 (1994)

21.145 S. Grondin, T. Rammsayer: Variable foreperiods and temporal discrimination, Q. J. Exp. Psychol. **56A**, 731–765 (2003)

21.146 S.W. Brown: Time and attention: A review of the literature. In: *Psychology of Time*, ed. by S. Grondin (Emerald, Bingley 2008) pp. 111–138

21.147 S. Tobin, N. Bisson, S. Grondin: An ecological approach to prospective and retrospective timing of long durations: A study involving gamers, PLoS ONE **5**(2), e9271 (2010)

21.148 D. Zakay, R.A. Block: Temporal cognition, Curr. Dir. Psychol. Sci. **6**, 12–16 (1997)

21.149 J. Gibbon, R.M. Church, W.H. Meck: Scalar timing in memory, Annals New York Acad. Sci. **423**, 52–77 (1984)

21.150 M.R. Jones, M.G. Boltz: Dynamic attending and responses to time, Psychol. Rev. **96**, 459–491

(1989)
21.151 E.W. Large, M.R. Jones: The dynamics of attending: How we track time varying events, Psychol. Rev. **106**, 119–159 (1999)
21.152 S. Teki, M. Grube, T.D. Griffiths: A unified model of time perception accounts for duration-based and beat-based timing mechanisms, Front. Integr. Neurosci. **5**, 91–97 (2012)
21.153 S. Teki, M. Grube, S. Kumar, T.D. Griffiths: Distinct neural substrates of duration-based and beat-based auditory timing, J. Neurosci. **31**, 3805–3812 (2011)
21.154 J.A. Grahn, J.D. McAuley: Neural bases of individual differences in beat perception, Neuroimage **47**, 1894–1903 (2009)

22. Automatic Processing of Musical Sounds in the Human Brain

Elvira Brattico, Chiara Olcese, Mari Tervaniemi

This chapter introduces neurophysiological evidence on the dissociation between unconscious and conscious aspects of musical sound perception. The focus is on research conducted with the event-related potential (ERP) technique, which allows chronometric investigation of information-processing stages during music listening. Findings suggest that automatic processes are confined to the auditory cortex and might even involve the discrimination of deviations from simple musical scale rules. In turn, voluntary, cognitive processes, likely originating from the inferior prefrontal cortex, are necessary to understand more complex musical rules, such as tonality and harmony. The implications of understanding how and to what extent music is processed below the level of consciousness are discussed in rehabilitation and therapeutic settings.

22.1 Perceiving the Music Around Us: An Attentive or Automatic Process?

In psychology, perception of music is typically conceptualized as a conscious, attentive process. Auditory attention is what enables us to process sound information from the world and is also defined as a spotlight on sounds around us to make them stand out. When a sound or sound sequence stands out, we notice and process it. Attention can change rapidly, switching from one thing to another. For instance, attention can be steered by our intentions (*top–down*) or by features of sounds in the environment (*bottom–up*) [22.1]. Hence, outside the focus of attention, when our intentions are diverted elsewhere, we might be attracted by an object or a sound in the environment, which is then processed further and consciously. Outside the focus of attention, preattentive or automatic processes help us decide what to pay attention to and what to filter out [22.2].

Hillyard [22.3, p. 180] wrote:

> *A stimulus set preferentially admits all sensory input to an attended channel* [...] *for further perceptual analysis while blocking or attenuating input arriving over irrelevant channels* [...] *at an early stage of processing.*

According to early filter theories from which the statement by Hillyard was inspired [22.4, 5], after early sensory-memory stages, characterized by parallel and fast sensorial (nonintelligent) analysis of the different stimulus features, there are higher level stages of information processing leading to feature integration and conscious perception. These processes are controlled by top–down voluntary attention and they permit cognitive operations on sounds. Additional neural resources, particularly in the parietal lobe, midbrain and pulvinar, are allocated to selective stimuli of the environment, allowing only relevant information to be further processed by the cerebral cortex [22.6–9]. Subsequently, the neurophysiological signal from the auditory channel is amplified (*Posner* [22.9]; for brain-imaging evidence concerning the auditory system, see [22.10, 11]).

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_22

To study the chronometric succession of neural activity going from fast automatic to slow attentive processes, the most appropriate method is the event-related potential (ERP; also more generally, sometimes referred to as evoked response), which measures the neural responses to a stimulus or a mental act by averaging those portions of an electroencephalography (EEG) or magnetoencephalography (MEG) signal that are locked to each stimulus or act presentation (termed trial) in the experimental session. As compared with other brain research methods, such as functional magnetic resonance imaging (fMRI) or positron emission tomography (PET), that measure the slow metabolic changes in oxygenated hemoglobin and other molecule consumption deriving from neural activity, the ERP method allows one to study with millisecond accuracy the different stages of neural sound processing [22.6, 12, 13]. The ERP data, deficient in spatial resolution but refined in latency information, can then be combined with anatomical and functional knowledge to infer the localization of the recorded neural activity. ERP responses or components to sounds are typically measured as the voltage potential between a reference electrode and the electrode placed at the scalp vertex as early as within the first tens of milliseconds from sound onset.

With regard to neurocognitive processes involved in music listening, the intervention of voluntary attention allows the sophisticated analysis of music according to the high-level hierarchical rules of a musical system, such as those of Western tonal music [22.14, 15]. The attribution of meaning to musical sounds is also supposed to take place during the attentional stage of pitch processing [22.16]. However, while attentional processes are important to understand complex musical rules and likely to attribute musical meaning, the way they are studied with brain research methods, with attentive tasks on the sound stimuli of interest, can be regarded as artificial compared with the everyday experience of music listening [22.17]. Relevant to this, the most common way of consumption of music, amounting to about 76% of all music consumption, consists in hearing music in the background while attending to a primary activity [22.18]. Thus, in order to extract the neural processing specific to the encoding of musical pitch relations, allowing generalization to the most typical music consumption situation, it is preferable to adopt a paradigm in which subjects direct their attention away from the sounds. While this experimental procedure does not bear on all of the music consumption situations in which we find ourselves, it nonetheless includes most of them, at the same time permitting one to avoid contaminations from top–down processes related to attention.

22.2 The MMN as a Measure of Automatic Sound Processing in the Auditory Cortex

The mismatch negativity or MMN [22.19] is an ERP component, which permits one to probe auditory brain processes without the need of attentive listening. The MMN is measured with EEG as a negative wave over frontocentral scalp regions and a potential reversal at mastoidotemporal regions, peaking at around 100–200 ms in response to *deviant* sound stimuli inserted in a sequence of repetitive *standard* stimuli. The potential reversal at the mastoidal sites suggests the generation of the MMN in the bilateral auditory cortex, which is located in the supratemporal lobe. Typically, the MMN indexes the automatic change detection capacity in the auditory cortex and the auditory processing accuracy in the human brain.

The MMN is considered an automatic brain response since it is elicited even when the subject is focused on a primary (e.g., visual) task and not attending to the auditory stimulation. Furthermore, the MMN is observed even during sleep, in coma patients and in newborn infants, confirming its independence from neural processes related to awareness [22.20]. Nevertheless, many studies show that the attention-independent MMN recording can be used as a prediction of the individual discrimination accuracy in behavioral and sensory tasks that use different kinds of sounds and sound-features, such us simple isolated tones changing in pitch [22.21–23] and complex spectro-temporal patterns changing in their contour [22.24].

The MMN depends on the presence of a neural memory trace of a brief duration accurately representing each simple and complex feature of the standard stimulus, including its temporal aspects, and on a neurally distinct mechanism comparing the incoming stimulus with the stored trace [22.25]. Relative to this, *Tervaniemi* et al. [22.26] observed an MMN following the omission of the second tone of a tone pair only when the interstimulus interval was shorter than 200 ms [22.27, 28]. This finding demonstrated that the memory trace for the standard stimulus has a duration, called the temporal window of integration. The MMN is also described to reflect the cerebral cortex function of

predicting the upcoming events by continuously updating a model of the (auditory) environment [22.29–31]. This brain capacity of executing neural matches and mismatches between consecutive sounds at a preattentive level is a fundamental survival tool. It regulates adaptive behavior, allowing humans to detect potentially dangerous inputs.

Classically, the MMN is measured in experiments using the one-deviant *oddball* paradigm: each sequence of repetitive stimuli contains unexpected variations (10–20% occurrence) differing in only one physical feature from the *standard*. The eliciting changes in the auditory input include frequency, pitch, intensity, duration, spatial location and rhythm, as well as more complex features, such as the presence of one phoneme instead of another or an unexpected auditory pattern violating a rule of the auditory sequence. More recently, a *multifeature* paradigm has been introduced, consisting of several deviant sounds deviating from the standard in one feature each (50% occurrence) in the same sound sequence [22.32]. This allows the recording of several MMNs for different auditory features in less than 15 min. The *musical multifeature* paradigm was adapted from the *multifeature* one to mimic a musical context, by integrating the feature changes in 4-tone patterns selected from the pitches of the chromatic scale and modulated into different keys and modes [22.2]. Also a *melody multifeature* paradigm, with changes in melodic contour, key, timbre, rhythm, and tuning has been implemented [22.33, 34]. These fast multifeature paradigms permit one to study automatic musical processing skills even in children and clinical populations, using experiments that are not uncomfortable in terms of duration [22.35–37].

By performing experiments with different stimulation intervals, it has been estimated that the memory system indexed by the MMN lasts almost 10 s [22.38, 39]. This has traditionally been the duration of the long-term auditory sensory memory [22.40]. Furthermore, the so-called *backward-masking* effect has to be considered: *Winkler* and *Näätänen* [22.41] observed that the MMN is elicited only when the *intertone interval* between the first and second tone of a pair is long enough (150–400 ms). The MMNs seem to also reflect the long-term memory, for example in the case of mother-tongue phonemes or musical scale sounds (as will be described in the following sections). Therefore, the MMN is not only a measure of auditory discrimination accuracy but also an index to investigate the higher level mechanisms of neuroplasticity [22.42]. In fact, based on MMN evidence, new views of the human auditory system have been developed: instead of one-way flow of information from short-term memory storage towards long-term memory, we now consider the memory system to also include information flow from the long-term memory towards short-term subsystems [22.43].

22.3 Neural Generators of the MMN

The neural activity underlining the MMN depends on different generators located in different brain areas. They integrate their activity to create a network that works to detect violations and evaluate unexpected stimuli. The MMN generators have been studied by fMRI [22.44–51], PET [22.52–55], and optical imaging (OI [22.56, 57]) methods. The first phase of the MMN mechanism, the detection of a change in the auditory space, originates in the superior temporal gyrus, namely in the nonprimary auditory cortex. Indeed, the auditory cortex is the area mainly involved in the MMN and it is fundamental to creating the short-term memory trace underlying it, as well as to estimating the degree of its violation by the incoming sound [22.58].

The second phase of the MMN consists in the orientation of attention toward sound deviants automatically processed in the auditory cortex [22.59, 60]. The inferior frontal cortex (mainly in the right hemisphere) is the neural generator of the second MMN phase. The frontal activation is delayed with respect to the auditory-cortex processing [22.61], supporting the hypothesis that the auditory cortex change-detection process activates the frontal attention-orientation mechanism [22.20, 25, 62, 63].

The MMN elicited by variations of pure sinusoidal tones originates in the auditory cortex, where different feature changes, such us frequency, intensity, and duration are processed in distinct areas [22.2, 64]; this is analogous to data from MEG in *Levänen* et al. [22.65, 66]. The generation sources of MMNm (magnetic mismatch negativity) responses to a change in different features of complex sound information were found to be differently located. This suggests that distinct supratemporal neuronal populations exist, each involved in the processing of a specific sound change and probably located in different portions of the auditory cortex [22.67]. *Tervaniemi* et al. [22.68] investigated the location and activation of phonetic versus musical sound processing. They presented the subjects with sequences of frequent and infrequent phonemes (/e/ and /o/) or chords (A major and A minor) and observed that the MMN in the right hemi-

sphere was stronger when elicited by an infrequent chord change than by a phoneme change, while it did not significantly differ within the left hemisphere. Furthermore, the MMNm sources for a phoneme or a chord change were found to be differently located in spatially distinct cortical areas, with the phoneme processing areas superior to those for the chord change in both hemispheres. In a subsequent fMRI study, pseudowords and musical sounds were found to activate distinct parts of the auditory as well as thalamic areas [22.69]. Thus, in healthy adult listeners, auditory areas seem to have a highly specialized structure to encode sounds of different complexity and informational content.

22.4 The MMN for Studying Automatic Processing of Simple Musical Rules

Using the MMN one can study whether the auditory cortex can process a violation of a rule-based sound sequence even outside of attentional focus. For instance, *Tervaniemi* et al. found that a significant MMN is elicited by an ascending tone or a repeating tone randomly presented in descending tone sequences [22.70]. It was concluded that not only physical stimulus properties, but also its abstract features are encoded in a sensory memory trace and automatically compared at a preattentive level [22.71, 72]. The MMN elicited by rule violations was defined as an abstract-feature MMN indicating how physical and abstract deviations are encoded even preattentively from the standard stimulation [22.71, 73–78].

In a more musical context, a local rule of Western tonal music is represented by the chromatic scale, which identifies the 12 pitches of the scale as the basic elements for composing a musical piece. As *Levitin* and *Tirovolas* [22.79] remarked, *composers sometimes reward and sometimes violate listener expectations [22.80, p. 216] but do so within this system of legal tones for their culture's music*. *Brattico* et al. [22.81] showed that a violation of the rules of the chromatic scale elicits an MMN in the auditory cortex at the preattentive level. In this case, the prediction for the incoming sound is based not on a repetitive sound heard before but on the relations between sounds, which are governed by the musical scale [22.81]. The ERPs were measured in participants with minimal musical education while they were presented with unfamiliar melodies, containing out-of-tune or out-of-key violations in a passive condition (reading a book) or an attentive condition (they rated the congruity or incongruity of the melodies). An early frontal MMN-like negativity was elicited in response to both pitch violations, but it was stronger for the out-of-tune than for the out-of-key pitches, independently of the subjects' attention. These results showed that the musical scale features are automatically processed and compared with the representations for the musical scale pitches already at the level of the auditory cortex (although a second generator was also noticed in the inferior prefrontal cortex). These findings evidence *intelligent* cognitive processes related to (simple) musical rule extraction already at the preattentive level [22.81, 82].

Furthermore, two important dichotomies governing Western music – major versus minor and consonance versus dissonance – have also been studied using MMN. *Brattico* et al. [22.83] found that two-tone intervals characterized by a simple frequency ratio (such as consonant intervals: octave, major fifth) are more easily discriminated than intervals characterized by complex frequency ratios (such as dissonant intervals: minor sevenths, minor second). She and her colleagues also showed with magnetoencephalographic recordings that in professional musicians or students from music academies (as opposed to nonmusicians) the magnetic MMN (MMNm) originated from the bilateral auditory cortex and was enhanced specifically in response to nonconventional chords, such as dissonant or mistuned chords. In contrast, their MMNm did not significantly differ from that of nonmusicians in response to minor chords, which are more frequently encountered in Western tonal music. The study pointed at a chord-selective neuroplasticity of the auditory cortex dependent on the exposure to a musical culture. *Virtala* et al. [22.84] further tested the MMN of healthy nonmusicians exposed to four different chords, major, dissonant, minor and inverted major, composed of the same tones and transposed to 12 different levels but with a different tone order. The minor and dissonant chords among major chords elicited an MMN while the inverted major chord did not. This finding, thoroughly investigated also in newborn infants, adolescents, and adult musicians [22.85–87] suggests that the knowledge of traditional music categories is implicitly learned from exposure to a musical culture and consequently preattentively encoded from an early age. The MMN enhancement in musicians demonstrates that explicit training in music might further facilitate and strengthen these processes.

22.5 ERAN as an Index of Semiautomatic Processing of Musical Rules

The violation of the less-local, hierarchical rule of harmony, dictating how chords should succeed each other (for instance, that a musical piece is always ended by a tonic chord preceded by a dominant chord), elicits a different brain response, the early right anterior negativity (ERAN) originating selectively from distinct anatomical regions of the frontal lobe, namely the inferior frontal gyrus (BA44; Brodmann area 44) and the ventral premotor cortex [22.88]. In *Brattico* et al. [22.81] the MMN to the violation of the musical scale was comparable in attended and unattended conditions, showing the automaticity of the neural process of neural representation of that basic musical rule and of the discrimination of its violation. In contrast, the ERAN process, underlying the neural representation of chord succession rules, is only partially automatic, since it is typically elicited when subjects pay attention to deviant timbres inserted in the musical sequences. Moreover, while in *Koelsch* et al. [22.89] an ERAN was elicited even when subjects were distracted from the sound stimulation by a primary task unrelated to the sounds, there was no ERAN to the chords in the middle of the cadence but only to the final ones, which are more salient. Similarly, in *Loui* et al. [22.90] the ERAN was larger when subjects were paying attention to the sounds as opposed to when they were distracted by a reading comprehension task. The formal knowledge of music enhances the neural representations of musical rules as indexed by the enhancement of the ERAN in musicians as opposed to nonmusicians [22.91]. In relation to this, when musical expertise is complex, such as in the case of Finnish folk musicians trained to perform pieces of various musical cultures, sometimes not even containing chords typical of Western tonal harmony (such as the Neapolitan chord not much used in Finnish folk music), the ERAN response is affected and diverges from that of typical listeners [22.92].

22.6 Environmental Exposure Modulates the Automatic Neural Representations of Musical Sounds

The brain modifies its neural activity to adapt to the environmental demands [22.93]. This process is most visible during childhood and adolescence, when it intermingles with developmental processes such as neuronal apoptosis, mielinization and overall body growth, but it exists to a certain extent even in adulthood. For instance, the exposure to repetitive sounds modifies plastically the auditory-cortex representations in order to implement their autonomic processing, saving neural resources, and to quickly discriminate the deviant incoming sound events. The MMN indexes this plastic process in the auditory domain. For instance, an enhanced MMN was elicited in response to infrequent voices familiar to the subjects (a close relative or a long-term friend of each subject recorded their utterance of the vowel *a*) as compared to infrequent unfamiliar ones [22.94], or to an infrequent tongue-clucking sound versus a similar but unfamiliar sound [22.95].

The MMN sensibility to familiar sounds might also imply that musicians, exposed everyday to musical sounds and sequences, would show a larger MMN response to any musical sound than nonmusicians, and particularly to those musical sounds to which they are most accustomed. This would indicate that the musicians' auditory cortex has an increased preattentive ability to discriminate violations of musical sound features and perhaps even of musical rules (which could be viewed as instances of sound sequence features). This hypothesis was confirmed by data from various studies [22.29, 96–98]. In particular, *Koelsch* et al. [22.96] presented major chords and single tones to both professional violinists and nonmusicians, in both passive and active conditions. Slightly mistuned chords elicited an MMN only in professional musicians, but not in nonmusicians. In contrast, the musicians and nonmusicians did not differ in their MMN elicited by sinusoidal tones. These results indicate that the sensory memory function can be trained and modulated, however, this modulation being specific to musical sounds.

This neural sensitivity to encode particular music features in a highly specialized manner has prompted research in which musicians with different practice history are compared. Such an approach is also motivated by the fact that musicians' perceptual and motor skills depend highly on the level of expertise as well as on the instrument, practice strategies, and musical genre. The *musical multifeature* paradigm defined by *Vuust* et al. [22.2] allowed one to study systematically and in a time-efficient manner the influence of a specific musical training on musicians' skills. In this paradigm, six types of acoustic changes relevant for different musical genres (pitch, mistuning, intensity, timbre, sound-source location, and rhythm) were presented in the same sound sequence including 4-tone

patterns resembling Alberti's bass. A pitch slide typical for improvisational (jazz) music (less present in classical music) was inserted in order to compare the neural responses of musicians with those of other musicians playing different musical genres [22.99]. The MMN was measured in three groups of musicians playing classical, jazz, or rock/pop music; also a group of nonmusicians was included. The jazz musicians had the larger MMN amplitude across the six different features, indicating a greater overall sensitivity to sound feature violations, but particularly to pitch-related feature deviants, such as the pitch slide. These data suggest that not only the musical training, but also the characteristics of the genre of music played by musicians might increase and influence their auditory cortex preattentive skills [22.99]. However, one has to keep in mind that the advanced sound discrimination skills of musicians might reflect either an innate prerequisite or an effect of music practice. In fact, superior performance in musicality tests is also reflected as enhanced MMN responses [22.21, 100]. Thus, longitudinal studies using the MMN need to be conducted in order to determine the sources of individual abilities in music perceptual skills.

Recently, Tervaniemi's group obtained pioneering evidence on the effects of music training in sound discrimination accuracy [22.101] in a semilongitudinal setting with the majority of the children involved in the study being followed up on many times over several years. In their study, children aged 7–13 years were presented with a multifeature paradigm (with deviances in pitch, perceived location, intensity, and duration; additionally a gap was placed within the sounds) and with a major–minor chord paradigm. Half of them were actively involved in musical training and received individual teaching in an instrument as well as lessons on music theory on a weekly basis, while half were active in other hobbies but not in any music-related ones. When the recordings were conducted for the first time at the onset of the musical training at the age of 7 years, the groups did not differ in any MMN parameters in the multifeature or chord paradigms. However, during the accumulated training in music until the age of 13, the MMN grew faster in the musically active children than in the musically inactive children. This observation was specific to the chord paradigm; in the multifeature paradigm, a similar statistically nonsignificant trend was observed for the location deviance alone. This result extends to childhood previous observations made with adult musicians [22.33, 99, 102] according to which music training can selectively modulate sound discrimination of those sound features and sound types that are most common in a certain musical culture and/or most salient in a particular instrumental practice, without being generalized to all sounds and sound types.

The selectivity of auditory-cortex discrimination skills does not, however, exclude the possibility of a cross-modal transfer of those skills to other domains. It has been found that musicians have superior automatic discrimination skills for acoustic features of language sounds, as a consequence of musical training, as the features involved are those spectral (pitch) features relevant also in music [22.103]; even the brainstem frequency-following response that indexes neuronal phase locking to the F0 acoustic cue – of a pitch note or a phoneme – is enhanced in musicians compared to nonmusicians [22.104, 105]. This spectral sound discrimination superiority dependent on musical training and thus on a *highly trained auditory system* has also been noticed in the discrimination of speech sounds, namely vowels and temporal variations of consonant-vowel syllables [22.106], although whether this is strictly a result of musical training or of inborn musicality skills [22.107] remains to be determined in future investigations.

22.7 Disrupted Automatic Discrimination of Musical Sounds

In the population, it is not uncommon to find individuals with self-reported or objectively suboptimal abilities for musical perception. Peretz et al. identified these individuals as congenital amusics or tone-deaf by means of a listening test, the Montreal battery for evaluation of amusia [22.108], combined with other perceptual and cognitive tests that isolate the music deficit from any other behavioral or cognitive difficulty. Scoring below two standard deviations from the mean obtained with the normal population is the commonly used criterion for a diagnosis of congenital amusia. The percentage of amusics among the general population has been estimated to be around 4, hence similar to other developmental disorders, such as dyslexia.

Neurophysiological and structural evidence shows that congenital amusic individuals have an abnormal connectivity between the superior temporal gyrus and the inferior prefrontal cortex, namely between the two brain regions that allow the processing of musical rules and their violations [22.109]. The first study reporting related functional brain evidence was by *Peretz* et al. [22.110]. In that study, eight amusic individuals

and ten controls were measured with the ERP technique while they listened to patterns of five repeated tones, in which the fourth tone was replaced 30% of the time, displaced either upward or downward by a quartertone to 3 semitones. In an attentive condition, the most evident ERP components reflecting the conscious detection of a sound change are typically the N200 and the P300. In amusic individuals the N200–P300 complex was enhanced for large pitch changes, whereas for quartertone and semitone changes the P300 was significantly reduced, as compared with controls (for whom the N200 was not even visible). Remarkably, the N100, an earlier obligatory ERP component reflecting initial sensory encoding of any incoming sound, did not distinguish the two experimental groups, confirming that the origin of the neural deficit for music perception can be traced down to the later, conscious stages of auditory processing [22.110, 111]. Seemingly discrepant results were obtained by *Albouy* et al. [22.109] in an MEG study using a working memory task where nine amusic individuals and nine control subjects were asked to detect if two 1.5 sec melodies of six tones, separated by a 2 sec silent interval, were the same or different. The change, presented in half of the melodies, could appear in the second to fifth tone of the second melody. The N100 response to the first tone in the melodies was similar in amusic and normal individuals, replicating previous findings, whereas the N100 response to the subsequent changed tones was diminished in amusic individuals as compared with controls (with a later enhanced response in amusic individuals, due to a delayed N100 latency). The authors claimed that they traced the music-specific problem down to the level of the auditory cortex. However, the neural generator of the N100 was identified in the inferior frontal gyrus, and indeed the neural response was obtained in a task involving conscious and effortful retrieval from working memory of musical material. Behavioral data obtained in the same study (replicating other findings: e.g., *Tillmann* et al. [22.112]) confirms a specific working memory deficit for pitch in amusic individuals, which is reflected in the N100 response and hints at a later stage of processing than the auditory-cortex one indexed by these findings.

The possibility that preattentive auditory-cortex skills might be preserved in amusia was further explored in a study by *Peretz* et al. [22.113]. They found even more surprising evidence about the dissociation between preattentive skills of amusic individuals and their conscious perception of musical pitch. Amusic individuals showed an early negative ERP response (MMN or N200) to mistuned pitches, which were at a quartertone interval from the preceding sound, when those pitches were inserted in unfamiliar tonal melodies, but not to out-of-key pitches. In contrast, amusics did not show the later positive ERP response (P600) to mistuned or out-of-key pitches, which were observed in controls, suggesting the presence in amusic individuals of a neural dysfunction for cognitively extracting and becoming aware of the musical violation, in spite of their preattentive ability to discriminate it. The authors commented on the findings as follows [22.113]:

> *This early neural information, typically computed in the right auditory cortex* [...] *appears, however, insufficient to make contact or to build schemas about musical keys for conscious detection of violations in the right inferior frontal region.*

The neurophysiological studies evidencing residual neural reactions to wrong pitch in congenital amusia have encouraged attempts for rehabilitation towards compensating the neurogenetic vulnerability. For instance, Peretz's group investigated 10–13-year-old children with a marked deficit in perceiving small pitch intervals (but with normal general intelligence [22.114]). The children were asked to listen to music for at least 30 min per day during a period of 4 weeks. Such passive music intervention, though, did not have any effect on the neural or behavioral correlates of the defective pitch processing in amusic children, leading the researchers to conclude that *musical deprivation in childhood does not seem to be the cause of congenital amusia* [22.114]. However, the authors left open the question as to whether active and motivational musical training would produce any positive rehabilitative effects on amusic children.

Recently, altered musical sound processing has also been studied in cochlear implant (CI) users. A CI is a device comprising a microphone, a speech processor and a transmitter coil, for conversion of acoustic signals into electric pulses. The CI is implanted above the basilar membrane of the inner ear in patients suffering from severe and profound hearing loss, allowing them to substantially increase speech intelligibility, especially if implanted early in life, but with limited improvements in regard to music perception, particularly in the domains of timbre and pitch [22.115, 116]. Using the fast *multifeature* paradigm *Torppa* et al. [22.36] showed the presence of an MMN to sound features in prelingually deaf children with a CI, which was diminished as compared with control children but only in timbre, sound duration, and gap. Diminished MMN amplitudes to timbre and pitch deviants inserted in a *musical multifeature* paradigm were also obtained with postlingually deaf CI adults [22.117], and even with 14-year-old CI adolescents who had been prelingually deaf [22.37].

Remarkably, in the latter experiment the CI adolescents followed an intensive active ear-training program lasting two weeks, and this short-term intensive program was sufficient to modify the neural responses to rhythm deviants, showing the potentiality of a musical rehabilitation program for young CI users.

22.8 Conclusions

The findings presented above highlight the role of preattentive neural mechanisms in music perception, as indexed by the MMN. Indeed, a study with stroke patients showed that the increase in the magnetic MMN to pitch changes correlated with the recovery of cognitive functions, such as behavioral improvement in the story recall and mental subtraction tasks [22.51]. This and other evidence presented also in the above text depicts preattentive auditory skills as prerequisites for more complex operations, which require cognitive functions.

The MMN literature reviewed, nevertheless, also evidenced dissociation between unconscious and conscious aspects of musical sound perception, suggesting that voluntary, cognitive processes, likely originating from the inferior prefrontal cortex, are necessary to understand more complex musical rules, such as tonality and harmony. This neural dissociation points at the presence of subliminal sound events never reaching the level of awareness which results in a behavioral music-specific deficit, but at the same time represent a hope for rehabilitation (such as for amusic individuals and CI users). In the auditory domain, then, the concept of unconscious by *Jung* still holds and describes our phenomena well [22.118, p. 964]:

> *The unconscious contains all those psychic events which, because of the lack of the necessary intensity of their functioning, are unable to pass the threshold which divides the conscious from the unconscious; so that they remain in effect below the surface of the conscious, and flit by in subliminal phantom forms.*

Subliminal effects of music on cognitive activities or even emotional responses in another modality have been proved, and the general mood state of subjects can be improved without realizing it while being exposed to relaxing music in the background without paying attention to it [22.119, 120]. Hence, understanding how and to what extent music is processed below the level of consciousness has potential implications not only for applications of music in rehabilitation of perceptual and cognitive functions but also in therapeutic settings or for our wellbeing in everyday life.

References

22.1 J.R. Anderson: *Cognitive Psychology and its Implications* (Worth, Duffield 2004)

22.2 P. Vuust, E. Brattico, E. Glerean, M. Seppänen, S. Pakarinen, M. Tervaniemi, R. Näätänen: New fast mismatch negativity paradigm for determining the neural prerequisites for musical ability, Cortex **47**(9), 1091–1098 (2011)

22.3 S.A. Hillyard, R.F. Hink, V.L. Schwent, T.W. Picton: Electrical signs of selective attention in the human brain, Science **182**, 180 (1973)

22.4 U. Neisser: *Cognitive Psychology: CT, US* (Appletion-Century-Crofts, East Norwalk 1967)

22.5 A.M. Treisman, G. Gelade: A feature-integration theory of attention, Cogn. Psychol. **12**(1), 97–136 (1980)

22.6 R. Näätänen: *Attention and Brain Function* (Lawrence Erlbaum Associates, Hillsdale 1992)

22.7 N. Cowan: *Attention and Memory: An Integrated Framework* (Oxford Univ. Press, New York 1995)

22.8 J.T. Coull: Neural correlates of attention and arousal: insights from electrophysiology, functional neuroimaging and psychopharmacology, Prog. Neurobiol. **55**(4), 343–361 (1998)

22.9 M. Posner: Neuropsychology: Modulation by instruction, Nature **373**(6511), 198–199 (1995)

22.10 L. Jäncke, S. Mirzazade, N.J. Shah: Attention modulates activity in the primary and the secondary auditory cortex: A functional magnetic resonance imaging study in human subjects, Neurosci. Lett. **266**(2), 125–128 (1999)

22.11 C.I. Petkov, X. Kang, K. Alho, O. Bertrand, E.W. Yund, D.L. Woods: Attentional modulation of human auditory cortex, Nat. Neurosci. **7**(6), 658–663 (2004)

22.12 T.W. Picton, A. Durieux-Smith: Auditory evoked potentials in the assessment of hearing, Neurol. Clin. **6**(4), 791–808 (1988)

22.13 M.D. Rugg, M.G.H. Coles (Eds.): *Electrophysiology of Mind: Event-Related Brain Potentials and Cognition* (Oxford Univ. Press, Oxford 1995)

22.14 C.L. Krumhansl, P. Toivanen, T. Eerola, P. Toiviainen, T. Järvinen, J. Louhivuori: Cross-cultural music cognition: Cognitive methodology applied to north sami yoiks, Cognition **76**(1), 13–58 (2000)

22.15 B. Snyder: *Music and Memory: An Introduction* (MIT Press, Cambridge 2000)

22.16 D. Schön, M. Besson: Audiovisual interactions in music reading. A reaction times and event-related potentials study, Ann. N.Y. Acad. Sci. **999**, 193–198 (2003)

22.17 V.J. Schmithorst: Separate cortical networks involved in music perception: Preliminary functional MRI evidence for modularity of music processing, Neuroimage **25**(2), 444–551 (2005)

22.18 A.J. Lonsdale, A.C. North: Why do we listen to music? A uses and gratification analysis, Br. J. Psychol. **102**(1), 108–134 (2011)

22.19 R. Näätänen, A.W. Gaillard, S. Mäntysalo: Early selective-attention effect reinterpreted, Acta Psychologica **42**, 313–329 (1978)

22.20 R. Näätänen: Mismatch negativity: Clinical research and possible applications, Int. J. Psychophysiol. **48**, 179–188 (2003)

22.21 H. Lang, T. Nyrke, M. Ek, O. Aaltonen, I. Raimo, R. Näätänen: Pitch discrimination performance and auditory event-related potentials. In: *Psychophysiological Brain Research*, ed. by C.H.M. Brunia, A.W.K. Gaillard, A. Kok, G. Mulder, M.N. Verbaten (Tilburg Univ. Press, Tilburg 1990) pp. 294–298

22.22 M. Tervaniemi, M. Huotilainen, E. Brattico: Melodic multi-feature paradigm reveals auditory profiles in music-sound encoding, Front. Hum. Neurosci. **8**, 496 (2014)

22.23 H. Tiitinen, P. May, K. Reinikainen, R. Näätänen: Attentive novelty detection in humans is governed by pre-attentive sensory memory, Nature **372**, 90–92 (1994)

22.24 R. Näätänen, E. Schröger, S. Karakas, M. Tervaniemi, P. Paavilainen: Development of a memory trace for a complex sound in the human brain, NeuroReport **4**, 503–506 (1993)

22.25 E. Schröger: On the detection of auditory deviations: A pre-attentive activation model, Psychophysiology **34**(3), 245–257 (1997)

22.26 M. Tervaniemi, J. Saarinen, P. Paavilainen, N. Danilova, R. Näätänen: Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity, Biol. Psychol. **38**, 157–167 (1994)

22.27 H. Yabe, M. Tervaniemi, K. Reinikainen, R. Näätänen: Temporal window of integration revealed by MMN to sound omission, NeuroReport **8**, 1971–1974 (1997)

22.28 H. Yabe, M. Tervaniemi, J. Sinkkonen, M. Huotilainen, R.J. Ilmoniemi, R. Näätänen: The temporal window of integration of auditory information in the human brain, Psychophysiology **35**, 615–619 (1998)

22.29 E. Brattico, M. Tervaniemi, R. Näätänen: Context effects on pitch perception in musicians and nonmusicians: Evidence from event-related potential recordings, Music Perception **19**, 1–24 (2001)

22.30 I. Winkler, G. Karmos, R. Näätänen: Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event – related potential, Brain Res. **742**, 239–252 (1996)

22.31 K. Friston: A theory of cortical responses, Philos. Trans. R. Soc. B **360**, 815–836 (2005)

22.32 R. Näätänen, S. Pakarinen, T. Rinne, R. Takegata: The mismatch negativity (MMN): Towards the optimal paradigm, Clin. Neurophysiol. **115**(1), 140–144 (2004)

22.33 V. Putkinen, M. Tervaniemi, K. Saarikivi, N. De Vent, M. Huotilainen: Investigating the effects of musical training on functional brain development with a novel melodic MMN paradigm, Neurobiol. Learn. Mem. **110**, 8–15 (2014)

22.34 T. Särkämö, M. Tervaniemi, S. Laitinen, A. Numminen, M. Kurki, J.K. Johnson, P. Rantanen: Cognitive, emotional, and social benefits of regular musical activities in early dementia: Randomized controlled study, Gerontologist **54**(4), 634–650 (2014)

22.35 E. Partanen, R. Torppa, J. Pykäläinen, T. Kujala, M. Huotilainen: Children's brain responses to sound changes in pseudo words in a multifeature paradigm, Clin. Neurophysiol. **124**(6), 1132–1138 (2013)

22.36 R. Torppa, E. Salo, T. Makkonen, H. Loimo, J. Pykäläinen, J. Lipsanen, A. Faulkner, M. Huotilainen: Cortical processing of musical sounds in children with Cochlear Implants, Clin. Neurophysiol. **123**(10), 1966–1979 (2012)

22.37 B. Petersen, E. Weed, P. Sandmann, E. Brattico, M. Hansen, S. Derdau Sørensen, P. Vuust: Brain responses to musical feature changes in adolescent cochlear implant users, Front. Hum. Neurosci. **9**, 7 (2015), https://doi.org/10.3389/fnhum.2015.00007

22.38 M. Sams, R. Hari, J. Rif, J. Knuutila: The human auditory sensory memory trace persists about 10 msec: Neuromagnetic evidence, J. Cogn. Neurosci. **5**, 363–370 (1993)

22.39 C. Böttscher-Gandor, P. Ullsperger: Mismatch negativity in event-related potentials to auditory stimuli as a function of varying interstimulus interval, Psychophysiology **29**, 546–550 (1992)

22.40 N. Cowan: On short and long auditory stores, Psychol. Bull. **96**, 341–370 (1984)

22.41 I. Winkler, R. Näätänen: Event-related potentials in auditory backward recognition masking: A new way to study the neurophysiological basis of sensory memory in humans, Neurosci. Lett. **140**, 239–242 (1992)

22.42 R. Näätänen: Mismatch negativity (MMN) as an index of central auditory system plasticity, Int. J. Audiol. **47**(2), S16–S20 (2008)

22.43 E. Schröger, M. Tervaniemi, M. Huotilainen: Bottom-up and top-down flows of information within auditory memory: Electrophysiological evidence. In: *Psychophysics Beyond Sensation: Laws and Invariants of Human Cognition*, ed. by C. Kaernbach, E. Schröger, H. Müller (Erlbaum, Hillsdale 2004) pp. 389–407

22.44 P. Celsis, K. Boulanouar, B. Doyon, J.P. Ranjeva, I. Berry, J.L. Nespoulous, F. Chollet: Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and

tones, Neuroimage **9**, 135–144 (1999)
22.45 U. Schall, P. Johnston, J. Todd, P.B. Ward, P.T. Michie: Functional neuroanatomy of auditory mismatch processing: An event-related fMRI study of duration-deviant oddballs, Neuroimage **20**, 729–736 (2003)
22.46 S. Molholm, A. Martinez, W. Ritter, D.C. Javitt, J.J. Foxe: The neural circuitry of pre-attentive auditory change-detection: An fMRI study of pitch and duration mismatch negativity generators, Cereb. Cortex **15**, 545–551 (2005)
22.47 T. Rinne, A. Degerman, K. Alho: Superior temporal and inferior frontal cortices are activated by infrequent sound duration decrements: An fMRI study, Neuroimage **26**, 66–72 (2005)
22.48 A.K. Lee, E. Larson, R.K. Maddox, B.G. Shinn-Cunningham: Using neuroimaging to understand the cortical mechanisms of auditory selective attention, Hearing Res. **307**, 111–120 (2014)
22.49 C. Lappe, O. Steinsträter, C. Pantev: A beamformer analysis of MEG data reveals frontal generators of the musically elicited mismatch negativity, PLoS One **8**(4), e61296 (2013)
22.50 K. Alho, T. Rinne, T.J. Herron, D.L. Woods: Stimulus-dependent activations and attention-related modulations in the auditory cortex: a metaanalysis of fMRI studies, Hear. Res. **307**, 29–41 (2014)
22.51 T. Särkämö, E. Pihko, S. Laitinen, A. Forsblom, S. Soinila, M. Mikkonen, T. Autti, H.M. Silvennoinen, J. Erkkilä, M. Laine, I. Peretz, M. Hietanen, M. Tervaniemi: Music and speech listening enhance the recovery of early sensory processing after stroke, J. Cogn. Neurosci. **22**(12), 2716–2727 (2010)
22.52 M. Tervaniemi, E. Shröger, M. Saher, R. Näätänen: Effects of spectral complexity and sound duration in complex-sound pitch processing in humans-a mismatch negativity study, Neurosci. Lett. **290**, 66–70 (2000)
22.53 A. Dittmann-Balcar, M. Juptner, W. Jentzen, U. Schall: Dorsolateral prefrontal cortex activation during automatic auditory duration mismatch processing in humans: A positron emission tomography study, Neurosci. Lett. **308**, 119–122 (2001)
22.54 B.W. Müller, M. Juptner, W. Jentzen, S.P. Müller: Cortical activation to auditory mismatch elicited by frequency deviant and complex novel sounds: A pet study, NeuroImage **17**, 231–239 (2002)
22.55 C.F. Doeller, B. Opitz, A. Mecklinger, C. Krick, W. Reith, E. Schröger: Prefrontal cortex involvement in preattentive auditory deviance detection: neuroimaging and electrophysiological evidence, NeuroImage **20**, 1270–1282 (2004)
22.56 C.Y. Tse, T.B. Penney: On the functional role of temporal and frontal cortex activation in passive detection of auditory deviance, NeuroImage **41**, 1462–1470 (2008)
22.57 C.Y. Tse, T. Rinne, K.K. Ng, T.B. Penney: The functional role of the frontal cortex in pre-attentive auditory change detection, NeuroImage **19**(83C), 870–879 (2013)
22.58 K. Alho: Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes, Ear Hearing **16**, 38–51 (1995)
22.59 M.H. Giard, F. Perrin, J. Pernier, P. Bouchet: Brain generators implicated in processing of auditory stimulus deviance: A topographic event-related potential study, Psychophysiology **27**, 627–640 (1990)
22.60 R. Näätänen, P.T. Michie: Early selective attention effects on the evoked potential: A critical review and reinterpretation, Biol. Psychol. **8**, 81–136 (1979)
22.61 T. Rinne, R.J. Ilmoniemi, J. Sinkkonen, J. Virtanen, R. Näätänen: Separate time behaviors of the temporal and frontal MMN sources, Neuroimage **12**, 14–19 (2000)
22.62 R. Näätänen: The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function, The Behav. Brain Sci. **13**, 201–288 (1990)
22.63 K. Alho, C. Escera, R. Diaz, E. Yago, J.M. Serra: Effects of involuntary auditory attention on visual task performance and brain activity, NeuroReport **8**, 3233–3237 (1997)
22.64 M.H. Giard, J. Lavikainen, K. Reinikainen, F. Perrin, O. Bertrand, J. Pernier, R. Näätänen: Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: An event-related potential and dipole-model analysis, J. Cogn. Neurosci. **7**, 133–143 (1995)
22.65 S. Levänen, A. Ahonen, R. Hari, L. McEvoy, M. Sams: Deviant auditory stimuli activate human left and right auditory cortex differently, Cereb. Cortex **6**, 288–296 (1996)
22.66 S. Levänen, R. Hari, L. McEvoy, M. Sams: Responses of the human auditory cortex to changes in one versus two stimulus features, Exp. Brain Res. **97**, 177–183 (1993)
22.67 K. Alho, M. Tervaniemi, M. Huotilainen, J. Lavikainen, H. Tiitinen, R.J. Ilmoniemi, J. Knuutila, R. Näätänen: Processing of complex sounds in the human auditory cortex as revealed by magnetic brain responses, Psychophysiology **33**, 369–375 (1996)
22.68 M. Tervaniemi, A. Kujala, K. Alho, J. Virtanen, R.J. Ilmoniemi, R. Näätänen: Functional specialization of the human auditory cortex in processing phonetic and musical sounds: A magnetoencephalographic (MEG) study, NeuroImage **9**, 330–336 (1999)
22.69 M. Tervaniemi, A.J. Szameitat, S. Kruck, E. Schröger, K. Alter, W. De Baene, A.D. Friederici: From air oscillations to music and speech: Functional magnetic resonance imaging evidence for fine-tuned neural networks in audition, J. Neurosci. **26**(34), 8647–8652 (2006)
22.70 M. Tervaniemi, S. Maury, R. Näätänen: Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity, NeuroReport **5**, 844–846 (1994)

22.71 P. Paavilainen, M. Jaramillo, R. Näätänen: Binaural information can converge in abstract memory traces, Psychophysiology **35**, 483–487 (1998)

22.72 P. Paavilainen, J. Saarinen, M. Tervaniemi, R. Näätänen: Mismatch negativity to changes in abstract sound features during dichotic listening, Int. J. Psychophysiol. **9**, 243–249 (1995)

22.73 J. Saarinen, P. Paavilainen, E. Schröger, M. Tervaniemi, R. Näätänen: Representation of abstract stimulus attributes in human brain, NeuroReport **3**, 1149–1151 (1992)

22.74 O.A. Korzyukov, I. Winkler, V.I. Gumenyuk, K. Alho: Processing abstract auditory features in the human auditory cortex, NeuroImage **20**(4), 2245–2258 (2003)

22.75 P. Paavilainen, P. Arajärvi, R. Takegata: Preattentive detection of nonsalient contingencies between auditory features, NeuroReport **18**, 159–163 (2007)

22.76 P. Paavilainen, A. Degerman, R. Takegata, I. Winkler: Spectral and temporal stimulus characteristics in the processing of abstract auditory features, NeuroReport **14**(5), 715–718 (2003)

22.77 E. Schröger, A. Bendixen, N.J. Trujillo–Barreto, U. Roeber: Processing of abstract rule violations in audition, PLoS ONE **2**, e1131 (2007)

22.78 P. Paavilainen, J. Simola, M. Jaramillo, R. Näätänen, I. Winkler: Preattentive extraction of abstract feature conjunctions from auditory stimulation as reflected by the mismatch negativity (MMN), Psychophysiology **38**(2), 359–365 (2001)

22.79 D.J. Levitin, A.K. Tirovolas: Current advances in the cognitive neuroscience of music, Ann. N.Y. Acad. Sci. **1156**, 211–231 (2009)

22.80 E. Narmour: *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model* (Univ. of Chicago Press, Chicago 1990)

22.81 E. Brattico, M. Tervaniemi, R. Näätänen, I. Peretz: Musical scale properties are automatically processed in the human auditory cortex, Brain Res. **1117**(1), 162–174 (2006)

22.82 R. Näätänen, M. Tervaniemi, E. Sussman, P. Paavilainen, I. Winkler: Primitive intelligence in the auditory cortex, Trends Neurosci. **24**(5), 283–288 (2001)

22.83 E. Brattico, R. Näätänen, T. Verma, V. Välimäki, M. Tervaniemi: Processing of musical intervals in the central auditory system: An eventrelated potential (ERP) study on sensory consonance. In: *Proc. Sixth Int. Conf. Music Percept. Cognit., Keele*, ed. by C. Woods, G. Luck, R. Brochard, F. Seddon, J.A. Sloboda (Keele University, Department of Psychology, Keele 2000) pp. 1110–1119, CD-ROM

22.84 P. Virtala, V. Berg, M. Kivioja, J. Purhonen, M. Salmenkivi, P. Paavilainen, M. Tervaniemi: The preattentive processing of major vs. minor chords in the human brain: An event-related potential study, Neurosci. Lett. **487**(3), 406–410 (2011)

22.85 P. Virtala, V. Putkinen, M. Huotilainen, T. Makkonen, M. Tervaniemi: Musical training facilitates the neural discrimination of major vs. minor chords in 13-year-old children, Psychophysiology **49**, 1125–1132 (2012)

22.86 P. Virtala, M. Huotilainen, E. Partanen, V. Fellman, M. Tervaniemi: Newborn infants' auditory system is sensitive to Western music chord categories, Front. Psychol. **4**, 492 (2013)

22.87 P. Virtala, M. Huotilainen, E. Partanen, M. Tervaniemi: Musicianship facilitates the processing of Western music chords – An ERP and behavioural study, Neuropsychologia **61**, 247–258 (2014)

22.88 S. Koelsch: Toward a neural basis of music perception – A review and updated model, Front. Psychol. **2**, 110 (2013), https://doi.org/10.3389/fpsyg.2011.00110

22.89 S. Koelsch, E. Schröger, T.C. Gunter: Music matters: Preattentive musicality of the human brain, Psychophysiology **39**(1), 38–48 (2002)

22.90 P. Loui, T. Grent-'T-Jong, D. Torpey, M. Woldorff: Effects of attention on the neural processing of harmonic syntax in Western music, Cogn. Brain Res. **25**(3), 678–687 (2005)

22.91 S. Koelsch, B.H. Schmidt, J. Kansok: Effects of musical expertise on the early right anterior negativity: An event-related brain potential study, Psychophysiology **39**(5), 657–663 (2002)

22.92 E. Brattico, T. Tupala, E. Glerean, M. Tervaniemi: Modulated neural processing of Western harmony in folk musicians, Psychophysiology **50**(7), 653–663 (2013)

22.93 E.R. Kandel: The molecular biology of memory storage: A dialogue between genes and synapses, Science **294**(5544), 1030–1038 (2001)

22.94 M. Beauchemin, L. De Beaumont, P. Vannasing, A. Turcotte, C. Arcand, P. Belin, M. Lassonde: Electrophysiological markers of voice familiarity, Eur. J. Neurosci. **23**, 3081–3086 (2006)

22.95 O. Hauk, Y. Shtyrov, F. Pulvermuller: The sound of actions as reflected by mismatch negativity: Rapid activation of cortical sensory-motor networks by sounds associated with finger and tongue movements, Eur. J. Neurosci. **23**, 811–821 (2006)

22.96 S. Koelsch, E. Schröger, M. Tervaniemi: Superior attentive and pre-attentive auditory processing in musicians, NeuroReport **10**, 1309–1313 (1999)

22.97 M. Seppänen, E. Brattico, M. Tervaniemi: Practice strategies of musicians modulate neural processing and the learning of sound–patterns, Neurobiol. Learn. Mem. **87**(2), 236–247 (2007)

22.98 T. Fujioka, L.J. Trainor, B. Ross, R. Kakigi, C. Pantev: Musical training enhances automatic encoding of melodic contour and interval structure, J. Cogn. Neurosci. **16**(6), 1010–1021 (2004)

22.99 P. Vuust, E. Brattico, M. Seppänen, R. Näätänen, M. Tervaniemi: The sound of music: Differentiating musicians using a fast, musical multi-feature mismatch negativity paradigm, Neuropsychologia **50**(7), 1432–1443 (2012)

22.100 M. Tervaniemi, T. Ilvonen, K. Karma, K. Alho, R. Näätänen: The musical brain: Brain waves reveal the neurophysiological basis of musicality in human subjects, Neurosci. Lett. **226**, 1–4 (1997)

22.101 V. Putkinen, M. Tervaniemi, K. Saarikivi, P. Ojala, M. Huotilainen: Enhanced auditory change detection in musically trained school-aged children: A longitudinal event-related potential study, Development. Sci. **17**, 282–297 (2014)

22.102 E. Brattico, K.J. Pallesen, O. Varyagina, C. Bailey, I. Anourova, M. Järvenpää, T. Eerola, M. Tervaniemi: Neural discrimination of nonprototypical chords in music experts and laymen: An MEG study, J. Cogn. Neurosci. **21**(11), 2230–2244 (2009)

22.103 G. Musacchia, D. Strait, N. Kraus: Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians, Hearing Res. **241**(1-2), 34–42 (2008)

22.104 P.C. Wong, E. Skoe, N.M. Russo, T. Dees, N. Kraus: Musical experience shapes human brainstem encoding of linguistic pitch patterns, Nature Neurosci. **10**(4), 420–422 (2007)

22.105 G.M. Bidelman, M.W. Weiss, S. Moreno, C. Alain: Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians, Eur. J. Neurosci. **40**(4), 2662–2673 (2014)

22.106 J. Kühnis, S. Elmer, M. Meyer, L. Jäncke: The encoding of vowels and temporal speech cues in the auditory cortex of professional musicians: An EEG study, Neuropsychologia **51**(8), 1608–1618 (2013)

22.107 R. Milovanov, M. Huotilainen, V. Välimäki, P.A. Esquef, M. Tervaniemi: Musical aptitude and second language pronunciation skills in school-aged children: Neural and behavioral evidence, Brain Res. **1194**, 81–89 (2008)

22.108 J. Ayotte, I. Peretz, K. Hyde: Congenital amusia: A group study of adults afflicted with a music-specific disorder, Brain **125**(2), 238–251 (2002)

22.109 P. Albouy, J. Mattout, R. Bouet, E. Maby, G. Sanchez, P.E. Aguera, S. Daligault, C. Delpuech, O. Bertrand, A. Caclin, B. Tillmann: Impaired pitch perception and memory in congenital amusia: The deficit starts in the auditory cortex, Brain **136**(5), 1639–1661 (2013)

22.110 I. Peretz, E. Brattico, M. Tervaniemi: Abnormal electrical brain responses to pitch in congenital amusia, Ann. Neurol. **58**(3), 478–482 (2005)

22.111 P. Moreau, P. Jolicœur, I. Peretz: Pitch discrimination without awareness in congenital amusia: Evidence from event-related potentials, Brain Cogn. **81**(3), 337–344 (2013)

22.112 B. Tillmann, K. Schulze, J.M. Foxton: Congenital amusia: A short-term memory deficit for non-verbal, but not verbal sounds, Brain Cogn. **71**, 259–264 (2009)

22.113 I. Peretz, E. Brattico, M. Järvenpää, M. Tervaniemi: The amusic brain: In tune, out of key, and unaware, Brain **132**(5), 1277–1286 (2009)

22.114 G. Mignault Goulet, P. Moreau, N. Robitaille, I. Peretz: Congenital amusia persists in the developing brain after daily music listening, PLoS One **7**(5), e36860 (2012)

22.115 S. Koelsch, M. Wittfoth, A. Wolf, J. Muller, A. Hahne: Music perception in cochlear implant users: An event-related potential study, Clin. Neurophysiol. **115**, 966–972 (2004)

22.116 C.J. Limb, J.T. Rubinstein: Current research on music perception in cochlear implant users, Otolaryngol. Clin. N. Am. **45**, 129–140 (2012)

22.117 L. Timm, P. Vuust, E. Brattico, D. Agrawal, S. Debener, A. Büchner, R. Dengler, M. Wittfoth: Residual neural processing of musical sound features in adult cochlear implant users, Front. Hum. Neurosci. **8**, 181 (2014)

22.118 C.G. Jung: On the importance of the unconscious in psychopathology, Br. Med. J. **2**, 964–968 (1914)

22.119 B. Gold, M.J. Frank, B. Bogert, E. Brattico: Pleasurable music affects reinforcement learning according to the listener, Front. Psychol. **4**, 541 (2013), https://doi.org/10.3389/fpsyg.2013.00541

22.120 T. Quarto, G. Blasi, K.J. Pallesen, A. Bertolino, E. Brattico: Implicit processing of visual emotions is affected by sound-induced affective states and individual affective traits, PLoS One **9**(7), e103278 (2014), https://doi.org/10.1371/journal.pone.0103278

23. Long-Term Memory for Music

Lola L. Cuddy

This chapter begins with a brief overview of traditional approaches to long-term memory in general. It highlights the memory system known as semantic memory, and then in subsequent sections explores the proposal that a purely musical semantic memory may be identified and is isolable from semantic memory in other, nonmusical, domains. Finally, neuropsychological evidence for the selective sparing of the musical semantic memory system in dementia is reviewed.

23.1 Long-Term Memory and the Semantic System

Research in long-term memory has led to a theoretical description of memory organization according to three interdependent systems – semantic, episodic, and procedural memory systems [23.1]. The concept of *semantic memory* was famously introduced to psychology by *Tulving* [23.2] as follows:

> *A mental thesaurus, organized knowledge a person possesses about words and other verbal symbols, their meaning and referents, about relations among them, and about rules, formulas and algorithms for the manipulation of these symbols, concepts, and relations.*

It was to be distinguished from the type of memory typically studied in laboratories of the time, namely, *episodic memory* [23.2, 3]. Episodic memories are memories for specific events or life-time situations, tagged with reference to specific temporal-spatial locations. Recalling that California is a state of the U.S. requires semantic memory, whereas recalling whether you have ever been in California requires episodic memory. Recalling a list of unrelated words in a memory experiment involves episodic memory – that is, the meanings of the words are already known and represented in semantic memory, but the participant in the experiment must remember that a given word was presented at a certain place and time in a word list spoken by a specific individual. Autobiographical memories – self-referential life-time memories – are a type of episodic memory that may contain semantic features. Semantic and episodic memories have been studied traditionally in the verbal and visual domains.

The third system, *procedural memory*, is skill memory – the performance in multiple domains of particular types of action such as riding a bike, reading, tying one's shoes, or, perhaps, singing a tune. It is thought that procedural memory is below the level of conscious awareness, because it is easy to produce these actions, though difficult to explain how and where they were learned. Thus, procedural memories do not require conscious control or attention. They are acquired through repetition of a complex activity and so become automatic.

Another distinction arising from theories of memory organization is that of explicit and implicit memory. Whereas *explicit memory* involves conscious recollections of events, *implicit memory* involves the effect of prior learning on performance without conscious awareness of effect. Procedural memory, with its characteristic definition of unconscious processing, has been identified with implicit memory. By contrast, semantic and episodic memories have been identified with explicit, conscious, memory. However, implicit memory is not always related to procedural memory. Both semantic and episodic memories have implicit features, a point which has implications for describing the contents of musical semantic memory, as discussed below.

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_23

23.2 Semantic Memory for Music

This section deals with the issue whether there is a musical semantic memory that is behaviorally, and possibly neurally, distinct from verbal semantic memory. It addresses the questions how such a memory might be described and defined. It is a matter of debate whether musical semantic memory is contained within a general verbal semantic system or is itself a separate entity.

Slevc and *Patel* [23.4] argue that it is not possible to disclose the semantic meaning of music specifically. According to *Patel* [23.5], the range of musical semantics is rather limited, because music, unlike words, does not have semantic referents that link to meaningful concepts. Such a perspective implies that musical semantic memory may be loosely tied to a general semantic memory, but lacks the resources to organize a vast semantic network of its own.

It is true that musical knowledge may be tied to linguistic elements – such as song lyrics, titles, music-theoretic, and cultural concepts. *Steinbeis* and *Koelsch* [23.6] have demonstrated that the emotions expressed by single musical features activate the processing of related emotional words. It is also plausible, to borrow an example from *Patel* [23.5], to conclude that it takes little effort to decide whether Beethoven's *Appassionata* connotes *explosive fury* or *passive contemplation*. However, word-associations to music are extramusical and failure to retrieve these linguistic descriptors need not signal failure of the musical semantic memory system.

Apart from extramusical referents, however, musical meaning *can emerge from the interpretation of intramusical structural relations* contained in the music itself [23.7, pp. 170–171]. For *Omar* et al. [23.8, p. 468]), semantic musical memory is a *self-contained, abstract, purely musical* [...] nonreferential system *meaningful on its own terms* containing knowledge of familiar tunes, instrumental timbres, symbols (notation), rules of musical structure, and emotion. For *Platel* et al. [23.9, p. 245]:

> *Semantic memory allows us to identify or to have a strong feeling of knowing for familiar songs or melodies* [...] *Musical semantic memory may represent a musical lexicon, separate of a verbal lexicon, even though strong links certainly exist between them.*

The musical lexicon has been formally described as a component of a modular music recognition system [23.10] (see also [23.11, 12]).

> *The musical lexicon is presently best understood as a perceptual representational system for isolated tunes, much in the same way as the mental word lexicon represents isolated words.* [23.13, p. 257]

This perspective on musical semantic memory emphasizes the point that the musical lexicon enables a strong sense of familiarity for a previously encountered tune interacting with, but relatively independently of, verbal associations with the tune.

In accord with this notion, neuroimaging studies have proposed distinct neural networks for musical and verbal semantic memory. For example, *Groussard* et al. [23.14] reported an fMRI study comparing activation for verbal and musical semantic representations. Within the left temporal cortex, musical material mainly activated the superior temporal gyrus and verbal material mainly activated the middle and inferior gyri.

The contents of musical semantic memory are acquired and processed implicitly – that is without conscious effort and awareness. In an extensive and important critical review, *Rohrmeier* and *Rebuschat* [23.15] discuss evidence that implicit musical knowledge is formed through mere exposure to musical structures without explicit instruction about the rules for building such structures. See for example, the studies on implicit knowledge of musical regularities and the results of artificial grammar learning studies (reviewed in [23.15, 16]). Knowledge of the principles governing tonality is implicit knowledge. This knowledge, implicitly acquired, may be tapped by semantic and episodic memory.

Formal music training is not required to recognize familiar tunes, generate musical expectancies and respond to the emotional properties – arousal and valence – of music [23.17]. *Krumhansl* [23.18] proposed that statistically frequent patterns in music are guides to understanding the organizing principles of hierarchical tonal structure in music. *Krumhansl* and *Cuddy* [23.19] note that these principles are implicitly picked up by preschool-age children, though they cannot describe this knowledge. The process of acquisition may be similar to that of natural language acquisition [23.20].

Moreover, when presented with musical excerpts that do not correspond to Western musical tonality, adult listeners implicitly re-orient to the unfamiliar pitch hierarchies of cross-cultural musical systems. Their pitch judgments correspond to the statistical distributions of tones (for the music of North India, see [23.21], for the music of Bali [23.22], for Korean court music [23.23]).

23.3 Evidence from Neuropsychology

There are many underlying causes of amnesia and memory loss – for example, stroke, tumor, psychiatric depression, vitamin deficiency, and neurodegenerative disease. Patients with brain damage following stroke have provided evidence for an isolated musical semantic memory or musical lexicon that is selectively lost or alternatively spared. *Peretz* and *Coltheart* [23.10] summarize case reports in which brain-damaged nonmusicians have lost the ability to recognize tunes while maintaining the ability to recognize words, including the lyrics that accompany the tunes. Similarly, the patient KB, who became amusic and aprosodic following right-hemisphere stroke, was unable to recognize instrumental tunes despite preserved cognitive abilities [23.24] (see also [23.25]). KB was able to recognize lyrics of popular tunes, when spoken, but unable to produce a single response to the request to sing the tune that would normally accompany the spoken lyrics. The reverse situation, loss of ability to recognize words but preserved ability to recognize tunes, is found more rarely, but was reported for three brain-damaged patients who were afflicted with word deafness yet able to recognize music [23.10]. However, selective preservation of musical memory is more commonly and strikingly found among the dementias.

Evidence from dementias studies: Alzheimer Disease (AD): Of the cases of neurodegenerative disease leading to dementia, Alzheimer's disease (AD) is the most prevalent accounting for about 2/3 of cases of dementia. The course of AD is relentlessly progressive through mild, moderate, and severe stages. There is no known cure. It primarily affects cholinergic transmissions to all cortical areas; some may be affected at differential rates but eventually all become impaired. The diagnosis is confirmed only by postmortem findings of characteristic plaques and tangles in brain tissue.

Participants in early studies of sparing in AD were representative of a selected population – professional or skilled musicians (pianists, violinists, a trombonist; for a detailed summary table of references, see [23.8, Table 1]; also see [23.26, Table 1]). Performance sparing for these individuals might be seen as evidence of preserved procedural, not semantic, musical memory.

Our first study was a case study [23.27] of an 84-year-old woman (EN), a former music lover, but not a professional musician, EN was severely demented; she was unable to carry on an intelligible conversation and did not recognize family members. Yet, as the family members reported to us, she still enjoyed and remembered music. We did not attempt imaging because of the difficulty of explaining the procedure to EN and our wish to avoid creating obvious distress. However, a postmortem examination two years following our testing sessions revealed neuropathological changes that confirmed relatively advanced AD – moderately severe atrophy of medial temporal lobe (MTL) structures and expansion of temporal horns of lateral ventricles. Immunostaining of representative sections showed significant neuronal loss and reactive gliosis in MTL together with the presence of moderate to numerous diffuse and neuritic amyloid plaques and prominent neurofibrillary degeneration (Braak Stage V of VI: [23.28]).

We had administered three tests in which we assessed EN's familiarity with melodies and one in which we had assessed her ability to detect a pitch distortion in familiar melodies. The test employed was the Distorted Tunes Test (DTT) constructed by *Dennis Drayna* [23.29] as an update of a test intended to identify tune-deafness [23.30]. Twenty-six short melodies were presented of which 17 were distorted by pitch errors. Pitch errors changed two to nine notes generally within one or two semitones of the correct note. The distortion maintained the contour (ups and downs in pitch) and rhythm of the familiar melody but the distortions violated the tonal structure of the melody. An online version of the DTT is available where the melodies may be heard [23.31].

Familiarity was assumed if EN sang along correctly with a short test tune and continued to sing after the tune stopped. *Unfamiliarity* was assumed if she sat with a blank stare and made no effort to sing along with the test tune. Evidence of correct detection of a pitch distortion was grimacing, frowning, and other facial signs of displeasure. No such displeasure was taken to indicate that she found the tune to be correctly played.

EN's test results provided encouraging support for the possibility that sparing of musical memory may be detected in dementia and may be reliably and quantitatively assessed through behavioral observation. Allowing assessment by behavioral observation corroborated with an independent judge, we found that EN's scores were high and typically in the control range. We noted dramatic contrast between EN's response to music and her mental status. We also noted that evidence of sparing would not have been recovered with conventional assessment – conventional assessment requiring verbal communication would suggest severe musical difficulties.

Subsequent work [23.32] followed the case study with an assessment of 50 AD persons at various stages of the disease. Their performance on six tests was compared with a larger sample of healthy young and older adults and showed intact recognition of familiar melodies along with preserved ability to detect pitch

distortions in the mild and moderate stages. Difficulties with detecting grammatical distortions in spoken song lyrics were evident even at the mild stage of AD.

Among some AD persons there is also sparing of ability to detect tonal violations even with unfamiliar melodies [23.33]. Where this is found, it may be said to reveal a preservation of implicit tonal knowledge. It implies that the detection of a distortion is not merely a process of matching a heard melody with a stored template of a melody and detecting a mismatch. It implies that there is an ability to detect a deviation from the tonal system of pitches, an ability that reflects musical knowledge of the Western tonal-hierarchical system of pitch organization.

A recent study by *Kerer* et al. [23.26] assessed 10 participants with MCI and 10 with early stage AD and compared their results with 28 cognitively normal older adults. They tested identification of familiar musical melodies and recognition of a pitch distortion in the melodies. Pitch distortions were introduced by rules similar to those followed in the DTT described above. Compared to healthy controls, their patient sample was poorer on verbal tasks such as naming the titles of the melodies, thus exhibiting a failure of the word-association system. However, patients were superior to controls on detecting a pitch distortion in familiar melodies and also in a timbre task, which was to decide whether melodies were normally played by instruments or presented vocally. The authors attribute the latter findings to a specialized memory for music.

In tandem with the above studies from our laboratory, we have also reported that recognition memory is compromised, not preserved, both in healthy aging, and dementia, as well as in mild cognitive impairment (suspected prodromal to AD) [23.34, 35]. The paradigm for assessing recognition memory involves presenting a playlist of melodies to a participant followed by a test list in which the playlist is interleaved with new melodies. The participant must judge which melodies in the test list occurred in the playlist and which did not. Young adults find the test quite easy, but older adults, even cognitively healthy older adults, do not. The problem for older adults and AD adults is that it is difficult to distinguish melodies heard on the playlist (where the participant must remember the episode of hearing the melodies in a certain place at a certain time) from melodies heard before outside the laboratory. The kind of memory being tested in a recognition memory paradigm may be characterized as episodic memory, and its failure in AD may be partly attributable to aging, rather than the disease.

Other studies with the recognition paradigm have obtained similar results [23.36, 37]. Halpern and O'Connor used an explicit and an implicit memory task. Whether test instructions were explicit (yes/no responses to a previously heard playlist of melodies mingled with new melodies) or implicit (pleasantness) ratings for the melodies, AD participants show poorer performance than healthy controls. An exception to this finding is a study by *Quoniam* et al. [23.38] who reported a positive affective bias (liking) for previously heard melodies in AD. Their results are puzzling, however, in that healthy controls showed little effect of repetition on likingness with the AD group surpassing the controls; the finding deserves to be replicated.

Evidence from other neuropsychological sources: A number of researchers have studied a somewhat rarer form of dementia, known as fronto-temporal dementia and its temporal variant semantic dementia. Frontotemporal dementia involves damage to the brain's frontal and temporal lobes and affects planning and judgment, emotions, speaking and understanding speech, walking and other basic movements. Both frontotemporal degeneration (FTD) and Alzheimer's disease (AD) are characterized by atrophy of the brain, and a gradual, progressive loss of brain function. However, FTD is primarily a disease of behavior and language dysfunction, while the hallmark of Alzheimer's disease is loss of memory, especially episodic memory. FTD patients exhibit behavioral and personality changes but retain features of memory such as keeping track of day-to-day events, and orientation to space and time. AD patients display increasing memory deficits, but typically retain socially appropriate behavior in the early (mild to moderate) stages. Some FTD patients may have language dysfunction only as revealed in the two types of progressive aphasia: semantic dementia and progressive nonfluent aphasia. The language decline seen in AD patients is of a milder nature [23.39].

Semantic dementia (SD) is associated with predominantly temporal lobe atrophy (left greater than right). The hallmark of the aphasia in semantic dementia is difficulty generating or recognizing familiar words. Patients exhibit language difficulties, not in producing speech but with a loss of the meaning or semantics of words, faces, and objects. Although semantic memory impairment occurs in AD, AD is almost always accompanied by severe impairment of episodic memory, and eventually disruption of many other cognitive domains, including visuospatial and frontal executive function [23.40, 41].

A few studies have assessed long-term musical memory in FTD and SD; some have drawn comparisons with AD cases [23.42–45]. In general, SD participants showed deficits in familiarity decisions, especially where naming of tunes is involved. *Hailstone* et al. [23.46] present a case study of SD with convincing evidence that musical knowledge was relatively

preserved. In the study by *Hsieh* et al. [23.42] variability in SD results was high and three of the 13 SD patients performed in the normal range on their tune familiarity test. AD patients, as might be expected from the AD findings cited earlier above, did not differ significantly from controls.

Weinstein et al. [23.47] report a case study of profound impairment in SD – a computer programmer in his early 60s who had also been a semiprofessional harpsichordist. The man was able to perform both familiar and previously unfamiliar Baroque pieces with accuracy and musicality. Of particular importance was the presence in his performances of novel, additional, notes that were style-appropriate embellishments. He also introduced a stylistically appropriate retard in the final, compared with the initial, presentation of a theme. The authors suggest that their case study may reflect the preservation of the abstract concept of musical style in SD; that is, preservation was not limited to the habitual striking of keys and was not solely procedural. They propose the hypothesis that:

> *knowledge represented in semantic memory is partly modular, dissociating long-term representations of object and word concepts from meaningful knowledge in other domains, such as music.* [23.47, p.250]

A case of musical sparing despite profound semantic and episodic impairment following herpes encephalitis was presented by *Finke* et al. [23.48]. Their patient was a a 69-year-old professional cellist (PM) who, despite his severe memory impairment, performed normally on a battery of musical tests, including discriminability of famous from nonfamous pieces, and learning complex new musical material as revealed by an incidental memory task. *Finke* et al. [23.48, p. 592] note that for PM: *learning and memory of complex musical information constitutes an island of intact cognition within a severe amnesic syndrome* and suggest this information involves a brain network independent of other types of (nonmusical) episodic and semantic memory.

23.4 Concluding Comments

Neuropsychological studies of neuropathology have yielded valuable behavioral/cognitive evidence supporting the notion of a musical semantic system that may be dissociated from semantic memories in other domains. This musical system may be preferentially spared in the face of loss of cognitive abilities. What my colleagues and I [23.27] once felt was a puzzle to decipher – whether there was loss or sparing of the musical system in AD – is resolved by considering that loss was found where musical episodic memories were tested, preservation when musical semantic memory was tested. Here it is doubtless relevant that *Platel* et al. [23.9] have uncovered separate neural substrates for musical episodic and musical semantic memories.

This chapter therefore proposes that a *musical semantic memory* can be dissociated from the *traditional*, or nonmusical, *semantic memory*. From patient studies, results indicate that musical semantic memory may be further characterized in three ways. First, musical semantic memory contains or accesses implicit tonal knowledge allowing the detection of tonal irregularities or violations. Second, memories may be acquired independently of performance. The patient studied by *Finke* et al. [23.48], reported above, was capable of learning music composed, but not practiced, after the onset of his encephalitis. Third, musical semantic memories may activate other personal, autobiographical memories in both AD [23.49–51] and brain-damaged patients [23.52]. The music may never have been performed by the patient but was experienced under memorable circumstances. Hence an associative link was established between a musical semantic memory and a life-time episode.

For future research what is needed is a standardized battery of musical tests that includes both musical episodic, musical semantic, and musical procedural tasks. Comparison of research results, including neuroimaging, would then be facilitated. More attention should be paid to the implicit features of musical knowledge, as outlined by *Rohrmeier* and *Rebuschat* [23.15].

Behavioral findings of spared long-term musical memory in neurodegenerative disease have important implications for palliative care while the cure remains elusive. Care givers can rely on the emotional associations with familiar music to promote a reduction of patient anxiety and agitation.

Acknowledgments. Lola L. Cuddy's research was supported by a Discovery Grant from NSERC (RGPIN/333), the GRAMMY foundation, and the Alzheimer Society of Canada. Ritu Sikka provided invaluable assistance in the preparation of this chapter.

References

23.1 E. Tulving: How many memory systems are there?, Am. Psychol. **40**(4), 385–398 (1985)

23.2 E. Tulving: Episodic and semantic memory. In: *Organization of Memory*, ed. by E. Tulving, W. Donaldson (Academic, London 1972) pp. 381–404

23.3 E. Tulving: What is episodic memory?, Curr. Dir. Psychol. Sci. **2**(3), 67–70 (1993)

23.4 L.R. Slevc, A.D. Patel: Meaning in music and language: Three key differences: Comment on "Towards a neural basis of processing musical semantics" by Stefan Koelsch, Phys. Life Rev. **8**(2), 110–111 (2011), https://doi.org/10.1016/j.plrev.2011.05.003

23.5 A.D. Patel: Sharing and nonsharing of brain resources for language and music. In: *Language, Music, and the Brain: A Mysterious Relationship*, ed. by M.A. Arbib (MIT Press, Cambridge 2013) pp. 329–355

23.6 N. Steinbeis, S. Koelsch: Affective priming effects of musical sounds on the processing of word meaning, J. Cogn. Neurosci. **23**(3), 604–621 (2011), https://doi.org/10.1162/jocn.2009.21383

23.7 S. Koelsch: *Brain and Music* (Wiley, Hoboken 2012) p. 308

23.8 R. Omar, J.C. Hailstone, J.D. Warren: Semantic memory for music in dementia, Music Percept. **29**(5), 467–477 (2012), https://doi.org/10.1525/MP.2012.29.5.467

23.9 H. Platel, J.-C. Baron, B. Desgranges, F.A. Bernard, F. Eustache: Semantic and episodic memory of music are subserved by distinct neural networks, NeuroImage **20**(1), 244–256 (2003), https://doi.org/10.1016/S1053-8119(03)00287-8

23.10 I. Peretz, M. Coltheart: Modularity of music processing, Nat. Neurosci. **6**(7), 688–691 (2003), https://doi.org/10.1038/nn1083

23.11 I. Peretz, A.S. Champod, K.L. Hyde: Varieties of musical disorders, Ann. New York Acad. Sci. **999**, 58–75 (2003)

23.12 I. Peretz, R.J. Zatorre: Brain organization for music processing, Annu. Rev. Psychol. **56**, 89–114 (2005), https://doi.org/10.1146/annurev.psych.56.091103.070225

23.13 I. Peretz, N. Gosselin, P. Belin, R.J. Zatorre, J. Plailly, B. Tillmann: Music lexical networks: The cortical organization of music recognition, Ann. New York Acad. Sci. **1169**, 256–265 (2009), https://doi.org/10.1111/j.1749-6632.2009.04557.x

23.14 M. Groussard, G. Rauchs, B. Landeau, F. Viader, B. Desgranges, F. Eustache, H. Platel: The neural substrates of musical memory revealed by fMRI and two semantic tasks, NeuroImage **53**(4), 1301–1309 (2010), https://doi.org/10.1016/j.neuroimage.2010.07.013

23.15 M. Rohrmeier, P. Rebuschat: Implicit learning and acquisition of music, Top. Cogn. Sci. **4**(4), 525–553 (2012), https://doi.org/10.1111/j.1756-8765.2012.01223.x

23.16 M. Rohrmeier, Z. Dienes, X. Guo, Q. Fu: Implicit learning and recursion. In: *Language Recursion*, ed. by F. Lowenthal, L. Lefebvre (Springer, New York 2014) pp. 67–85

23.17 E. Bigand, B. Poulin-Charronnat: Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training, Cognition **100**(1), 100–130 (2006), https://doi.org/10.1016/j.cognition.2005.11.007

23.18 C.L. Krumhansl: *Cognitive Foundations of Musical Pitch*, Vol. 17 (Oxford Univ. Press, New York 1990)

23.19 C.L. Krumhansl, L.L. Cuddy: A theory of tonal hierarchies in music. In: *Music Perception: Springer Handbook of Auditory Research*, Vol. 36, ed. by M. Riess Jones, R.R. Fay, A.N. Popper (Springer, New York 2010) pp. 51–87

23.20 C.L. Krumhansl, F.C. Keil: Acquisition of the hierarchy of tonal functions in music, Mem. Cogn. **10**, 243–251 (1982)

23.21 M.A. Castellano, J.J. Bharucha, C.L. Krumhansl: Tonal hierarchies in the music of North India, J. Exp. Psychol. Gen. **113**(3), 394–412 (1984)

23.22 E.J. Kessler, C. Hansen, R.N. Shepard: Tonal schemata in the perception of music in Bali and in the West, Music Percept. **2**, 131–165 (1984)

23.23 M.E. Lantz, J.-K. Kim, L.L. Cuddy: Perception of a tonal hierarchy derived from Korean music, Psychol. Music **42**(4), 580–598 (2013), https://doi.org/10.1177/0305735613483847

23.24 W.R. Steinke, L.L. Cuddy, L.S. Jakobson: Dissociations among functional subsystems governing melody recognition after right-hemisphere damage, Cogn. Neuropsychol. **18**(5), 411–437 (2001)

23.25 A.D. Vanstone, L.L. Cuddy, J.M. Duffin, E. Alexander: Exceptional preservation of memory for tunes and lyrics: Case studies of amusia, profound deafness, and Alzheimer's disease, Ann. N.Y. Acad. Sci. **1169**, 291–294 (2009), https://doi.org/10.1111/j.1749-6632.2009.04763.x

23.26 M. Kerer, J. Marksteiner, H. Hinterhuber, G. Mazzola, G. Kemmler, H.R. Bliem, E.M. Weiss: Explicit (semantic) memory for music in patients with mild cognitive impairment and early-stage Alzheimer's disease, Exp. Aging Res. **39**(5), 536–564 (2013), https://doi.org/10.1080/0361073X.2013.839298

23.27 L.L. Cuddy, J.M. Duffin: Music, memory, and Alzheimer's disease: Is music recognition spared in dementia, and how can it be assessed?, Med. Hypotheses **64**, 229–235 (2005), https://doi.org/10.1016/j.mehy.2004.09.005

23.28 H. Braak, E. Braak: Neuropathological stageing of Alzheimer-related changes, Acta Neuropathol. **82**, 239–259 (1991), https://doi.org/10.1007/BF00308809

23.29 D. Drayna, A. Manichaikul, M. de Lange, H. Snieder, T. Spector: Genetic correlates of musical pitch recognition in humans, Science **291**, 1969–1972 (2001)

23.30 H. Kalmus, D. Fry: On tune deafness (dysmelodia): Frequency, development, genetics and musical background, Ann. Hum. Genet. **43**, 369–382 (1980)

23.31 National Institute on Deafness and Other Communication Disorders: Distorted Tunes Test, https://www.nidcd.nih.gov/tunestest/take-distorted-tunes-test (2014) reproduced courtesy of D.T. Drayna

23.32 L.L. Cuddy, J.M. Duffin, S.S. Gill, C.L. Brown, R. Sikka, A.D. Vanstone: Memory for melodies and lyrics in Alzheimer's Disease, Music Percept. **29**(5), 479–491 (2012), https://doi.org/10.1525/MP.2012.29.5.479

23.33 A.D. Vanstone, L.L. Cuddy: Musical memory in Alzheimer disease, Aging, Neuropsychol. Cogn. **17**(1), 108–128 (2010), https://doi.org/10.1080/13825580903042676

23.34 M. Collett, R. Sham, A.D. Vanstone, A. Garcia, L. L. Cuddy: Episodic memory for melodies in mild cognitive impairment. Presented to the Rotman Research Institute, Toronto (2012)

23.35 A.D. Vanstone, R. Sikka, L. Tangness, R. Sham, A. Garcia, L.L. Cuddy: Episodic and semantic memory for melodies in Alzheimer's disease, Music Percept. **29**(5), 501–507 (2012), https://doi.org/10.1525/MP.2012.29.5.501

23.36 A.R. Halpern, J.C. Bartlett, W.J. Dowling: Aging and experience in the recognition of musical transpositions, Psychol. Aging **10**(3), 325–342 (1995)

23.37 A.R. Halpern, M.G. O'Connor: Implicit memory for music in Alzheimer's disease, Neuropsychology **14**(3), 391–397 (2000), https://doi.org/10.1037/0894-4105.14.3.391

23.38 N. Quoniam, A.-M. Ergis, P. Fossati, I. Peretz, S. Samson, M. Sarazin, J.-F. Allilaire: Implicit and explicit emotional memory for melodies in Alzheimer's disease and depression, Ann. N.Y. Acad. Sci. **999**, 381–384 (2003), https://doi.org/10.1196/annals.1284.047

23.39 The Association for Frontotemporal Degeneration (AFTD): https://www.theaftd.org/understandingftd/diagnosis (2007–2017)

23.40 J.R. Hodges: Memory in the dementias. In: *The Oxford Handbook of Memory*, Vol. 3 (Oxford University Press, New York 2000) pp. 441–459

23.41 J.R. Hodges, D.P. Salmon, N. Butters: Semantic memory impairment in Alzheimer's disease: Failure of access or degraded knowledge?, Neuropsychologia **30**(4), 301–314 (1992), https://doi.org/10.1016/0028-3932(92)90104-T

23.42 S. Hsieh, M. Hornberger, O. Piguet, J.R. Hodges: Neural basis of music knowledge: Evidence from the dementias, Brain **134**(9), 2523–2534 (2011), https://doi.org/10.1093/brain/awr190

23.43 J.K. Johnson, C.-C. Chang, S.M. Brambati, R. Migliaccio, M.L. Gorno-Tempini, B.L. Miller, P. Janata: Music recognition in frontotemporal lobar degeneration and Alzheimer disease, Cogn. Behav. Neurol. **24**(2), 74–84 (2011), https://doi.org/10.1097/WNN.0b013e31821de326

23.44 R. Omar, J.C. Hailstone, J.E. Warren, S.J. Crutch, J.D. Warren: The cognitive organization of music knowledge: A clinical analysis, Brain **133**(4), 1200–1213 (2010), https://doi.org/10.1093/brain/awp345

23.45 R. Omar, S.M.D. Henley, J.W. Bartlett, J.C. Hailstone, E. Gordon, D.A. Sauter, C. Frost, S.K. Scott, J.D. Warren: The structural neuroanatomy of music emotion recognition: Evidence from frontotemporal lobar degeneration, NeuroImage **56**(3), 1814–1821 (2011), https://doi.org/10.1016/j.neuroimage.2011.03.002

23.46 J.C. Hailstone, R. Omar, J.D. Warren: Relatively preserved knowledge of music in semantic dementia, J. Neurol. Neurosurg. Psychiatry **80**(7), 808–809 (2009), https://doi.org/10.1136/jnnp.2008.153130

23.47 J. Weinstein, P. Koenig, D. Gunawardena, C. McMillan, M. LBonner, M. Grossman: Preserved musical semantic memory in semantic dementia, Arch. Neurol. **68**(2), 248–250 (2011), https://doi.org/10.1001/archneurol.2010.364

23.48 C. Finke, N.E. Esfahani, C.J. Ploner: Preservation of musical memory in an amnesic professional cellist, Curr. Biol. **22**(15), R591–R592 (2012), https://doi.org/10.1016/j.cub.2012.05.041

23.49 L.L. Cuddy, R. Sikka, A.D. Vanstone: Preservation of musical memory and engagement in healthy aging and Alzheimer's disease, Ann. N.Y. Acad. Sci. **1337**, 223–231 (2015)

23.50 M. El Haj, L. Fasotti, P. Allain: The involuntary nature of music-evoked autobiographical memories in Alzheimer's disease, Conscious. Cogn. **21**, 238–246 (2012)

23.51 M. El Haj, V. Postal, P. Allain: Music enhances autobiographical memory in mild Alzheimer's disease, Educ. Gerontol. **38**, 30–41 (2012)

23.52 A. Baird, S. Samson: Music evoked autobiographical memory after severe acquired brain injury: Preliminary findings from a case series, Neuropsychol. Rehabil. **24**(1), 125–143 (2014), https://doi.org/10.1080/09602011.2013.858642

24. Auditory Working Memory

Katrin Schulze, Stefan Koelsch, Victoria Williamson

This chapter reviews behavioral and neuroimaging findings on:

1. The comparison between verbal and tonal working memory (WM)
2. The impact of musical training
3. The role of sound mimicry for auditory memory
4. The influence of long-term memory (LTM) on auditory WM performance, i. e., the effect of strategy use on auditory WM.

Whereas the core structures, namely Broca's area, the premotor cortex, and the inferior parietal lobule, show a substantial overlap, results in musicians suggest that there are also different subcomponents involved during verbal and tonal WM. If confirmed, these results indicate that musicians develop either independent tonal and phonological loops or unique processing strategies that allow novel interactive use of the WM systems. We furthermore present and discuss data that provide substantial support for the hypothesis that motor-related processes assist auditory WM, and as a result we propose a strong link between sound mimicry and auditory WM.

24.1 The Baddeley and Hitch WM Model: Theoretical Considerations and Empirical Support

Working memory (WM) describes a brain system that is thought to be responsible for the temporary storage and simultaneous manipulation of information [24.1–4]. This system underpins higher cognitive functions like planning, problem solving, comprehension and reasoning, and is critical for understanding or appreciating speech and music.

There are many theories and models of short-term memory (STM) or WM (for an overview see [24.5–11]). Our chapter, however, is theoretically embedded in the highly influential *Baddeley* and *Hitch* [24.3] WM model due to the fact that studies that have examined WM for music and language have overwhelmingly focused on this framework [24.3, 12–18]. In addition, although parts of this model are still debated, no other verbal memory model is as well investigated, developed, or corroborated by research results [24.4].

There is little consensus in the literature regarding use of the terms STM and WM [24.19]. In some publications the term STM is used to describe the temporary storage of information, while WM refers to the maintenance and manipulation of information [24.6, 19]. Often, however, it is difficult to know whether a task needs further processing and/or manipulation in addition to the passive storage of information, e.g., if one tone has to be compared with several tones played in a sequence. Therefore, in this chapter no distinction is made between STM and WM; we refer only to WM. In

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_24

addition, the term *auditory WM* will be used to describe WM processes for stimuli that were presented auditorily. Finally, it should be noted that whereas verbal WM has been investigated using both recall and recognition tasks, WM for tonal material has primarily been studied within recognition paradigms (but see [24.18] for a recall task for musical stimuli).

As mentioned, the present review is based on the *Baddeley* and *Hitch* WM model, the first version of which was described almost 40 years ago [24.3]. This initial WM model was developed from earlier frameworks such as the *Atkinson* and *Shiffrin* model [24.20], which assumed that:

1. Successful conversion to long-term storage required information to be held in the short-term store.
2. The short-term store was essential for access to long-term memory (LTM).
3. The short-term store was a seat of intellectual ability.

These assumptions were challenged by evidence that long-term memory (LTM) was dissociable from STM, as seen in neurological patients [24.21], and that complex encoding strategies, as opposed to simple rehearsal, predicted recall success [24.22].

Baddeley and *Hitch* [24.3] conducted their own series of dual task studies where they demonstrated that people could manage a concurrent task while holding items in memory for later recall with just a small decrement in processing speed, something that should not have been possible according to previous memory models. The concept of a unitary, passive short-term store was consigned to history and replaced with the WM model, which separated out attentional control from storage processes in memory. The original WM model proposed an attentional control system (the *central executive*) that operated in conjunction with two subsidiary systems: the visuospatial sketchpad and the phonological loop. Each of the subsidiary systems contained tightly coupled storage and rehearsal elements.

The central executive component was underresearched for many years, leaving the *homunculus* of the WM system poorly understood. After a decade, *Baddeley* [24.23] drew on a new theory proposed by *Norman* and *Shallice* [24.24] to suggest that the executive comprised a supervisory attentional system that allocated limited resources between the storage systems depending on task demands. More recent studies have confirmed the central executive's role in complex processes such as focusing and dividing attention, and task switching [24.6]. The true extent of the central executive's role in higher cognition however remains a highly contentious issue [24.25].

The *phonological loop* represents speech-based or verbal materials in WM. The loop comprises a short-term store and an articulatory rehearsal mechanism, both of which are described in more detail below. Theory regarding the development of the phonological loop framework was directly informed by behavioral research, which demonstrated that:

1. Storage within the loop is phonological as opposed to visual (the phonological similarity effect [24.26]).
2. Spoken material gains obligatory access to the storage component (the articulatory suppression effect [24.27]).
3. Subvocal rehearsal is likely to occur in real time (the word length effect [24.28]).

The *visuospatial sketchpad* is responsible for processing and storing both visual and spatial information. There is a degree of dissociation between visual and spatial memory within this WM component as evidenced from findings of behavioral dual task paradigms, [24.29] and the neuroimaging literature [24.30].

The *episodic buffer* was introduced as a fourth component to the WM model to acknowledge the mutual interaction between LTM and WM [24.31]. As a limited capacity system, the episodic buffer is thought to:

1. Bind information from the subsidiary systems
2. Store information in a multimodal code
3. Enable the interaction of resources between the WM and LTM systems [24.6, 32].

24.2 WM: Behavioral Data

24.2.1 Verbal Information

Baddeley and *Hitch's* [24.3] multicomponent WM model predicts that verbal information is processed by the phonological loop. As described above, this component can be further subdivided into (i) the *phonological store*, a passive storage component and (ii) the *articulatory loop*, an active rehearsal mechanism. The store can maintain speech-based information only for a few seconds [24.2, 5]. If longer maintenance is required,

then the articulatory loop can rehearse the information, a process comparable to subvocal speech [24.1].

The phonological loop component of the WM model continues to provide a parsimonious account of a number of verbal memory phenomena. Firstly, the phonological similarity effect describes how memory span is lower for lists of visually presented acoustically similar verbal items compared to lists of dissimilar sounding verbal items (B, V, P, T, G versus X, R, J, Y, S). Furthermore, errors made in this task are typically phonological rather than visual (i. e., recalling a C instead of a T). As a result of this finding, *Conrad* [24.33] argued that short-term storage of verbal materials must rely on recoding visual letters into subvocal auditory codes. (Note that such subvocal processing presumably involves sensorimotor-related codes, i. e., codes related to vocalization. This issue will become important when we discuss the *tonal loop*.) This argument has remained a principle of the WM model ever since. Phonological similarity is disruptive to memory as the phonological codes become confused either during the storage or retrieval phases of recall [24.34, 35].

A second, well-established finding in verbal memory research is that maintenance and rehearsal of stored material in the articulatory loop can be interrupted by articulatory suppression [24.1, 2], and overt or covert movement of the articulators including the jaw, tongue and throat musculature [24.28, 34, 36–39]. According to theory, articulatory suppression gains obligatory control over at least part of the articulatory loop within WM, which then reduces the overall functionality of the phonological loop by preventing the basic recoding of visually presented materials into (sensorimotor-related) phonological codes and/or disrupting the rehearsal process [24.27, 40].

Finally, the word length effect demonstrates that subvocal rehearsal within the articulatory loop is likely to occur in real time. People achieve greater memory spans [24.28] and superior recognition accuracy [24.36] for lists of short words compared to lists of long words. Although the word length effect has proved to be one of the most contentious aspects of the WM model [24.41], both the effects of articulatory suppression and of word length suggest that verbal material is maintained within WM in a manner comparable to subvocal speech [24.1, 2, 6].

24.2.2 Tonal Information

Because the *Baddeley* and *Hitch* WM model [24.1, 3] was developed to explain mainly verbal and visual-spatial information, the framework does not specify whether nonphonological information is processed by the phonological loop, or whether, in addition to the phonological loop, different subsystems exist for other materials such as tones [24.12, 15]. We have seen above that verbal material is maintained in WM by internal articulatory rehearsal; can tonal pitch also be maintained in this way?

Studies have found that internal rehearsal improved WM performance for tones [24.14–16]. Other studies, however, report that internal rehearsal only leads to a small improvement of WM performance or no improvement at all [24.42–44]. These conflicting results could be explained by the fact that the experimental stimuli used across these studies differed to the degree to which participants can imitate them. Studies that found a small or no beneficial effect of internal rehearsal used auditory stimuli that were difficult to sing or repeat, because:

1. The pitches of the tones did not correspond to the Western chromatic scale [24.43, 44].
2. The pitch difference between the tones was smaller than one semitone [24.43, 44].
3. Chords were used, which consisted of several simultaneously played sine wave tones [24.42].

By comparison, studies that found a benefit of rehearsal used auditory stimuli with (i) pitches that corresponded to the Western chromatic scale [24.15, 16] and/or (ii) pitch differences larger than one semitone [24.14–16], and are thus easy to sing/repeat. Overall therefore, the data support the assumption that internal rehearsal mechanisms underlie WM for tones.

24.2.3 Comparison Between Verbal and Tonal WM

The majority of auditory WM studies to date have used verbal materials (phonemes, syllables, and words) to explore both storage and rehearsal processes. By contrast, research on WM for pitch, or research comparing verbal and tonal WM, is rather scant and to date provides an inconsistent picture.

Deutsch [24.45] was among the first to conduct an auditory WM experiment and reported that intervening tones interfered more strongly with memory for a single tone compared to intervening phonemes. Deutsch interpreted this result as evidence for a specialized tonal WM system, at least when it comes to the storage of single tones. Furthermore, *Salame* and *Baddeley* [24.46] observed that vocal music interfered more with verbal WM than instrumental music, again supporting the assumption of two independent WM systems for verbal and tonal stimuli.

Semal et al. [24.47], however, criticized the lack of control of the pitch relations between the standard tones

and the intervening verbal material in *Deutsch*'s [24.45] study, and argued that this could explain the missing interference between the to-be-remembered tones and the intervening verbal material. In their experiment, pitch similarity of the intervening stimuli, which were either words or tones, modified WM performance to a greater degree than the nature of their modality, suggesting that pitch for both verbal and tonal material was processed in the same WM system. This result was replicated by *Ueda* [24.48] using different stimuli. Furthermore, *Chan* et al. [24.49] showed that better verbal WM performance was associated with musical training, a result also pointing towards overlapping mechanisms underlying verbal and tonal WM.

Few behavioral studies have attempted to compare WM storage using sequences of verbal and tonal materials. Whilst it seems difficult to recreate classic WM storage effects such as phonological similarity in the musical domain, some studies have drawn close parallels in order to compare verbal and tonal WM. *Williamson* et al. [24.18] argued that the phonological similarity effect was based on acoustic confusion at the point of memory storage, an effect that could be represented in music by pitch proximity [24.47]. In *Williamson* et al. [24.18] nonmusician participants recalled sequences of phonologically similar (B, D, G) or dissimilar (M, Q, R) letters and pitch proximal (C4, D4, E4) or distant tones (C4, G4, B4), the notes being balanced according to theory of tonal hierarchy [24.50]. The results replicated the traditional phonological similarity effect with verbal items and saw in parallel a smaller effect of pitch proximity with tonal stimuli. Participants were pretested for their ability to perceive the difference between the tones, meaning that acoustic confusion within memory storage was the most likely explanation for the observed effect. This study supports the theory that storage of both verbal and tonal material in WM is based on acoustic representations of sound. Whilst this finding argues for similar processing constraints in verbal and tonal WM storage it does not, in of itself, support storage system overlap; the principles of storage may be similar for verbal and tonal materials but storage could still be separable, based on the individual codes of speech (phonological store) and tones (tonal store) [24.51].

Studies of rehearsal activity in WM for verbal and tonal material more clearly show support for overlap than is necessarily justified for storage processes. *Schendel* and *Palmer* [24.16] demonstrated that when musicians carry out music suppression (singing *la*) there was decreased recognition accuracy for both digit and tone sequences, in the same manner as with verbal suppression (producing the word *the*), indicating that musical or verbal suppression does not selectively impair verbal or tonal WM. (To our knowledge, there is lack of studies investigating effects of nonphonological sensorimotor suppression, such as moving the fingers of the left hand in string players during a pitch WM task.) *Williamson* [24.52] required nonmusician participants to carry out whispered articulatory suppression while encoding and recalling sequences of letters and tones. She reported that both verbal and tonal WM tasks showed a decrement under articulatory suppression, a finding that could not be accounted for by the dual-task nature of the experiment. This result suggests that subvocal rehearsal of verbal and tonal material may take place in a similar way within the articulatory loop.

24.3 Neural Correlates Underlying WM

24.3.1 Verbal Information

As with behavioral work, most neuroimaging studies that have explored the functional neuroarchitecture underlying auditory WM utilize verbal material. The internal rehearsal of verbal material has been found to be supported by a functional network comprised mainly of Broca's area and premotor areas (as well as the supplementary motor area (SMA) and pre-SMA) [24.53–57]. In addition, data indicates that both the insular cortex [24.54, 58, 59] and the cerebellum [24.57, 60, 61] are involved during the internal rehearsal of verbal information.

The crucial role of Broca's area and the premotor cortex for internal rehearsal has been corroborated by numerous studies (for an overview see [24.1]). Findings regarding the phonological store, however, are much less conclusive. It has been suggested that parietal areas, particularly the inferior parietal lobule [24.53, 54, 56, 60–64], but also the superior parietal lobule [24.53, 60], serve as the underlying neural correlate of the phonological store. However, pinpointing the phonological store in the parietal lobe is controversial. Firstly, neural activity in the parietal lobe has been associated with attention, and therefore activation in this area might also reflect increased engagement of attentional resources [24.65, 66]. Secondly, the coordinates of the reported activation during tasks requiring storage of verbal material differ greatly between studies [24.4]. And finally, passive listening does not activate the infe-

rior parietal lobule (IPL) [24.4, 67], a finding predicted by the WM model [24.3, 4] if this structure is involved in automatically storing incoming auditory information.

As a consequence of these theoretical and empirical inconsistencies, area Spt (Sylvian-parietal-temporal, i. e., left posterior planum temporale), has been suggested as an alternative neural structure underlying the temporary storage of verbal information during WM tasks [24.4] based on the following observations. First, activation in the left Spt was (i) enhanced during the delay period of a WM task [24.13, 68] and (ii) not influenced by the modality of the presented stimuli (auditory or visual [24.68]). In addition, area Spt is involved during the perception and production of speech. Based on these findings it has been suggested that area Spt acts as an auditory-motor interface for WM [24.4, 13, 68], meaning that perception, production and WM are more closely linked than previously believed. This proposition is corroborated by the dual-stream model of speech processing [24.69–72]. This model assumes a ventral stream that supports speech comprehension via lexical access and a left dominant dorsal stream comprising also area Spt, which enables humans to map perceived speech signals onto articulatory representations, and therefore supports sensory–motor integration.

24.3.2 Tonal Information

Far fewer neuroimaging studies have investigated WM for tones than have investigated the phonological loop. *Gaab* et al. [24.73] showed the involvement of the supramarginal gyrus (SMG), the intraparietal sulcus (IPS), the planum temporale, premotor regions encroaching on Broca's area, and cerebellar regions during a pitch memory task in participants that were not selected for musical expertise. The neural network observed in this study is surprisingly similar to the network supporting the phonological loop described above. Another group had previously observed activation of a similar network during a WM task for tones, including the inferior frontal and insular cortex, the planum temporale, and the SMG [24.74].

24.3.3 Comparison Between Verbal and Tonal WM

To our knowledge, only three neuroimaging studies have compared the neuroarchitecture of auditory WM for sequential tonal and verbal material [24.13, 14, 17]. In a functional magnetic resonance imaging (fMRI) experiment, *Hickok* et al. [24.13] studied nonmusicians and compared the neural correlates underlying verbal and tonal WM. The authors presented piano melodies during a tonal condition and pseudoword sentences during a verbal condition. The internal rehearsal of the verbal and tonal materials involved activation of area Spt and premotor regions encroaching on Broca's area [24.13]. A similar functional network was uncovered in a WM recognition study by *Koelsch* et al. [24.14] in which neural similarities between WM for verbal (syllables) and tonal (pitch) material were explored, again in nonmusicians. The functional network observed during verbal rehearsal included the premotor cortex, the anterior insula, the SMG/IPS, the planum temporale, the inferior frontal gyrus, pre-SMA and the cerebellum. Importantly, internal rehearsal of the tonal stimuli relied on a virtually identical network to that activated during verbal rehearsal.

In a follow-up study, *Schulze* et al. [24.17] employed a recognition task to compare the neuroarchitecture supporting the internal rehearsal of verbal and tonal WM. In a replication of the above-described experiment [24.14], both verbal and tonal WM tasks activated structures previously reported during either verbal [24.1, 53, 54, 56] or tonal [24.13, 73, 74] WM processes in nonmusicians, namely Broca's area, the left premotor cortex, (pre-)SMA, left insular cortex, and left IPL. This finding corroborates data from previous experiments, which have reported considerable overlap of the functional networks subserving verbal and tonal WM [24.13, 14]. Importantly, all these core structures activated during tonal WM were also activated during verbal WM in nonmusicians. By contrast, verbal but not tonal WM relied on additional structures, including the right ventrolateral premotor cortex, right IPL, right cerebellum and the left mid-dorsolateral prefrontal cortex (mid-DLPFC). These structures have previously been associated with verbal WM tasks [24.1, 75, 76]. This activation difference between verbal and tonal WM was also expressed in the behavioral data: nonmusicians showed superior performance during verbal than during tonal WM.

In summary, the few neuroimaging studies that have directly compared the structures underlying verbal and tonal WM [24.13, 14, 17] have obtained data from nonmusicians that point towards a considerable overlap of neural resources underlying WM for verbal and tonal material. This common (core) network is mainly left-lateralized and includes fronto-parietal structures, that is premotor cortex, Broca's area, and in two of the three studies also the IPL [24.14, 17] and the cerebellum [24.14, 17], as well as the planum temporale/area Spt [24.13, 14]. (*Schulze* et al. [24.17] did not scan continuously like in the other fMRI studies [24.13, 14], but employed a sparse-temporal sampling scanning technique. Thus, participants did not have to separate scanner noise from the information they had to rehearse, possibly leading to less (or no) activation of

area Spt. Another possibility is that the sparse temporal scanning method was not sensitive enough to capture the involvement of area Spt.)

24.3.4 Comparison Between Nonmusicians and Musicians

Speech is a fundamental human skill typically acquired during early childhood. Thus, while both musicians and nonmusicians are trained in processing and producing speech, nonmusicians' expertise in the music domain is comparatively far less developed. This begs the question of whether tonal WM would be different in musicians compared to nonmusicians, where a group difference in verbal WM would not be predicted. In support of this hypothesis *Williamson* et al. [24.18] found that nonmusicians showed effects of both phonological similarity and pitch proximity, whereas musicians showed the former but not the latter. It was suggested that musicians, thanks to their specialized training, develop systems and/or strategies for storing tones that negate the impact of pitch proximity. Nonmusicians, on the other hand, are more likely to store a basic acoustic representation of the tonal sounds that they hear, which are then more vulnerable to the impact of pitch proximity.

Schulze et al. [24.17] directly studied the neural underpinnings of verbal and tonal WM in highly trained musicians to see how they compared to the established WM network in nonmusicians. The aim was to determine if musicians had developed a specialized or alternative system for processing tones. Many of the structures (Broca's area, left premotor cortex, left insular cortex, (pre-)SMA, and left IPL) that were activated more strongly in nonmusicians during verbal WM showed an increased involvement in musicians compared to nonmusicians during tonal WM. That is, the functional network that supported verbal WM in nonmusicians was also used by musicians for tonal WM tasks. In addition, unlike nonmusicians, musicians recruited a number of structures exclusively for either verbal or tonal WM: the left cuneus, the right globus pallidus, the right caudate nucleus, and the left cerebellum supported WM for tonal material; the right insular cortex was exclusively involved in processing verbal information.

On top of these differences, activation distinctions between verbal and tonal WM tasks were reported for a number of structures in musicians, thus supporting the existence of two WM systems, potentially a phonological loop maintaining and processing phonological information and a tonal loop maintaining and processing tonal information. Both WM systems showed a considerable overlap in activation because the same WM core structures were activated. Differences, however, were also observed for both systems in that they relied on different neural subcomponents. Importantly, it is not possible to explain the observed functional differences between the verbal and tonal WM tasks in musicians by citing performance differences between both tasks, because several brain structures were recruited selectively for verbal or tonal WM [24.17, 77]. Instead, musical expertise might facilitate the development of a more extended network underlying tonal WM, which shows similarities to the functional network supporting verbal WM, but also substantial differences.

24.4 Sensorimotor Codes – Auditory WM and the Motor System

To acknowledge the interconnection of verbal WM with speech perception and production, the underlying representations of verbal WM have been named sensorimotor codes [24.78]. In the following we explain and discuss results that indicate internal WM rehearsal for verbal material shares characteristics with speech production.

The word length effect and the articulatory suppression effect, which were described above, support the assumption that verbal WM is similar to subvocal speech (for an overview see [24.1]). In addition, the phonological loop has been conceptualized as a memory system that involves internal articulatory speech actions implemented by motor-related areas such as Broca's area, premotor and insular cortices [24.1, 58], SMA [24.53, 54, 57], and the cerebellum [24.1, 57, 60]. As discussed above, results suggest that tonal WM can be improved by internal rehearsal only if participants are able to imitate and repeat these auditory stimuli [24.14–16, 42–44]. WM thus seems more closely linked to production (i. e., action-related) processes than previously believed. This assumption has been supported by data that compares the neural correlates of verbal and tonal WM between musicians and nonmusicians described above [24.17]. The better performance of nonmusicians for verbal compared to the tonal WM tasks, and the superior performance of musicians compared to nonmusicians for tonal WM tasks, was primarily associated with activation differences in neural structures that have been described to support planning, programming, executing and controlling actions, like Broca's area, premotor cortex, (pre-)SMA,

left insular cortex, intraparietal sulcus and IPL, and the cerebellum.

Initially we interpreted these observed behavioral and neurophysiological differences between verbal and tonal WM in nonmusicians as a consequence of the extensive production and rehearsal of verbal material in their daily life, which for nonmusicians is not the case, or is to a lesser degree, for musical material. An alternative but related interpretation is that musical training could lead to a long-term learning of associations between pitch information and motor actions [24.79–83]. As a result of this process, musicians could have developed more elaborate sensorimotor codes supporting the internal rehearsal of tones compared to nonmusicians. Such codes probably include, e.g., finger representations, and codes of finger movements, in piano and string players.

To our knowledge no studies have investigated whether the ability to repeat an auditory stimulus might facilitate its internal rehearsal and therefore increase auditory WM capacity, but there is some indication that mimicry is important for auditory WM [24.84–86]. For example, WM processes for timbres, which are difficult or impossible to mimic, differ from those for tones and words, which are easy to mimic: Whereas WM for timbre seems to rely on a passive sensory trace, WM for verbal and tonal stimuli appears to be maintained by active internal rehearsal processes [24.85]. Furthermore, distractors that share features with hard-to-mimic auditory stimuli are likely to overwrite and degrade their memory trace [24.87, 88].

Interestingly, monkeys, who are not vocal learners and therefore cannot learn to mimic unfamiliar auditory stimuli, seem to rely only on a *passive form of auditory STM* (the authors [24.89] referred to a passive form of STM to explicitly distinguish it from WM, therefore this term has also been used here. Otherwise, as explained in the introduction, in the present chapter no distinction has been made between STM and WM): *Scott* et al. [24.89] used a small set of sounds from different categories (pure tones, environmental sounds, monkey calls, etc.) and observed that monkeys performed very poorly on an auditory serial delayed match-to-sample task. The authors observed an *overwriting* effect for auditory stimuli (including monkey vocalizations) that was far greater compared to that in the visual domain, indicating that the observed performance in monkeys depended on the passive form of STM and not on WM. Interestingly, when humans are tested on auditory stimuli that they cannot mimic, they too seem to rely on passive auditory STM [24.87, 88].

The idea that auditory WM relies on the assistance of the motor system is further supported by patient studies. Speech disorders such as speech apraxia, a disorder of speech planning and programming [24.90], are associated with decreased performance in a verbal WM task [24.91]. Another piece of evidence comes from the investigation of the three-generational KE family, half of whom suffer from a speech and language disorder caused by a mutation of the FOXP2 gene [24.92, 93]. Although the speech impairments of the affected KE (aKE) family members are widespread, their core deficit is in executing orofacial, especially articulatory, sequences. The structural and functional cortical abnormalities in the aKE also include the inferior frontal (Broca's) area and ventral premotor cortices [24.92, 94–96]. The first study to compare WM in aKE and controls used a test [24.97] based on the Baddeley and Hitch WM model and analyzed the different components of WM separately – the central executive, the phonological loop, and the visuospatial sketchpad [24.98]. Compared to controls (the control group was matched to the aKE for age and performance IQ), the aKE performed only worse for the tasks related to the phonological loop. Importantly, the aKE members were also impaired in the recognition-based subtest of the phonological loop *word list matching*, in which repetition (i. e., motor output) of speech-based material is not required. Results thus suggest that the aKE, who show both structural and functional abnormalities in Broca's area, a structure underlying phonological and auditory WM (see above), could be specifically impaired in phonological WM but not other domains of WM. This indicates an association between the speech difficulties of the aKE members and their representations underlying internal rehearsal of speech-based material in phonological WM [24.98].

Interestingly, just as with auditory WM, auditory LTM appears to require the assistance of the (oro)motor system. *Schulze* et al. [24.99] tested whether a sound that can be neither mimicked nor labeled can be stored as a long-lasting representation for subsequent recognition by comparing participants' ability to recognize different auditory stimuli that varied widely in the degree to which they could be reproduced or labeled (words, pseudowords, nonverbal sounds, and reversed words, i. e., words played backwards, were presented auditorily). Participants listened to a list of 10 stimuli once for familiarization (study list), and then, after a 5 min interval filled with a counting task to block WM, were presented with these 10 items from the study list again in random order intermixed with 10 new items (recognition list), and had to make an old-new judgment for each of these items. The results were clear; participants showed recognition difficulty only for the reversed words, which were hard either to mimic or to label with an associate. Importantly, a control experiment demonstrated that recognition difficulty was

not due to a perceptual failure. The participants' poor memory performance of the reversed words supports the proposal that a sound's pronounceability – its potential to activate subvocally the speech production system, potentially via an auditory mirror-neuron system [24.100, 101] – may be essential for generating a long-lasting representation of that sound.

In summary, sensorimotor processes are thought to be involved in the rehearsal and representation of information in auditory, verbal and tonal WM [24.13, 14, 17, 102, 103] and thus may play in important role during the representation and manipulation of auditory information. These action-related sensorimotor codes are assumed to rely on motor knowledge – how to produce the auditory stimulus (e.g., syllable, tone). The dual-stream model of speech processing [24.69, 71, 72] suggests that sensory–motor integration, i. e., mapping the perceived speech signals onto articulatory representations is one of the functions of the dorsal path of the auditory system. Results indicate that the superior longitudinal as well as the arcuate fasciculi [24.104] form the dorsal stream. Speech production requires motor speech representations as well as representations of sensory speech targets to compare between predicted and actual consequences of motor speech acts [24.70]. Sensorimotor integration also seems to play a role during singing [24.105, 106]. Therefore, we propose that internal rehearsal associated with auditory WM relies on sensorimotor representations, which also underlie singing and speaking.

24.5 The Influence of LTM on Auditory WM Performance

WM is a limited-capacity system in terms of how much information can be stored and for how long [24.1, 28, 107]. However, the use of a strategy based on information stored in LTM can improve WM performance, for example by chunking the to-be-remembered information [24.108, 109]. During chunking, items are organized into one unit or chunk [24.107], resulting in stronger associations between items within one chunk than between chunks [24.110]. This process is thought to be supported by the episodic buffer, which enables features from different sources to be bound [24.6]. Previously, the neural correlates underlying such strategy-based memorization were explored for the visual–spatial or verbal domain [24.111–114], but it was largely unknown whether a similar functional network was also involved during the strategy-based WM for tones.

By using structured (all tones belonged to one tonality) and unstructured (atonal) five-tone sequences, *Schulze* et al. [24.115] investigated: (i) whether musical structure influences performance on a nonverbal auditory WM task, and if so, (ii) how this is reflected in the brain of nonmusicians and musicians. Musicians, but not nonmusicians, performed better for the structured than for the unstructured sequences, indicating that musicians' knowledge about musical regularities helped them to maintain the structured sequences in WM [24.116–119]. In terms of brain responses, musicians showed stronger involvement of a lateral (pre)frontal–parietal network during the memorization of the structured sequences, including the right inferior precentral sulcus and the premotor cortex, as well as the left IPS. A similar network has been described previously to support strategy-based WM processing for visual and auditory–verbal stimuli [24.111, 113, 114]. The combined results point towards a modality-independent (pre)frontal–parietal network subserving strategy-based WM. A follow-up behavioral study by *Schulze* et al. [24.120] has confirmed the facilitating effect of tonality (structure) on tonal WM performance in both musicians and nonmusicians, but only during memory maintenance (forward task) and not complex memory manipulation (backward task).

24.6 Summary and Conclusion

This chapter reviewed and discussed research results demonstrating behavioral, structural, and functional differences and similarities between verbal and tonal WM. Whereas the core structures, namely Broca's area, premotor cortex, and IPL, show a substantial overlap, results in musicians suggest that there are also different subcomponents involved during verbal and tonal WM tasks. If confirmed, these results indicate that musicians develop either independent tonal and phonological loops or unique processing strategies that allow novel interactive use of the WM systems. In addition we discussed behavioral and neuroimaging results that provide substantial support for a strong link between sound mimicry and auditory

WM. Sensorimotor processes are thought to be involved in the rehearsal and representation of information in auditory WM. These action-related sensorimotor codes are assumed to rely on motor knowledge – how to produce the auditory stimulus (e.g., syllable, tone).

References

24.1 A.D. Baddeley: Working memory: Looking back and looking forward, Nat. Rev. Neurosci. **4**(10), 829–839 (2003)

24.2 A.D. Baddeley: Working memory, Science **255**(5044), 556–559 (1992)

24.3 A.D. Baddeley, G.J. Hitch: Working memory. In: *Recent Advances in Learning and Motivation*, ed. by G.A. Bower (Academic, New York 1974) pp. 47–89

24.4 B.R. Buchsbaum, M. D'Esposito: The search for the phonological store: From loop to convolution, J. Cogn. Neurosci. **20**(5), 762–778 (2008)

24.5 A.D. Baddeley: Working memory, Curr. Biol. **20**(4), R136–R140 (2010)

24.6 A.D. Baddeley: Working memory: Theories, models, and controversies, Annu. Rev. Psychol. **63**, 1–29 (2012)

24.7 N. Cowan: Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system, Psychol. Bull. **104**(2), 163–191 (1988)

24.8 N. Cowan: An embedded-processes model of working memory. In: *Models of Working Memory*, ed. by A. Miyake, P. Shah (University Press, Cambridge 1999) pp. 62–101

24.9 K.A. Ericsson, W. Kintsch: Long-term working-memory, Psychol. Rev. **102**(2), 211–245 (1995)

24.10 D.M. Jones: Objects, streams and threads of auditory attention. In: *Attention: Selection, Awareness and Control*, ed. by A.D. Baddeley, L. Weiskrantz (Clarendon, Oxford 1993) pp. 87–104

24.11 J.S. Nairne: A feature model of immediate memory, Mem. Cogn. **18**(3), 251–269 (1990)

24.12 W.L. Berz: Working memory in music: A theoretical model, Music Percept **12**(3), 353–364 (1995)

24.13 G. Hickok, B. Buchsbaum, C. Humphries, T. Muftuler: Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt, J. Cogn. Neurosci. **15**(5), 673–682 (2003)

24.14 S. Koelsch, K. Schulze, D. Sammler, T. Fritz, K. Muller, O. Gruber: Functional architecture of verbal and tonal working memory: An fMRI study, Hum. Brain Mapp. **30**(3), 859–873 (2009)

24.15 T. Pechmann, G. Mohr: Interference in memory for tonal pitch: Implications for a working-memory model, Mem. Cogn. **20**(3), 314–320 (1992)

24.16 Z.A. Schendel, C. Palmer: Suppression effects on musical and verbal memory, Mem. Cogn. **35**(4), 640–650 (2007)

24.17 K. Schulze, S. Zysset, K. Mueller, A.D. Friederici, S. Koelsch: Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians, Hum. Brain Mapp. **32**, 771–783 (2011)

24.18 V.J. Williamson, A.D. Baddeley, G.J. Hitch: Musicians' and nonmusicians' short-term memory for verbal and musical sequences: Comparing phonological similarity and pitch proximity, Mem. Cogn. **38**(2), 163–175 (2010)

24.19 N. Cowan: What are the differences between long-term, short-term, and working memory?, Essence Mem **169**, 323–338 (2008)

24.20 R.C. Atkinson, R.M. Shiffrin: Human memory: A propsed system and its control processes. In: *The Psychology of Learning and Motivation: Advances in Research and Theory*, ed. by K.W. Spence, J.T. Spence (Academic, New York 1968) pp. 89–195

24.21 T. Shallice, E.K. Warrington: Independent functioning of verbal memory stores: A neuropsychological study, Q. J. Exp. Psychol. **22**(2), 261–273 (1970)

24.22 F.I.M. Craik, R.S. Lockhart: Levels of processing–framework for memory research, J. Verbal Learn. Verbal Behav. **11**(6), 671–684 (1972)

24.23 A.D. Baddeley: *Working Memory Oxford Psychology Series*, Vol. 11 (Clarendon, Oxford 1986)

24.24 D.A. Norman, T. Shallice: Attention to action: Willed and automatic control of behaviour. In: *Consciousness and Self-Regulation. Advances in Research and Theory*, ed. by R.J. Davidson, G.E. Schwartz, D. Shapiro (Plenum, New York 1986) pp. 1–18

24.25 R.W. Engle, M.J. Kane: Executive attention, working memory capacity, and a two-factor theory of cognitive control, Psychol. Learn. Motiv. **44**, 145–199 (2004)

24.26 R. Conrad, A.J. Hull: Information, acoustic confusion and memory span, Br. J. Psychol. **55**, 429–432 (1964)

24.27 A.D. Baddeley, V. Lewis, G. Vallar: Exploring the articulatory loop, Q. J. Exp. Psychol. Sect. A **36**(2), 233–252 (1984)

24.28 A.D. Baddeley, N. Thomson, L. Buchanan: Word length and the structure of short-term memory, J. Verbal Learn. Verbal Behav. **14**(6), 575–589 (1975)

24.29 R.H. Logie: Visuo-spatial processing in working memory, Q. J. Exp. Psychol. A **38**(2), 229–247 (1986)

24.30 E.E. Smith, J. Jonides: Working memory: A view from neuroimaging, Cogn. Psychol. **33**(1), 5–42 (1997)

24.31 A.D. Baddeley: The episodic buffer: A new component of working memory?, Trends Cogn. Sci. **4**(11), 417–423 (2000)

24.32 R.J. Allen, A.D. Baddeley, G.J. Hitch: Is the binding of visual features in working memory resource-demanding?, J. Exp. Psychol. Gen. **135**(2), 298–313

(2006)
24.33 R. Conrad: Acoustic confusions in immediate memory, Br. J. Psychol. **55**, 75–84 (1964)
24.34 A.M. Surprenant, I. Neath, D.C. LeCompte: Irrelevant speech, phonological similarity, and presentation modality, Memory **7**(4), 405–420 (1999)
24.35 A. Baddeley (Ed.): *Human Memory: Theory and Practice* (Psychology, East Sussex 1997)
24.36 A.D. Baddeley, D. Chincotta, L. Stafford, D. Turk: Is the word length effect in STM entirely attributable to output delay? Evidence from serial recognition, Q. J. Exp. Psychol. Sect. A **55**(2), 353–369 (2002)
24.37 R.N.A. Henson, T. Hartley, N. Burgess, G. Hitch, B. Flude: Selective interference with verbal short-term memory for serial order information: A new paradigm and tests of a timing-signal hypothesis, Q. J. Exp. Psychol. Sect. A **56**(8), 1307–1334 (2003)
24.38 J.D. Larsen, A.D. Baddeley: Disruption of verbal STM by irrelevant speech, articulatory suppression, and manual tapping: Do they have a common source?, Q. J. Exp. Psychol. Sect. A **56**(8), 1249–1268 (2003)
24.39 I. Neath, A.M. Surprenant, D.C. LeCompte: Irrelevant speech eliminates the word length effect, Mem. Cogn. **26**(2), 343–354 (1998)
24.40 T. Meiser, K.C. Klauer: Working memory and changing state hypothesis, J. Exp. Psychol. Learn. Mem. Cogn. **25**(5), 1272–1299 (1999)
24.41 C. Hulme, A.M. Suprenant, T.J. Bireta, G. Stuart, I. Neath: Abolishing the word-length effect, J. Exp. Psychol. Learn. Mem. Cogn. **30**(1), 98–106 (2004)
24.42 L. Demany, G. Montandon, C. Semal: Pitch perception and retention: two cumulative benefits of selective attention, Percept. Psychophys. **66**(4), 609–617 (2004)
24.43 C. Kaernbach, K. Schlemmer: The decay of pitch memory during rehearsal, J. Acoust. Soc. Am. **123**(4), 1846–1849 (2008)
24.44 T.A. Keller, N. Cowan, J.S. Saults: Can auditory memory for tone pitch be rehearsed?, J. Exp. Psychol. Learn. Mem. Cogn. **21**(3), 635–645 (1995)
24.45 D. Deutsch: Tones and numbers: Specificity of interference in immediate memory, Science **168**(939), 1604–1605 (1970)
24.46 P. Salame, A.D. Baddeley: Effects of background music on phonological short-term memory, Q. J. Exp. Psychol. **41**(A), 107–122 (1989)
24.47 C. Semal, L. Demany, K. Ueda, P.A. Halle: Speech versus nonspeech in pitch memory, J. Acoust. Soc. Am. **100**(2 Pt 1), 1132–1140 (1996)
24.48 K. Ueda: Short-term auditory memory interference: The Deutsch paradigm revisited, J. Acoust. Soc. Japan **25**(6), 457–467 (2004)
24.49 A.S. Chan, Y.C. Ho, M.C. Cheung: Music training improves verbal memory, Nature **396**(6707), 128 (1998)
24.50 C.L. Krumhansl: *Cognitive Foundations of Musical Pitch* (Oxford Univ. Press, Oxford 1990)
24.51 A.D. Patel: Language, music, syntax and the brain, Nat. Neurosci. **6**(7), 674–681 (2003)
24.52 V. Williamson: *Comparing Short-Term Memory for Sequences of Verbal and Tonal Materials*, Ph.D. Thesis (Univ. of York, York 2008), available from http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.550492
24.53 E. Awh, J. Jonides, E.E. Smith, E.H. Schumacher, R.A. Koeppe, S. Katz: Dissociation of storage and rehearsal in verbal working memory: Evidence from positron emission tomography, Psychol. Sci. **7**(1), 25–31 (1996)
24.54 E. Paulesu, C.D. Frith, R.S. Frackowiak: The neural correlates of the verbal component of working memory, Nature **362**(6418), 342–345 (1993)
24.55 J.A. Fiez, E.A. Raife, D.A. Balota, J.P. Schwarz, M.E. Raichle, S.E. Petersen: A positron emission tomography study of the short-term maintenance of verbal information, J. Neurosci. **16**(2), 808–822 (1996)
24.56 O. Gruber, D.Y. von Cramon: The functional neuroanatomy of human working memory revisited. Evidence from 3-T fMRI studies using classical domain-specific interference tasks, Neuroimage **19**(3), 797–809 (2003)
24.57 S.M. Ravizza, M.R. Delgado, J.M. Chein, J.T. Becker, J.A. Fiez: Functional dissociations within the inferior parietal cortex in verbal working memory, Neuroimage **22**(2), 562–573 (2004)
24.58 D.E. Bamiou, F.E. Musiek, L.M. Luxon: The insula (Island of Reil) and its role in auditory processing, literature review, Brain Res. Brain Res. Rev. **42**(2), 143–154 (2003)
24.59 J.M. Chein, J.A. Fiez: Dissociation of verbal working memory system components using a delayed serial recall task, Cereb. Cortex **11**(11), 1003–1014 (2001)
24.60 S.H. Chen, J.E. Desmond: Cerebrocerebellar networks during articulatory rehearsal and verbal working memory tasks, Neuroimage **24**(2), 332–338 (2005)
24.61 M.P. Kirschen, S.H. Chen, P. Schraedley-Desmond, J.E. Desmond: Load- and practice-dependent increases in cerebro-cerebellar activation in verbal working memory: An fMRI study, Neuroimage **24**(2), 462–472 (2005)
24.62 S. Crottaz-Herbette, R.T. Anagnoson, V. Menon: Modality effects in verbal working memory: Differential prefrontal and parietal responses to auditory and visual stimuli, Neuroimage **21**(1), 340–351 (2004)
24.63 R.N.A. Henson, N. Burgess, C.D. Frith: Recoding, storage, rehearsal and grouping in verbal short-term memory: An fMRI study, Neuropsychologia **38**(4), 426–440 (2000)
24.64 J. Jonides, E.H. Schumacher, E.E. Smith, R.A. Koeppe, E. Awh, P.A. Reuter-Lorenz, C. Marshuetz, C.R. Willis: The role of parietal cortex in verbal working memory, J. Neurosci. **18**(13), 5026–5034 (1998)
24.65 R. Cabeza, L. Nyberg: Imaging cognition II: An empirical review of 275 PET and fMRI studies, J. Cogn. Neurosci. **12**(1), 1–47 (2000)

24.66 M. Corbetta, G.L. Shulman: Control of goal-directed and stimulus-driven attention in the brain, Natl. Rev. Neurosci. **3**(3), 201–215 (2002)

24.67 J.T. Becker, D.K. MacAndrew, J.A. Fiez: A comment on the functional localization of the phonological storage subsystem of working memory, Brain Cogn **41**(1), 27–38 (1999)

24.68 B.R. Buchsbaum, R.K. Olsen, P. Koch, K.F. Berman: Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory, Neuron **48**(4), 687–697 (2005)

24.69 G. Hickok: The functional neuroanatomy of language, Phys. Life Rev. **6**(3), 121–143 (2009)

24.70 G. Hickok, J. Houde, F. Rong: Sensorimotor integration in speech processing: Computational basis and neural organization, Neuron **69**(3), 407–422 (2011)

24.71 G. Hickok, D. Poeppel: The cortical organization of speech processing, Nat. Rev. Neurosci. **8**(5), 393–402 (2007)

24.72 J.P. Rauschecker, S.K. Scott: Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing, Nat. Neurosci. **12**(6), 718–724 (2009)

24.73 N. Gaab, C. Gaser, T. Zaehle, L. Jancke, G. Schlaug: Functional anatomy of pitch memory – An fMRI study with sparse temporal sampling, Neuroimage **19**(4), 1417–1426 (2003)

24.74 R.J. Zatorre, A.C. Evans, E. Meyer: Neural mechanisms underlying melodic perception and memory for pitch, J. Neurosci. **14**(4), 1908–1919 (1994)

24.75 R. Cabeza, L. Nyberg: Neural bases of learning and memory: Functional neuroimaging evidence, Curr. Opin. Neurol. **13**(4), 415–421 (2000)

24.76 M. Petrides, B. Alivisatos, E. Meyer, A.C. Evans: Functional activation of the human frontal cortex during the performance of verbal working memory tasks, Proc. Natl. Acad. Sci. US **90**(3), 878–882 (1993)

24.77 T.D. Wager, E.E. Smith: Neuroimaging studies of working memory: A meta-analysis, Cogn. Affect Behav. Neurosci. **3**(4), 255–274 (2003)

24.78 M. Wilson: The case for sensorimotor coding in working memory, Psychon. Bull. Rev. **8**(1), 44–57 (2001)

24.79 M. Bangert, T. Peschel, G. Schlaug, M. Rotte, D. Drescher, H. Hinrichs, H.J. Heinze, E. Altenmuller: Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction, Neuroimage **30**(3), 917–926 (2006)

24.80 A. D'Ausilio, E. Altenmuller, M. Olivetti Belardinelli, M. Lotze: Cross-modal plasticity of the motor cortex while listening to a rehearsed musical piece, Eur. J. Neurosci. **24**(3), 955–958 (2006)

24.81 U.C. Drost, M. Rieger, M. Brass, T.C. Gunter, W. Prinz: When hearing turns into playing: Movement induction by auditory stimuli in pianists, Q. J. Exp. Psychol. A **58**(8), 1376–1389 (2005)

24.82 U.C. Drost, M. Rieger, M. Brass, T.C. Gunter, W. Prinz: Action-effect coupling in pianists, Psychol. Res. **69**(4), 233–241 (2005)

24.83 B. Haslinger, P. Erhard, E. Altenmuller, U. Schroeder, H. Boecker, A.O. Ceballos-Baumann: Transmodal sensorimotor networks during action observation in professional pianists, J. Cogn. Neurosci. **17**(2), 282–293 (2005)

24.84 A.R. Halpern, R.J. Zatorre, M. Bouffard, J.A. Johnson: Behavioral and neural correlates of perceived and imagined musical timbre, Neuropsychologia **42**(9), 1281–1292 (2004)

24.85 K. Schulze, B. Tillmann: Working memory for pitch, timbre, and words, Memory **21**(3), 377–395 (2013)

24.86 A.R. Halpern, R.J. Zatorre: When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies, Cereb. Cortex **9**(7), 697–704 (1999)

24.87 D. McKeown, R. Mills, T. Mercer: Comparisons of complex sounds across extended retention intervals survives reading aloud, Perception **40**(10), 1193–1205 (2011)

24.88 T. Mercer, D. McKeown: Updating and feature overwriting in short-term memory for timbre, Aten. Percept. Psychophys. **72**(8), 2289–2303 (2010)

24.89 B.H. Scott, M. Mishkin, P. Yin: Monkeys have a limited form of short-term memory in audition, Proc. Natl. Acad. Sci. US **109**(30), 12237–12241 (2012)

24.90 F.J. Liégeois, A.T. Morgan: Neural bases of childhood speech disorders: Lateralization and plasticity for speech functions during development, Neurosci. Biobehav. Rev. **36**(1), 439–458 (2012)

24.91 G.S. Waters, E. Rochon, D. Caplan: The role of high-level speech planning in rehearsal – evidence from patients with apraxia of speech, J. Mem. Lang. **31**(1), 54–73 (1992)

24.92 F. Vargha-Khadem, D.G. Gadian, A. Copp, M. Mishkin: FOXP2 and the neuroanatomy of speech and language, Nat. Rev. Neurosci. **6**(2), 131–138 (2005)

24.93 C.S.L. Lai, S.E. Fisher, J.A. Hurst, F. Vargha-Khadem, A.P. Monaco: A forkhead-domain gene is mutated in a severe speech and language disorder, Nature **413**(6855), 519–523 (2001)

24.94 F. Liégeois, A.T. Morgan, A. Connelly, F. Vargha-Khadem: Endophenotypes of FOXP2: Dysfunction within the human articulatory network, J. Eur. Paediatr. Neurol. Soc. **15**(4), 283–288 (2011)

24.95 K.E. Watkins, F. Vargha-Khadem, J. Ashburner, R.E. Passingham, A. Connelly, K.J. Friston, R.S. Frackowiak, M. Mishkin, D.G. Gadian: MRI analysis of an inherited speech and language disorder: Structural brain abnormalities, Brain **125**, 465–478 (2002)

24.96 F. Liegeois, T. Baldeweg, A. Connelly, D.G. Gadian, M. Mishkin, F. Vargha-Khadem: Language fMRI abnormalities associated with FOXP2 gene mutation, Nat. Neurosci. **6**(11), 1230–1237 (2003)

24.97 S.J. Pickering, S.E. Gathercole: *Working Memory Test Battery for Children (WMTB-C)* (Pearson, London 2001)

24.98 K. Schulze, F. Vargha-Khadem, M. Mishkin: Phonological working memory and FOXP2 (2017)

submitted for publication
24.99 K. Schulze, F. Vargha-Khadem, M. Mishkin: Test of a motor theory of long-term auditory memory, Proc. Natl. Acad. Sci. USA **109**(18), 7121–7125 (2012)
24.100 G. Rizzolatti, L. Craighero: The mirror-neuron system, Annu. Rev. Neurosci. **27**, 169–192 (2004)
24.101 E. Kohler, C. Keysers, M.A. Umilta, L. Fogassi, V. Gallese, G. Rizzolatti: Hearing sounds, understanding actions: Action representation in mirror neurons, Science **297**(5582), 846–848 (2002)
24.102 C. Jacquemot, S.K. Scott: What is the relationship between phonological short-term memory and speech processing?, Trends Cogn. Sci. **10**(11), 480–486 (2006)
24.103 K. Schulze, S. Koelsch: Working memory for speech and music, Ann. N.Y. Acad. Sci. **1252**, 229–236 (2012), https://doi.org/10.1111/j.1749-6632.2012.06447.x
24.104 D. Saur, B.W. Kreher, S. Schnell, D. Kummerer, P. Kellmeyer, M.S. Vry, R. Umarova, M. Musso, V. Glauche, S. Abel, W. Huber, M. Rijntjes, J. Hennig, C. Weiller: Ventral and dorsal pathways for language, Proc. Natl. Acad. Sci. USA **105**(46), 18035–18040 (2008)
24.105 K. Schulze: How singing works, Front Psychol **3**, 51 (2012)
24.106 S.D.B. Dalla, M. Berkowska, J. Sowinski: Disorders of pitch production in tone deafness, Front. Psychol. **2**, 164 (2011)
24.107 G.A. Miller: The magical number seven plus or minus two: Some limits on our capacity for processing information, Psychol. Rev. **63**(2), 81–97 (1956)
24.108 K.A. Ericsson, W.G. Chase, S. Faloon: Acquisition of a memory skill, Science **208**(4448), 1181–1182 (1980)
24.109 F. Gobet: Chunking models of expertise: Implications for education, Appl. Cogn. Psychol. **19**(2), 183–204 (2005)
24.110 F. Gobet, P.C. Lane, S. Croker, P.C. Cheng, G. Jones, I. Oliver, J.M. Pine: Chunking mechanisms in human learning, Trends Cogn. Sci. **5**(6), 236–243 (2001)
24.111 D. Bor, N. Cumming, C.E. Scott, A.M. Owen: Prefrontal cortical involvement in verbal encoding strategies, Eur. J. Neurosci. **19**(12), 3365–3370 (2004)
24.112 C.R. Savage, T. Deckersbach, S. Heckers, A.D. Wagner, D.L. Schacter, N.M. Alpert, A.J. Fischman, S.L. Rauch: Prefrontal regions supporting spontaneous and directed application of verbal learning strategies: Evidence from PET, Brain **124**(Pt 1), 219–231 (2001)
24.113 D. Bor, J. Duncan, R.J. Wiseman, A.M. Owen: Encoding strategies dissociate prefrontal activity from working memory demand, Neuron **37**(2), 361–367 (2003)
24.114 D. Bor, A.M. Owen: A common prefrontal-parietal network for mnemonic and mathematical recoding strategies within working memory, Cereb. Cortex **17**(4), 778–786 (2007)
24.115 K. Schulze, K. Mueller, S. Koelsch: Neural correlates of strategy use during auditory working memory in musicians and non-musicians, Eur. J. Neurosci. **33**(1), 189–196 (2011)
24.116 S. Koelsch, E. Schroger, M. Tervaniemi: Superior pre-attentive auditory processing in musicians, Neuroreport **10**(6), 1309–1313 (1999)
24.117 C.L. Krumhansl, R.N. Shepard: Quantification of the hierarchy of tonal functions within a diatonic context, J. Exp. Psychol. Hum. Percept. Perform **5**(4), 579–594 (1979)
24.118 C.L. Krumhansl: The psychological representation of musical pitch in a tonal context, Cogn. Psychol. **11**, 346–374 (1979)
24.119 S. Koelsch, E. Schroger, T.C. Gunter: Music matters: Preattentive musicality of the human brain, Psychophysiology **39**(1), 38–48 (2002)
24.120 K. Schulze, W.J. Dowling, B. Tillmann: Working memory for tonal and atonal sequences during a forward and a backward recognition task, Music Percept **29**(3), 255–267 (2012)

25. Musical Syntax I: Theoretical Perspectives

Martin Rohrmeier, Marcus Pearce

The understanding of musical syntax is a topic of fundamental importance for systematic musicology and lies at the core intersection of music theory and analysis, music psychology, and computational modeling. This chapter discusses the notion of musical syntax and its potential foundations based on notions such as sequence grammaticality, expressive unboundedness, generative capacity, sequence compression and stability. Subsequently, it discusses problems concerning the choice of musical building blocks to be modeled as well as the underlying principles of sequential structure building. The remainder of the chapter reviews the main theoretical proposals that can be characterized under different mechanisms of structure building, in particular approaches using finite-context or finite-state models as well as tree-based models of context-free complexity (including the Generative Theory of Tonal Music) and beyond. The chapter concludes with a discussion of the main issues and questions driving current research and a preparation for the subsequent empirical chapter *Musical Syntax II*.

> *The idea that there is a grammar of music is probably as old as the idea of a grammar itself. Mark Steedman* [25.1, p. 1]

25.1 Outline

What distinguishes music from other sounds? One answer is to be found in the manner in which the elements are organized and related within a structural framework and, most importantly, the apprehension of this structure by a listener, so that the sound is *experienced as music* by that listener. Therefore, discovering the principles of musical structure building is one of the central questions for theoretical and empirical music research. Despite the strong historical (and methodological) divide between music-theoretical, computational, and psychological/neuroscientific approaches, questions about musical structure and the perception of it facilitate a close link across traditional divisions between disciplines [25.2]. Note that we use the term *computational* to describe a theory that is expressed in computational terms, whether or not it is actually implemented as a computer program.

Exploring the principles of musical structure building naturally requires us to distinguish between the goal of uncovering rules governing the structure of music (an

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_25

external goal) and the cognitive principles of the perception and production of these structures (an internal psychological goal). Yet both aspects form two sides of the same coin: the capacities and limitations of human perception and cognitive processes influence the possible structures that composers can use (for a similar argument about language, the reader is directed to [25.3]) and, together with other constraints (e.g., those imposed by cultural factors or the physical properties of instruments, constraints of the hands or the body, constraints of the performance and so on [25.4]) give rise to the musical structures that we find in music. In turn, musical structure is acquired implicitly by listeners from mere exposure and musical interaction and represented internally [25.5, 6] and, ultimately, reproduced in compositional practice (since composers are listeners before they become composers). However, there can be no learning without a hypothesis space and therefore theoretical models of musical structure, especially those grounded in computational modeling, provide a useful approach to understanding the hypothesis space that human learners are faced with when they acquire the syntactic structure of a musical style.

Finding a formal characterization of musical structure brings traditional music theory in close connection with computational modeling since the search for an optimal structural description (that relates to structures as *heard*) strongly implies *modeling* the internal structure of the music (whether it is a single composition, a part thereof, or a corpus). Since music is an inherently psychological phenomenon, we often use psychological understanding to guide the development of structural models of music, just as we use structural models of music to guide the development and testing of psychological theories of the perception of musical structure.

The disciplines involved in research on musical syntax range from musicology and music theory, through computational modeling, to psychology and neurobiology. Although the disciplinary perspectives are distinct (e.g., it is possible to develop a structural theory that is optimal according to some criterion, such as simplicity, but not according to the criterion of matching human perception and cognition), in this contribution, we focus on a converging picture that emerges when musical syntax is examined by triangulating between theory, computational modeling, and cognitive research. Here we focus on theoretical approaches to musical syntax while empirical research using computational models, psychological experimentation, and neuroimaging are covered in the companion chapter, *Musical Syntax II*.

25.2 Theories of Musical Syntax

25.2.1 The Concept of Musical Syntax

Berwick et al. [25.7, p. 89] give a brief account of syntax as:

> *the rules for arranging items (sounds, words, word parts, phrases) into their possible permissible combinations in a language.*

In human language the set of items (alphabet of symbols) may be words and morphosyntactic units, in birdsong they may be pitches, slides and other sounds. In music the symbols may be melodic notes, chords, voice-leading patterns or relationships between voices, timbral qualities and so on. Many music-theoretical approaches constitute informal, verbal accounts of syntactic models of music. Although the use of strict and well-defined formalisms is not (yet) common in music theory, there are some accounts that employ the notion of syntax in music theory. For instance, *Aldwell* and *Schachter* write the following in order to characterize *harmonic syntax* [25.8, p. 139]:

> *One way that music resembles language is that the order of things is crucial in both.* I went to the concert *is an English sentence, whereas* I concert went the to *is not. Similarly, I-VII6-I6-II6-V7-I* [...] *is a coherent progression of chords, whereas I-I6-VII6-II6-I-V7* [...] *is not, as you can hear if you play through the two examples. In the study of language, the word* syntax *is used to refer to the arrangement of words to form sentences; word order is a very important component of syntax. In studying music, we can use the term* harmonic syntax *to refer to the arrangement of chords to form progressions; the order of chords within these progressions is at least as important as the order of words in language. (Other components of harmonic syntax are the position of chords within phrases, the preparation and resolution of dissonances, and the relation of chord progressions to melody and bass lines.)*

A syntactic theory might be applied to any aspect of musical structure – melody, harmony, rhythm, metre, grouping structure, form, or even aspects such as timbre and dynamics. In practice, syntactic approaches have typically been applied to *what* happens in a musical sequence – e.g., predicting (combinations of) pitches and chords – rather than *when* it happens. Conversely,

theories of rhythm and metre often do not take an explicitly syntactic approach. By analogy, metrical and rhythmic features of language are often studied from the perspective of phonology rather than syntax. A well-formed harmonic sequence, for instance, may be assigned to metrical structure in a regular or irregular way. It is important to note that despite the predominance of Western music in theoretical and cognitive research [25.9], the general notion of musical syntax is *not* limited to Western tonal music – and there are approaches addressing non-Western music [25.10–12]. Different aspects of musical structure may be more or less important in different musical styles and cultures.

Several models for musical syntax have been proposed based on different levels of structural representation (melodic structure, harmony and chords, bass lines, outer voices, voice-leading, other types of categorized sound, polyphonic pitch structure and so on). Here, we reserve the term *syntax* for approaches presenting a formal system characterizing the sequential structure of such building blocks, in contrast to the more general term *musical structure* which captures the rich interaction of different musical features such as rhythm, metre, timbre, counterpoint, dynamics, phrasing, instrumentation, agogics, and so on. The precise identity of these building blocks is one of the central ongoing research questions in musical syntax.

The general term *musical structure* refers to the way in which one or more pieces of music may be represented in terms of their constituent parts, potentially reflecting a wide range of different musical features including rhythm, metre, timbre, counterpoint, dynamics, phrasing, instrumentation, agogics, and so on.

Musical syntax is a formal characterization of the principles governing permissible sequential structure in music. It characterizes sequences of musical events generated from a *lexicon of building blocks* and a set of *rules* governing how the building blocks are combined.

The *lexicon* (the set of building blocks) may consist of single events or patterns (schemata) of notes, glissandi, rests, chords, voice-leading patterns, timbres, or other noises. The *rules* may constitute any formal system that characterizes how sequences may (and may not) be formed by combining elements from the lexicon.

25.2.2 Foundations of Musical Syntax

Why do we need a syntax of music? When characterizing musical structure, and the cognitive representation and processing of that structure, several issues arise which motivate the development of a formal syntactical understanding of music. These include distinguishing regular and irregular musical structures (i. e., making grammaticality judgements), the fact that the space of possible musical compositions is theoretically infinite (or unbounded), the idea that we want to be able to describe structural relationships within musical sequences (i. e., focus on strong generative capacity compared to weak generative capacity). Syntax is also relevant to tasks such as compression, identifying the stability of events at different levels within music, and measuring musical similarity. We investigate these issues further in the following sections.

Grammaticality

One core foundation for the concept of musical syntax is the notion of regularity, permissibility, well-formedness or grammaticality, i. e., the characterization of structures that are regular or irregular with respect to a particular system (representing, for example, a musical style). If such a distinction were irrelevant, the characterization of musical syntax would be unnecessary since every sequence would be equally plausible. However, musical styles and idioms are implicitly characterized by regular and irregular sequences. Although categorical grammaticality decisions are often made, the distinction may be one of degree (compare [25.13–15], in linguistics). For instance, not every chord sequence or every musical form is regular in the 18th century Classical idiom [25.16]. Another example illustrates a regular and irregular common-practice harmonic sequence (Fig. 25.1 adapted from [25.17]). Musical regularity can be characterized empirically through (computational or hand-conducted) corpus analysis, which can provide information about frequent and less frequent regular patterns and indirectly about irregularity by the discovery of absent and low-probability patterns (although absence does not necessarily constitute evidence for ungrammaticality). Grammaticality can also be experimentally established through psychological experiments. Furthermore, introspective analyses by individual experts may be regarded as single-participant experiments, with some extrapolation to wider groups (whether expert or otherwise) assumed. In this context it is essential to understand the importance of negative evidence in the form of explicit instruction about implausible or irregular structures (note that this is different from the absence of positive evidence). While some regularities and rules may be inferred from positive data alone (i. e., the presence of well-formed structures), it is negative evidence that makes the strongest conclusions possible with regard to range, scope of generalization, and mutual interaction of grammatical systems. There is continuing debate about whether and how people might receive negative evidence in the development of language, but this topic has received little attention in the domain of music.

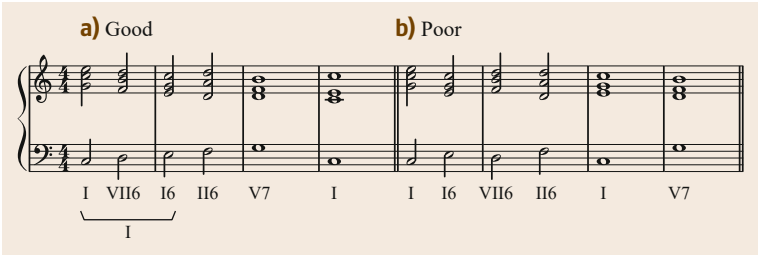


Fig. 25.1a,b The contrast between a good **(a)** and a poor **(b)** harmonic progression as discussed by *Aldwell* and *Schachter* [25.8, p. 140]. **(b)** is poor because the dependencies between the chords are disrupted – for instance, the II6 chord is not functionally well connected to its context (even though it features good voice-leading). Further note that in the analysis of the good example, the authors propose a hierarchical analysis of I VII6 I6 as a prolongation of a single I harmony

Unboundedness

The set of possible musical structures is unbounded – music, in Humboldt's famous words, makes infinite use of finite means. It is simple to demonstrate the unboundedness of musical structures: for every sequence we can imagine a longer one or a variation of it that inserts another element (tone, chord, etc.) into the sequence; we can further imagine a composition that never ends (such as ideal airport music). Hence, it is impossible to construct an exhaustive list of all musical sequences. Therefore, the only way to characterize musical structure is by employing a finite set of building blocks and recursive (or iterative) rules to generate grammatical sequences based on the recombination of building blocks using the rules. Generative grammars [25.18, 19] are *one kind* of formalism that embodies this principle, often used in theoretical approaches to the syntax of music (and other auditory sequences such as language or birdsong). Note that in this context the term *generative* does not refer to (human) music generation but to the description and analysis of a set of sequences by formal rules that are capable of generating them by a well-defined formal mechanism (such as a formal grammar).

Weak and Strong Generative Capacity

A syntactic model of a set of (musical) sequences may focus on the description of the surface sequences in order to reproduce exactly those sequences. For a given language (i. e., a set of strings), such a characterization may be accurate in terms of coverage (i. e., they can generate the set of strings). This is referred to as *weak generative capacity*. However, for most languages there is an infinite number of formal models that adequately describe the language, many of which are highly implausible. For this reason it is desirable that a syntactic theory matches theoretical insights as well as cognitively relevant (or adequate) structures, provides useful and testable generalizations, and achieves optimal compression (see below). This broader concept is known as *strong generative capacity*.

Compression

Characterizing a set of (musical) sequences using a generative theory allows us to capture a potentially infinite set of sequences using a finite set of rules. In this sense, we can think of generative theories in terms of the extent to which they enable compression through efficient representations of a set of sequences. Highly efficient, sparse encoding of the environment constitutes a core principle of cognitive systems [25.20, 21], and there is a close relationship between prediction and compression because we only need to store the information that is not predictable using a model [25.22]. Research in music information retrieval [25.23] and music psychology [25.24] has used general-purpose compression algorithms as models of musical complexity. A model that better captures structural regularities with general coverage in a given musical idiom is expected to be capable of more accurate prediction and, therefore, compression. Conversely, we can use compressibility (of unseen data to ensure generalizability) as a measure of the power and efficiency of a generative theory (and the latent structure that it postulates). However, a more complex model will itself consume more space, meaning that increased level of compression of the data must exceed the increased size of the model in order to be efficient. In this respect, approaches following the paradigms of minimum description length (MDL; [25.25]) and Bayesian model comparison provide closely related methods [25.22, Chap. 28] for comparing different candidate models taking into account differences in their complexity and the numbers of free parameters and so on [25.26, 27]. See *Mavromatis* [25.28] for an example of these principles applied to music.

Compression-based evaluation of a model of musical syntax is independent from other questions such as

grammaticality or weak/strong generative capacity. In particular, the criterion of optimal compression makes it possible to evaluate and compare syntactic models independently of the grammaticality distinction as well as independently of tests (such as pumping lemmata) that require grammaticality distinctions over sequences that are extremely improbable and do not generalize over corpora (such as *n*-th level center embeddings). In this context, compression provides a better way to provide a foundation for strong generative capacity and also assess the cognitive relevance of a proposed syntactic account of a (musical) language.

Stability, Similarity, and Semantics as Underpinnings of Syntax

There are several other ways to motivate syntactic structure in music. One of these is the proposal that we need an account of syntactic structure in music to be able to predict the relative stability of musical events. Many music theorists observe that in harmonic, melodic or voice-leading sequences, some events may be considered ornamental or accidental whereas others are structurally fundamental [25.8, 29, 30]. If this notion of relative structural stability – not to be confused with tonal stability and the tonal hierarchy [25.31] – is extended to a fully recursive structure (i. e., not just to individual notes or chords but also to motifs, phrases, and other larger scale components of musical form), it can be accounted for using a hierarchical syntactic formalism. Whether or not this type of structure is in turn coextensive with the above forms of establishing hierarchical structure remains a question open for further theoretical investigation.

Another, related avenue for establishing hierarchical structure is similarity. From a theoretical and psychological perspective musical similarity may be construed in terms of operations of omission or insertion of events with respect to a common core structure (for instance, differences between different cover versions of a song). In this context, it is important that such operations respect (hierarchical) structural boundaries of constituents (e.g., a tonic expansion) rather than comparisons between unstructured surface sequences. For example, *De Haas* et al. [25.32] implemented a similarity measure that is closely related to structural stability in terms of the largest common embeddable subtree between two compositions. This approach outperformed *edit distance* (a structure-free surface comparison between sequences) in predicting harmonic similarity between music sharing similar melodies. Similarity is also closely related to the concept of compression since we can train a syntactic model on one piece of music and use that model to predict another piece of music – greater degrees of predictability (and hence compressibility) indicate greater degrees of structural overlap between the pieces [25.21, 23].

Finally, semantics may constrain syntactic structures, particularly in linguistics. Whereas linguistic syntactic structures to a large extent serve the temporalization/linearization of semantic structure (in terms of form/meaning pairs), there is no immediate analogy in music. Although music may express meaning in terms of illocutionary acts like warnings, or aggression, or in terms of symbolic associations, it is agreed that music, in general, lacks complex, explicit propositional semantic forms ([25.33] and its discussion [25.33–36]). However, the patterns of relative stability outlined above (which are themselves related to syntactic structure) lead to perception and experience of tension and release by the listener, which can be viewed as a kind of semantic interpretation [25.37–40]. However, further research is required to examine these potential relationships between syntax and semantics in music.

25.3 Models of Musical Syntax

A model of musical syntax consists of two core components: first, a choice of the underlying representation for musical building blocks and how they relate to the musical surface; second, a formalism used to generate musical structure based on the set of building blocks.

25.3.1 Building Blocks

The choice of building block is fundamental for the syntactic model. In contrast to language, where the set of morphosyntactic features is largely accepted, syntactic models of music have made different choices of building block. This entails modeling musical structure at different levels of representation (or abstraction), such as: harmony and chord sequence, bass line, melodic line (diastematics), outer voices and voice-leading, or polyphonic pitch structure. Every choice involves selecting a distinction between structural and nonstructural items *with respect to the underlying model*. For instance, a model of harmonic syntax may regard different surface and melodic realizations of a chord sequence as equivalent; similarly, a theory of voice-leading would regard certain note repetitions, trills, or ornaments as nonstructural. Given the divergence of representations, styles, and level of abstraction adopted by different approaches in the literature, there is no consensus at

present how (and based on which principles) a *fundamental domain* of building blocks for a musical syntax could be established independently of the modeling goals.

We should mention here that some very interesting work has been done on representational spaces for various aspects of musical elements including, most notably, pitch spaces [25.37, 41–48] and metrical structure [25.49]. These theories define how these aspects of music may be expressed in algebraic ways and potentially represented by cognitive systems [25.50], but since they characterize the formal space of musical objects rather than explicitly specifying how sequences of elements may be combined, we do not consider them as theories of syntax proper.

25.3.2 Structure Building

Traditionally there have been a number of theoretical attempts to characterize the sequential structure of elements in a sequence, ranging from Markov models to context-free languages and corresponding probabilistic models. Many theories of structure have used explicit types of formal languages in the Chomsky hierarchy and its extensions [25.51]. Characterizations with models of different complexity involve a trade-off between the expressive (and compressive) power of the representation and corresponding processing requirements.

The languages generated by each class of grammar form proper subsets of the languages generated by classes of grammar higher up in the hierarchy. However, as we move up the hierarchy, the complexity of recognition and parsing increases in tandem with the increased expressive power of each class of grammar. In particular, while context-free grammars (and those higher in the hierarchy) are capable of capturing phenomena, such as embedded structure, which cannot be captured by finite-state grammars, they also bring with them many problems of intractability and undecidability, especially in the context of grammar induction [25.52].

It is fundamental to note that the Chomsky hierarchy and its extensions [25.51] constitute just *one way* of characterizing (musical) sequential structure. They are in no particular way primary or more natural than other approaches that characterize classes of infinite sets of strings, except in historical terms. There are many ways to characterize sequential structure, as any handbook of formal languages demonstrates (e.g., Handbook formal languages, [25.53]). Furthermore, computational models are, fundamentally, in no respect distinct from hand-crafted models by theorists in terms of their expressive power [25.5, 54, 55].

25.4 Syntactic Models of Different Complexity

25.4.1 Finite-Context Models

There is an interesting subclass of grammar contained within the class of finite-state grammars which are known as *finite context* grammars [25.56, 57]. In finite context automata, the next state is completely determined by testing a finite portion of length $n-1$ of the end of the already processed portion of the input sequence [25.57]. The core idea of these very local models is to characterize sequential structure by identifying possible element-to-element transitions (how elements may follow or precede each other). This characterization formally amounts to a table that lists grammatical relationships between each possible combination of elements (such as chord, note, or root transitions). This account is easily extended to larger context-lengths: the next element may be related not only to its predecessor, but also to the sequence of 2, 3, or more preceding elements. What such accounts have in common is the assumption that there are no (unbounded) dependencies between events longer than the relevant context of the model. In general, finite-context models correspond to the formal subcategory of strictly local languages (k-factor languages) and are also referred to as Markov or n-gram models.

A *k-factor language* is formally defined by a set of factors (strings of length k). A sequence is grammatical iff every subsequence of length k is part of the set of factors. Several models have been proposed in music theory and cognition that contain, in part, *k-factor* models [25.58–60]. It is important to note here that schema-theoretic approaches [25.61–63] do *not* naturally correspond with *k-factor languages* without modification (since they involve reductions, nonlocal patterns, and the ability to distinguish notes that are structurally important from those that are not).

These characterizations of structure, however, only draw a distinction between regular and irregular sequences, yet within those categories, they consider every possible sequence equally. For many theoretical purposes this is insufficient as some of these structures occur very frequently whereas other transitions are rare, less common, or unlikely. This theoretical requirement demands a characterization that is based not only on grammaticality but also on probability. It is straightforward to expand the above definition to incor-

porate probabilities: every entry in a transition matrix is associated with a probability. Probabilistic instantiations of the approach, therefore form a superset of the nonprobabilistic versions, which only allow the probabilities 0 (nongrammatical) and 1 (grammatical). Often these probabilities are estimated through analysis of frequency counts of events in a corpus [25.64, 65]. Such probabilistic extensions of k-factor languages are referred to as Markov models or n-gram models (for which n-grams correspond to probabilistic versions of k-factors). In an n-gram model, the sequence $e^{j}_{(j-n)+1}$ is called an n-gram (note that the subscript and superscript symbols denote the beginning and ending of a subsequence in the string; in the previous case it refers to the subsequence, from index $(j-n)+1$ to the index j) which consists, conceptually, of an initial subsequence, $e^{j-1}_{(j-n)+1}$, of length $n-1$ known as the *context* and a single symbol extension, e_j, called the *prediction*. The quantity $n-1$ is the *order* of the n-gram rewrite rule.

Such models are frequently used in computational models of music (see below), and also some music theoretical accounts (e.g., Piston's table of common root progressions, shown in Table 25.1).

By definition all types of strictly local or Markov models share the Markov assumption (25.1) and (25.2): the grammaticality (gr) of a subsequence or the probability (p) of a symbol appearing in a sequence depends only on its immediate preceding context of length k. This assumption means that these models cannot represent any nonconsecutive dependencies between musical elements beyond a fixed finite length.

$$gr\left(e^{i}_{1}\right) = gr\left(e^{i}_{i-n+1}\right) \tag{25.1}$$

$$p\left(e^{i}_{1}\right) \approx p\left(e^{i}_{i-n+1}\right) \tag{25.2}$$

Markov models provide powerful approximations to sequential structure for numerous practical applications independently of whether those sequences obey the Markov assumption. Nonetheless, such models are theoretically as well as practically limited in the extent to which they can capture and represent more complex structural features such as nonlocal dependencies, nested structures, and cross-serial dependencies. To some extent these limitations can be addressed by using sophisticated representation schemes such as the multiple-viewpoint formalism [25.64, 65] that extends the range of context that a Markov model can take into account by combining several Markov models over different feature spaces and (possibly) time scales, including nonadjacent events.

Table 25.1 Table of common root chord progressions (after [25.60])

	Is often followed by	Sometimes by	Less often by
I	IV or V	VI	II or III
II	V	IV or VI	I or III
III	VI	IV	I, II or V
IV	V	I or II	III or VI
V	I	VI or IV	III or II
VI	II or V	III or IV	I
VII	III	I	

25.4.2 Finite-State Models

Several theoretical approaches can be viewed as having equivalent representational power to *finite-state or regular grammars* in Chomsky's terminology. In contrast to *k-factor languages*, such models involve grammars that distinguish between (hidden) variables (nonterminal symbols) and surface symbols (terminals). Accordingly, *regular grammars* (i. e., grammars that only have rules of the form $A \rightarrow aB$; in which a refers to a terminal and A, B to nonterminals; see the appendix below) characterize sequential structure by building up a string from left to right. They form a true superset of *k-factor* languages. The formal machine that recognizes the set of strings generated by such a grammar is a *finite-state automaton* (informally, a flow-chart). The probabilistic counterpart to a regular grammar is the *Hidden Markov Model* (HMM; [25.66]).

25.4.3 Context-Free or Equivalent Models

There are several accounts of structure in music theory which go beyond the expressive power of finite-context and finite-state grammars (for further discussion [25.38]):

- Differences of structural importance
- Dependency structure, preparation, and ornamentation
- Headedness
- Nested structures
- Functional categories.

A useful starting point is the insight that the elements in a sequence may differ in structural importance, i. e., some can be left out without impairing grammaticality whereas others cannot. An early account by *Kostka* and *Payne* [25.67] refers to this as *levels* of harmony (note, however, that the observation is not restricted to harmony). Second, musical structure expresses dependencies: e.g., in a I II V or I III IV progression, the II or III chord may be understood as preparation for V or IV and not simply a sequential succession of I; accordingly, it is *dependent* on V or IV, not on I. This is expressed by the rules V $\rightarrow$ II V or *IV* $\rightarrow$ III IV

(for further details the reader is directed to [25.68]). This further entails that goals are structurally more fundamental than their preparations and, conversely, that ornamentation and variation adds new material to basic structure. This notion of dependency structure further entails a notion of *headedness*, namely that in the II V progression, V is the fundamental chord, i. e., the head (as expressed in the left-hand side of the rules above).

Another central formal concept concerns nested structures: the notion of dependency introduced above may lead to the formation of dependent subsubsequences within a dependent subsequence within a sequence (and so on). For instance, the II chord (which is the preparation of the V chord) in the above sequence may be further elaborated, ornamented, or prepared. This leads to recursive structures in the form of tail recursion (chains) and nested recursion (sequence in a sequence). One prominent example of nested structure in tonal music is modulation (e.g., an early account by [25.69]; the reader is also directed to [25.38, 68, 70]; for nested structure in music see [25.1, 38, 71, 72]). Finally, *Riemann* [25.73] introduced the notion that chords can be classified into different functional categories (such as tonic, dominants, and subdominants) that may be functionally interchangeable, such as II and IV leading to V. Riemann considered harmonic sequences to be hierarchical [25.74], and *Rohrmeier* [25.38, 68] developed a context-free formalism for the functional approach to harmony. In this formalism, tree structures represent different harmonic sequences that fulfill identical harmonic functions in the same way in higher parts of the tree.

Context-free languages and hierarchical tree-based accounts are well-suited for representing these kinds of structural dependencies in sequences. A number of theories account for music in such theoretical terms: *Schenker* [25.75], *Lerdahl* and *Jackendoff* [25.71]; *Keiler* [25.77, 78]; *Steedman* [25.1, 72]; *Narmour* [25.30]; *Lerdahl* [25.37]; *Tidhar* [25.79]; *Rohrmeier* [25.38, 68]. Various partial or full computational implementations of these theories exist, as discussed below.

Schenker [25.75] proposed a theoretical account of music based on reductional analysis that reveals different layers of musical structure ranging from surface to foreground, middle ground and Ursatz. Briefly construed, his account entails that principles of counterpoint (such as neighbor notes) may be used to distinguish the structural importance of musical events.

Lerdahl and *Jackendoff*'s Generative Theory of Tonal Music [25.71] (GTTM) provides an account that brings the ideas expressed by Schenker into a rule-based theoretical framework, inspired by the generative approach to grammar in linguistics. It is, for example, founded on the assumption that a piece of music can be partitioned into hierarchically organized segments which may be derived through the recursive application of the same rules at different levels of the hierarchy. Specifically, the theory is intended to yield a hierarchical, structural description of the final cognitive state of an experienced listener to that composition.

According to GTTM, a listener unconsciously infers four types of hierarchical structure in a musical surface (Fig. 25.2): first, *grouping structure* which corresponds to the segmentation of the musical surface into units (e.g., motives, phrases, and sections); second, *metrical structure* which corresponds to the pattern of periodically recurring strong and weak beats; third, *time-span reduction* which represents the relative structural importance of pitch events within contextually established rhythmic units; and finally, *prolongational reduction* reflecting patterns of tension and relaxation amongst pitch events at various levels of structure. According to the theory, grouping and metrical structure are largely derived directly from the musical surface and these structures are used in generating a time-span reduction which is, in turn, used in generating a prolongational reduction. Each of the four domains of organization is subject to *well-formedness rules* that specify which hierarchical structures are permissible and which themselves may be modified in limited ways by *transformational rules*. While these rules are abstract in that they define only formal possibilities, *preference rules* select which well-formed or transformed structures ac-

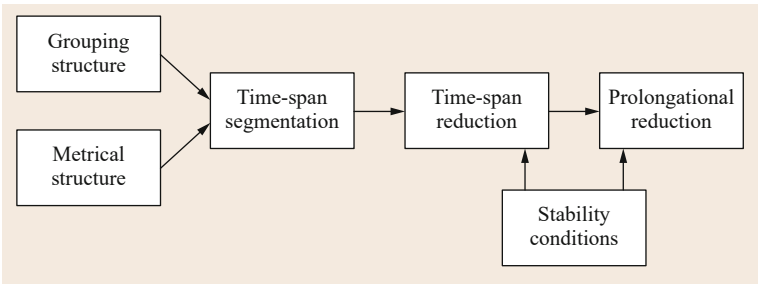


Fig. 25.2 The overall structure of GTTM (after [25.76, Fig. 10.6])

tually apply to particular aspects of the musical surface. Time-span and prolongational reduction additionally depend on tonal-harmonic *stability conditions* which are internal schemata induced from previously heard musical surfaces.

When individual preference rules reinforce one another, the analysis is stable and the passage is regarded as stereotypical whilst conflicting preference rules lead to an unstable analysis causing the passage to be perceived as ambiguous and vague. In this way, according to GTTM, the listener unconsciously attempts to arrive at the most stable overall structural description of the musical surface. Experimental studies of human listeners have found support for some of the preliminary components of the theory including the grouping structure [25.80] and the metrical structure [25.31].

The theory constitutes a formal predecessor to *Jackendoff*'s later parallel architecture framework of language [25.81, 82]. It is important to observe that the GTTM is not a *grammar* or a *syntax* of music: it provides a model of parsing but contains no generative rules or mechanisms to derive the musical surface, further it does not model a distinction between regular and irregular sequences. Rather than generating the musical surface, the GTTM is a theory of musical processing with only limited applicability as a theory of structural syntax per se.

It is highly challenging to develop formal context-free grammars that account for musical surface structure but several efforts have been made (e.g., [25.83, 84] for reviews). *Johnson-Laird* [25.85] used grammatical formalisms to investigate what has to be computed to produce acceptable rhythmic structure, chord progressions, and melodies in jazz improvisation. While a finite-state grammar (or equivalent procedure) can adequately compute the melodic contour, onset, and duration of the next note in a set of Charlie Parker improvisations, its pitch is determined by harmonic constraints derived from a context-free grammar modeling harmonic progressions. In a more recent approach, *Rohrmeier* [25.38, 68] introduces a set of context-free rules modeling the main features of tonal harmony from the common-practice period.

Context-free languages (and more complex formalisms) constitute supersets of regular and suprareglular languages. In fact, the latter constitute local boundaries of context-free languages (i. e., substrings that do not use the features of nested embedding are regular; it is further possible to derive precise models of local transitions from context-free models). Accordingly, the distinction between these types of languages does not imply a forced alternative – rather, context-free language models can result from the addition of the above structural features to regular language accounts. Put another way, we can add degrees of context-free character to regular grammars.

25.4.4 Beyond Context-Free Complexity

Are there aspects of musical structure that require greater than context-free power to be modeled? Debates in theoretical linguistics of the past 25 years have reached a fairly consensical view that human language is mildly context sensitive [25.86, 87]. It requires syntactic power that is stronger than context-free but considerably less strong than the immense computational power of full context-sensitive grammars. One example of this context-sensitive complexity is given by cross-serial dependencies (as in Dutch or Swiss German relative clauses [25.86, 88]) that cannot be expressed by context-free grammars. In the Chomskian tradition, minimalist grammars, that are equally mildly context-sensitive [25.89], adopted two mechanisms of external merge (similar to a context-free tree building operation) and internal merge (combining an already derived branch of a tree with different free positions in the tree). Internal merge may express features such as movement (*Sue wondered which book Peter read?*). *Katz* and *Pesetsky* [25.90] argue that musical and linguistic structure are formally equivalent in the sense that both require structure-building operations based on external and internal merge.

What about music? In his review, *Roads* [25.83] argues that the strict hierarchy characteristic of context-free grammars is difficult to reconcile with the ambiguity inherent in music. Faced with the need to consider multiple attributes occurring in multiple overlapping contexts at multiple hierarchical levels, even adding ambiguity to a grammar is unlikely to yield a satisfactory representation of musical context. The use of context-sensitive grammars can address these problems to some extent but this also brings considerable additional difficulties in terms of efficiency and complexity. There are several attempts to model music using grammatical formalisms which add some degree of context sensitivity to context-free grammars without adding significantly to the complexity of the rewrite rules. An example is the Augmented Transition Network (ATN) which extends a recursive transition network (formally equivalent to a context-free grammar) by associating state transition arcs (rewrite rules) with procedures which perform the necessary contextual tests. *Cope* [25.91] describes the use of ATNs to rearrange harmonic, melodic, and rhythmic structures in EMI (experiments in musical intelligence). Another example is provided by the pattern grammars developed by *Kippen* and *Bel* [25.10] for modeling improvisation in North Indian tabla drumming.

Steedman [25.1, 72] has developed a categorial grammar (augmented context-free) to account for the harmonic structure of 12-bar blues, based on a theory of tonal harmony due to *Longuet-Higgins* [25.45, 46]. Although it is less ambitious than that of *Johnson-Laird* [25.85], this allows a more elegant description of improvisational competence since it does not rely on substitution into a previously prepared skeleton. However, in using the grammar to generate structural descriptions of blues chord progressions, Steedman had to introduce implicit meta-level conventions not explicit in the production rules of the grammar. The extent of context-sensitivity required to adequately model musical structure requires further investigation.

25.5 Discussion

This discussion of theoretical accounts of musical syntax raises several issues and questions which are driving current research:

1. How powerful a grammar do we need to represent the relationships present in musical structure? Are there examples of syntactic musical structures that require (mild) context sensitivity? How can multiple, polyphonic streams be represented by formal approaches?
2. How does musical syntax interact with other aspects of musical structure such as rhythm, metre, and timing? Or are these aspects also best explained using syntactic formalisms?
3. To what extent does real music, and listeners' perception of music, exhibit features of recursion, nonlocal dependencies, single or multiple center-embedding?
4. Which kinds of formal structures are listeners (musicians or nonmusicians) sensitive to?
5. Can such syntactic structures and relationships be learned, and if so, how and which kinds of predispositions need to be assumed as innate?

The power of a particular syntactic formalism is independent of whether it is probabilistic or deterministic. Probabilistic models have distinct advantages in terms of the subtlety with which they can capture structural dependencies for application in prediction, classification, parsing, and learnability/inference as well as in terms of robustness and graded grammaticality. For each of the model classes of the extended Chomsky hierarchy probabilistic counterparts have been proposed (e.g., finite-context grammars: *n*-gram models; regular grammars: Hidden Markov models; context-free grammars: probabilistic context-free grammars). These developments suggest as a general strategy that it may be beneficial to move from deterministic to probabilistic models for implementation and evaluation. It is important to note here that the Chomsky hierarchy is just one way of characterizing the power of grammatical formalisms but it does not necessarily lend itself naturally to every aspect of musical structure. Furthermore, as we noted above degrees of context-free character may be added to regular grammars and degrees of context-sensitivity added to context-free grammars.

While Markov and *n*-gram models are easily learned and are useful for prediction, they are barely capable of modeling more complex structures, nonlocal and hierarchical dependencies in music described above that are essential in musical implicative structure, stability, tension, and similarity. Conversely, more powerful types of syntactic formalisms are considerably more difficult to infer from data. It is not currently clear that we can develop one overarching theoretical stance that generalizes across musical styles and cultures. As in other areas of empirical musicology, the majority of research on musical syntax has focused on Western music and harmony in particular (with a few notable exceptions including [25.10, 12] and recent work by *Mavromatis* [25.92]). Different musical styles or traditions may emphasize different kinds of building block or show different degrees and kinds of complexity in their syntactic structure. Cross-cultural comparisons may have implications for evolutionary theories of music. While each process of inference requires predetermined (innate) assumptions about at least the search space and the structure of the model, it must be noted that cross-cultural universality by no means implies innateness of more than these assumptions.

Many of these questions are best addressed by implementing a computational theory as a computer model that embodies a particular theoretical stance on musical syntax and testing the model by comparing its behavior with human behavior. Modeling requires the theory to make all its assumptions explicit and permits the analysis of complex examples and large corpora. It is also possible to conduct a quantitative comparison of the behavior of a computational model with the behavior of listeners, allowing a rigorous empirical test of the theory as a psychologically plausible model of cognitive representation and processing of musical syntax. We address these points in detail in the following chapter, *Musical Syntax II*.

25.A Appendix: The Chomsky Hierarchy

Noam Chomsky introduced a containment hierarchy of four classes of formal grammar in terms of increasing restrictions placed on the form of valid rewrite rules [25.52]. A formal grammar consists of a set of nonterminal symbols (variables), terminal symbols (elements of the surface), production rules, and a starting symbol to derive productions. In the following description, $a \in T^*$ denotes a (possibly empty) sequence of terminal symbols, $A, B \in V$ denote nonterminal symbols, $\alpha \in (V \cup T)^+$ denotes a nonempty sequence of terminal and nonterminal symbols, and $\beta, \beta' \in (V \cup T)^*$ denote (possibly empty) sequences of terminal and nonterminal symbols. The difference between different formal grammars in the Chomsky hierarchy relates to different possible production rules.

Every grammar in the Chomsky hierarchy corresponds with an associated automaton: while formal grammars generate the string language, formal automata specify constraints on processing or generating mechanisms that characterize the formal language. Automata provide an equivalent characterization of formal languages as formal grammars.

25.A.1 Type 3 (Regular)

Grammars in this class feature restricted rules allowing only a single terminal symbol, optionally accompanied by a single nonterminal, on the right-hand side of their productions

$$A \to a$$
$$A \to aB \quad \text{(right linear grammar) or}$$
$$A \to Ba \quad \text{(left linear grammar)} .$$

Regular grammars generate all languages which can be recognized by a finite-state automaton, which requires no memory other than a representation of its current state.

It is essential to note that regular grammars are *not* equivalent to Markov models or k-factor languages (see Sect. 25.4.1 above).

25.A.2 Type 2 (Context Free)

Grammars in this class only restrict the left-hand side of their rewrite rules to a single nonterminal symbol – i. e., the right-hand side can be any string of nonterminal symbols

$$A \to \alpha .$$

The equivalent automata characterization of a context-free language is a nondeterministic pushdown automaton, which is an extension of finite-state automata that employs memory using a stack and state transitions may depend on the current state as well as the top symbol in the stack.

25.A.3 Type 1 (Context Sensitive)

Grammars in this class are restricted only in that there must be at least one nonterminal symbol on the left-hand side of the rewrite rule and the right-hand side must contain at least as many symbols as the left-hand side, i. e., string length increases monotonically in the production sequence.

$$\beta A \beta' \to \alpha, \quad \left|\beta A \beta'\right| \leq \alpha$$

Context-sensitive languages are equivalently characterized by a linear bounded automaton, that is a state-machine extended by a linear bounded random access memory band, whose transitions depend on the state, the symbol on the memory band.

25.A.4 Type 0 (Unrestricted)

Grammars in this class place no restrictions on their rewrite rules

$$\alpha \to \beta$$

and generate all languages which can be equivalently characterized by a universal Turing machine (the recursively enumerable languages), which is the same as the linear bounded automaton for context-sensitive languages without bounds on the memory tape.

References

25.1 M. Steedman: The blues and the abstract truth: Music and mental models. In: *Mental Models in Cognitive Science*, ed. by A. Garnham, J. Oakhill (Erlbaum, Mahwah 1996) pp. 305–318

25.2 M.T. Pearce, M. Rohrmeier: Music cognition and the cognitive sciences, Top. Cogn. Sci. **4**(4), 468–484 (2012)

25.3 M. Christiansen, N. Chater: Toward a connectionist model of recursion in human linguistic performance, Cogn. Sci. **23**, 157–205 (1999)

25.4 D. Sudnow: *Ways of the Hand: The Organization of Improvised Conduct* (MIT Press, Cambridge 1978)

25.5 M.T. Pearce, G.A. Wiggins: Auditory expectation: The information dynamics of music perception and cognition, Top. Cogn. Sci. **4**(4), 625–652 (2012), https://doi.org/10.1111/j.1756-8765.2012.01214.x

25.6 M. Rohrmeier, P. Rebuschat: Implicit learning and acquisition of music, Top. Cogn. Sci. **4**(4), 525–553 (2012), https://doi.org/10.1111/j.1756-8765.2012.01223.x

25.7 R.C. Berwick, A.D. Friederici, N. Chomsky, J.J. Bolhuis: Evolution, brain, and the nature of language, Trends Cogn. Sci. **17**(2), 91–100 (2013), https://doi.org/10.1016/j.tics.2012.12.002

25.8 E. Aldwell, C. Schachter: *Harmony & Voice Leading* (Thomson Schirmer, New York 2003)

25.9 I. Cross: Cognitive science and the cultural nature of music, Top. Cogn. Sci. **4**(4), 668–677 (2012)

25.10 J. Kippen, B. Bel: Modelling music with grammars. In: *Computer Representations and Models in Music*, ed. by A. Marsden, A. Pople (Academic Press, London 1992) pp. 207–238

25.11 S. Marcus: The Eastern Arab system of melodic modes: A case study of Maqam Bayyati. In: *The Garland Encyclopedia of World Music. The Middle East* (Routledge, New York 2003) pp. 33–44

25.12 D.R. Widdess: Aspects of form in North Indian ālāp and dhrupad. In: *Music and Tradition: Essays on Asian and Other Musics Presented to Laurence Picken* (Cambridge Univ. Press, Cambridge 1981) pp. 143–182

25.13 A. Sorace, F. Keller: Gradience in linguistic data, Lingua **115**(11), 1497–1524 (2005)

25.14 B. Aarts: *Fuzzy Grammar: A Reader* (Oxford Univ. Press, Oxford 2004)

25.15 B. Aarts: *Syntactic Gradience: The Nature of Grammatical Indeterminacy* (Oxford Univ. Press, Oxford 2007)

25.16 W. Caplin: *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven* (Oxford Univ. Press, New York, Oxford 1998)

25.17 E. Aldwell, C. Schachter: *Harmony and Voice Leading*, 2nd edn. (Harcourt Brace Jovanovich, San Diego 1989)

25.18 N. Chomsky: *Syntactic Structures* (Mouton, The Hague 1957)

25.19 N. Chomsky: *Aspects of the Theory of Syntax* (MIT Press, Cambridge 1965)

25.20 N. Chater: Reconciling simplicity and likelihood principles in perceptual organisation, Psychol. Res. **103**, 566–581 (1996)

25.21 N. Chater, P. Vitanyi: The generalized universal law of generalization, J. Math. Psychol. **47**, 346–369 (2003)

25.22 D.J.C. MacKay: *Information Theory, Inference and Learning Algorithms* (Cambridge Univ. Press, Cambridge 2003)

25.23 R. Cilibrasi, P.M.B. Vitanyi, R. de Wolf: Algorithmic clustering of music based on string compression, Comput. Music J. **28**, 49–67 (2004)

25.24 M.M. Marin, H. Leder: Examining complexity across domains: Relating subjective and objective measures of affective environmental scenes, paintings and music, PLoS ONE **8**(8), e72412 (2013)

25.25 P.D. Grünwald: *The Minimum Description Length Principle* (MIT Press, Cambridge 2007)

25.26 A. Perfors, J.B. Tenenbaum, T. Regier: The learnability of abstract syntactic principles, Cognition **118**(3), 306–338 (2011)

25.27 C. Kemp, J.B. Tenenbaum: The discovery of structural form, Proc. Natl. Acad. Sci. **105**(31), 10687–10692 (2008)

25.28 P. Mavromatis: Minimum description length modelling of musical structure, J. Math. Music **3**(3), 117–136 (2009)

25.29 S. Kostka, D. Payne: *Tonal Harmony* (Alfred A. Knopf, New York 1984)

25.30 E. Narmour: *The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model* (University of Chicago Press, Chicago, London 1992)

25.31 C.L. Krumhansl: *Cognitive Foundations of Musical Pitch* (Oxford Univ. Press, Oxford 1990)

25.32 B. De Haas, M. Rohrmeier, R. Veltkamp, F. Wiering: Modeling harmonic similarity using a generative grammar of tonal harmony. In: *Proc. 10th Int. Soc. Music Inf. Retr. Conf. (ISMIR 2009), Kobe*, ed. by K. Hirata, G. Tzanetakis, K. Yoshii (2009) pp. 549–554

25.33 S. Koelsch: Towards a neural basis of processing musical semantics, Phys. Life Rev. **8**(2), 89–105 (2011)

25.34 L.R. Slevc, A.D. Patel: Meaning in music and language: Three key differences: Comment on "Towards a neural basis of processing musical semantics" by Stefan Koelsch, Phys. Life Rev. **8**(2), 110–111 (2011)

25.35 U. Reich: The meanings of semantics: Comment on 'Towards a neural basis of processing musical semantics' by Stefan Koelsch, Phys. Life Rev. **8**(2), 120–121 (2011)

25.36 W.T. Fitch, B. Gingras: Multiple varieties of musical meaning: Comment on "Towards a neural basis of processing musical semantics" by Stefan Koelsch, Phys. Life Rev. **8**(2), 108–109 (2011)

25.37 F. Lerdahl: *Tonal Pitch Space* (Oxford Univ. Press, New York 2001)

25.38 M. Rohrmeier: Towards a generative syntax of tonal harmony, J. Math. Music **5**(1), 35–53 (2011)

25.39 M. Lehne, M. Rohrmeier, S. Koelsch: Tension-related activity in the orbitofrontal cortex and amygdala: An fMRI study with music, Soc. Cogn. Affect. Neurosci. **9**(10), 1515–1523 (2013)

25.40 M. Rohrmeier, W. Zuidema, G.A. Wiggins, C. Scharff: Principles of structure building in music, language and animal song, Phil. Trans. R. Soc. B (2015), https://doi.org/10.1098/rstb.2014.0097

25.41 G.J. Balzano: The pitch set as a level of description for studying musical pitch perception. In: *Music, Mind and Brain*, ed. by M. Clynes (Plenum, New York 1982) pp. 321–351

25.42 L. Euler: *Tentamen Novae Theoriae Musicae* (Academia Scientiae, St. Petersburg 1739), reprint: Broude Bros., New York 1968

25.43 G. Weber: *Versuch einer geordneten Theorie der Tonsetzkunst*, Vol. 1–4 (Schott, Mainz 1830)

25.44 J. Pressing: Cognitive isomorphisms between pitch and rhythm in world musics: West Africa, the Balkans and Western tonality, Stud. Music **17**, 38–61 (1983)

25.45 H.C. Longuet-Higgins: Letter to a musical friend, Music Rev. **23**, 244–248 (1962)

25.46 H.C. Longuet-Higgins: Second letter to a musical friend, Music Rev. **23**, 271–280 (1962)

25.47 R.N. Shepard: Structural representations of musical pitch. In: *Psychology of Music*, ed. by D. Deutsch (Academic Press, New York 1982) pp. 343–390

25.48 D. Tymoczko: *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice* (Oxford Univ. Press, Oxford 2011)

25.49 J. London: *Hearing in Time* (Oxford Univ. Press, Oxford 2004)

25.50 P. Janata, J.L. Birk, J.D. van Horn, M. Leman, B. Tillmann, J.J. Bharucha: The cortical topography of tonal structures underlying Western music, Science **298**(5601), 2167–2170 (2002)

25.51 G. Jäger, J. Rogers: Formal language theory: refining the Chomsky hierarchy, Philos. Trans. R. Soc. B **367**(1598), 1956–1970 (2012)

25.52 J.E. Hopcroft, J.D. Ullman: *Introduction to Automata Theory, Languages and Computation* (Addison-Wesley, Reading 1979)

25.53 G. Rozenberg, A. Salomaa (Eds.): *Handbook of Formal Languages* (Springer, New York 1997)

25.54 G. Wiggins: Computer models of (music) cognition. In: *Language and Music as Cognitive Systems*, ed. by P. Rebuschat, M. Rohrmeier, I. Cross, J. Hawkins (Oxford Univ. Press, Oxford 2012) pp. 169–188

25.55 M.A. Rohrmeier, S. Koelsch: Predictive information processing in music cogntion. A critical review, Int. J. Psychophysiol. **83**(2), 164–175 (2012)

25.56 T.C. Bell, J.G. Cleary, I.H. Witten: *Text Compression* (Prentice Hall, Englewood Cliffs 1990)

25.57 S. Bunton: *On-Line Stochastic Processes in Data Compression*, Doctoral Dissertation (University of Washington, Seattle 1996)

25.58 D. Huron: *Sweet Anticipation: Music and the Psychology of Expectation* (MIT Press, Cambridge 2006)

25.59 J.-P. Rameau: *Traite de l'harmonie reduite a ses principes naturels* (J.B.C. Ballard, Paris 1722)

25.60 W. Piston: *Harmony* (W.W.Norton, New York 1948)

25.61 R.O. Gjerdingen: Learning syntactically significant temporal patterns of chords, Neural Netw. **5**, 551–564 (1992)

25.62 V. Byros: Meyer's anvil: Revisiting the schema concept, Music Anal. **31**(3), 273–346 (2012)

25.63 V. Byros: Towards an "archaeology" of hearing: schemata and eighteenth-century consciousness, Musica Humana **1**(2), 235–306 (2009)

25.64 D. Conklin, I.H. Witten: Multiple viewpoint systems for music prediction, J. New Music Res. **24**(1), 51–73 (1995)

25.65 M.T. Pearce: *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition*, Doctoral Dissertation (Department of Computing, City University, London 2005)

25.66 L.R. Rabiner: A tutorial on Hidden Markov Models and selected applications in speech recognition, Proc. IEEE **77**(2), 257–285 (1989)

25.67 S. Kostka, D. Payne: *Tonal Harmony* (McGraw-Hill, New York 1995)

25.68 M. Rohrmeier: A generative grammar approach to diatonic harmonic structure. In: *Proc. 4th Sound Music Comput. Conf.*, ed. by Spyridis, Georgaki, Kouroupetroglou, Anagnostopoulou (2007) pp. 97–100

25.69 D.R. Hofstadter: *Goedel, Escher, Bach* (Basic Books, New York 1979)

25.70 I. Giblin: *Music and the Generative Enterprise: Situating a Generative Theory of Tonal Music in the Cognitive Sciences*, Doctoral Dissertation (University of New South Wales, Sydney 2008)

25.71 F. Lerdahl, R. Jackendoff: *A Generative Theory of Tonal Music* (MIT Press, Cambridge 1983)

25.72 M. Steedman: A generative grammar for jazz chord sequences, Music Percept **2**(1), 52–77 (1984)

25.73 H. Riemann: *Musikalische Syntaxis* (Breitkopf Härtel, Leipzig 1877)

25.74 T. Christensen: The Schichtenlehre of Hugo Riemann, Theory Only **6**(4), 37–44 (1982)

25.75 H. Schenker: *Der Freie Satz. Neue musikalische Theorien und Phantasien* (Margada, Liège 1935)

25.76 F. Lerdahl: Cognitive constraints on compositional systems. In: *Generative Processes in Music: The Psychology of Performance, Improvisation and Composition*, ed. by J.A. Sloboda (Clarendon, Oxford 1988) pp. 231–259

25.77 A. Keiler: Bernstein's "The Unanswered Question" and the problem of musical competence, Music. Q. **64**(2), 195–222 (1978)

25.78 A. Keiler: Two views of musical semiotics. In: *The Sign in Music and Literature*, ed. by W. Steiner (Univ. Texas Press, Austin 1981) pp. 138–168

25.79 D. Tidhar: *A Hierarchical and Deterministic Approach to Music Grammars and its Application to Unmeasured Preludes* (dissertation.de, Berlin 2005)

25.80 I. Deliège: Grouping conditions in listening to music: An approach to Lerdahl and Jackendoff's grouping preference rules, Music Percept **4**(4), 325–360 (1987)

25.81 R. Jackendoff: *Foundations of Language – Brain, Meaning, Grammar, Evolution* (Oxford Univ. Press, Oxford 2003)

25.82 R. Jackendoff: A parallel architecture perspective on language processing, Brain Res **1146**, 2–22 (2007)

25.83 C. Roads: Grammars as representations for music. In: *Foundations of Computer Music*, ed. by C. Roads, J. Strawn (MIT Press, Cambridge 1985) pp. 403–442

25.84 J. Sundberg, B. Lindblom: Generative theories for describing musical structure. In: *Representing Musical Structure*, ed. by P. Howell, R. West, I. Cross (Academic Press, London 1991) pp. 245–272

25.85 P.N. Johnson-Laird: Jazz improvisation: A theory at the computational level. In: *Representing Musical Structure*, ed. by P. Howell, R. West, I. Cross (Academic Press, London 1991) pp. 291–325

25.86 S.M. Shieber: Evidence against the context-freeness of natural language. In: *The Formal Complexity of Natural Language*, ed. by W.J. Savitch, E. Bach, W. Marsh, G. Safran-Navah (Springer Netherlands, Dordrecht 1987) pp. 320–334
25.87 A.K. Joshi, K.V. Shanker, D. Weir: *The Convergence of Mildly Context-Sensitive Grammar Formalisms*. Technical Report No. MS- CIS-90-01 (Univ. of Pennsylvania, Department of Computer and Information Science 1990)
25.88 M. Steedman: *The Syntactic Process* (MIT Press, Cambridge 2001)
25.89 E.P. Stabler: Computational perspectives on minimalism. In: *Oxford Handbook of Linguistic Minimalism*, ed. by C. Boeckx (Oxford Univ. Press, Oxford 2011) pp. 617–643
25.90 J. Katz, D. Pesetsky: The Identity Thesis for Language and Music. http://ling.auf.net/lingbuzz/000959 (2010)
25.91 D. Cope: Computer modelling of musical intelligence in EMI, Comput. Music J. **16**(2), 69–83 (1992)
25.92 P. Mavromatis: A hidden Markov model of melody production in Greek church chant, Comput. Musicol. **14**, 93–112 (2005)

26. Musical Syntax II: Empirical Perspectives

Marcus Pearce, Martin Rohrmeier

Efforts to develop a formal characterization of musical structure are often framed in syntactic terms, sometimes but not always with direct inspiration from research on language. In Chap. 25, we present syntactic approaches to characterizing musical structure and survey a range of theoretical issues involved in developing formal syntactic theories of sequential structure in music. Such theories are often computational in nature, lending themselves to implementation and our first goal here is to review empirical research on computational modeling of musical structure from a syntactic point of view. We ask about the motivations for implementing a model and assess the range of approaches that have been taken to date. It is important to note that while a computational model may be capable of deriving an optimal structural description of a piece of music, human cognitive processing may not achieve this optimal performance, or may even process syntax in a different way. Therefore we emphasize the difference between developing an optimal model of syntactic processing and developing a model that simulates human syntactic processing. Furthermore, we argue that, while optimal models (e.g., optimal compression or prediction) can be useful as a benchmark or yardstick for assessing human performance, if we wish to understand human cognition then simulating human performance (including aspects that are nonoptimal or even erroneous) should be the priority. Following this principle, we survey research on processing of musical syntax from the perspective of computational modeling, experimental psychology and cognitive neuroscience. There exists a large number of computational models of musical syntax, but we limit ourselves to those that are explicitly cognitively motivated, assessing them in the context of theoretical, psychological and neuroscientific research.

26.1 Computational Research

26.1.1 Foundations

Different approaches to building computer models of musical structure can be characterized, and distinguished, in terms of how expressive they are in terms of the degree of structural complexity they are capable of representing. Therefore, there is a direct link between the theoretical characterization of musical syntax (discussed in Chap. 25) and the implementation and testing of these theories as computational models of cognition, discussed here. Implementing a theory of musical syntax and processing has several potential advantages,

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_26

following well-known examples in cognitive science:

1. Implementing a theory as a computer program (which must run, generating output from the data and parameters supplied as input) ensures that it takes as little as possible for granted and any assumptions are explicitly stated [26.1–3].
2. Experiments can be run to evaluate the implemented theory by comparison of its behavior with the output expected from theoretical accounts or directly with human behavior given the same inputs [26.3, 4].

It is important to distinguish between models whose knowledge is provided to them by a human expert (in the style of good, old-fashioned artificial intelligence, GOFAI) and those that acquire knowledge about musical structure by learning (either supervised or unsupervised) from experience of music, given some predefined structural representation, the parameters of which are learned (in the tradition of machine learning), be it a neural network, a Markov model or grammatical inference. The appropriate approach depends on the goal of the modeling. However, in many cases the two approaches are complementary in that successfully learning complex representations (e.g., context-free rules) is extremely challenging but the alternative approach can result in models that do not generalize far beyond the musical domain for which they were developed (e.g., *Steedman*'s [26.5, 6] grammars for blues harmony). In the past, tasks that required context-free representations have usually been hand coded while tasks that require simpler representational relationships have been able to benefit from the flexibility of machine learning. However, methods have now been developed in computational linguistics to learn certain kinds of context-free representation [26.7]. Note that if we are interested in cognitive representations of music, there is the additional issue of the extent to which the representations in question are actually learned or inherited (i. e., innately specified) by human beings ([26.8] as, e.g., discussed in the poverty of the stimulus debate [26.9]).

As mentioned above, we must also emphasize the distinction between finding an optimal structural description of a piece of music and modeling the cognitive representation of that piece in the mind of a listener. The former bounds the latter but listeners are likely to be subject to constraints of perception and cognition (e.g., limitations of working memory load), which would prevent them reaching an optimal structural description. Note also that it is problematic to assume the existence of an *average listener* without understanding all the factors (e.g., musical training, environmental context, degree of attention etc.) that could influence the structural descriptions that listeners form. Nonetheless, it is often useful to identify theoretical bounds on structural complexity using an optimal model. Furthermore, theoretical models of musical structure can help us understand the hypothesis space that human learners are faced with when they acquire the syntactic structure of a musical style. In artificial intelligence research, we distinguish between the representation defining the search space and algorithms for traversing the space. Similarly in machine learning, we distinguish between the hypothesis space and learning mechanisms for traversing the space. In the case of musical syntax, the hypotheses correspond to potential stylistic grammars generating structural descriptions of music and we can think of the learning as traversing the space of possible grammars, specifying the parameters distinguishing those grammars along the way.

The following sections illustrate these points, using different kinds of computational models that have been proposed for understanding musical syntax.

26.1.2 Early Approaches: Pattern Processing

We begin with a review of two early approaches that are of historical importance because they laid the foundations for symbolic models that subsequently became influential. One early approach was based on the assumption that listeners use pattern induction processes to develop predictions for successive events in melodies [26.10, 11]. These models attempt to define formal languages for describing the patterns perceived by humans in temporal sequences (such as music) and use them to explain how these patterns are applied for prediction. *Simon* and *Sumner* [26.11], for example, begin with ordered alphabets for representing the range of possible values for a particular musical dimension (e.g., note names, note durations). Simon and Sumner restrict their attention to the dimensions of melody, harmony, rhythm and form and use alphabets for diatonic notes, triads, duration, stress and formal structure. The operations they consider are *same* (when the subsequent symbol is identical to the previous one) and *NEXT* (the subsequent symbol is obtained by taking the next symbol in the specified alphabet a specified number of times). Sequences of symbols may then be described more compactly as a sequence of these operations.

Deutsch and *Feroe* [26.10] extended the model of Simon and Sumner in several ways, in particular defining structures as sequences of elementary operators and sequence operators such as prime, retrograde, inversion and alternation. They apply their pattern language to various alphabets, corresponding to collections of pitches, such as the major and (natural, harmonic and melodic) minor scales, major, minor and diminished tri-

ads, and seventh chords. They argue that the pattern language facilitates processing in four ways:

1. Reduced redundancy of representation
2. Distinct alphabets may be invoked at different levels
3. Embedded sequence structures and their associated alphabets may be encoded as chunks
4. The chunking of structures allows for the representation of configurations that satisfy proximity and the differentiation of different members of the alphabet in terms of frequency.

Deutsch and Feroe propose that multiple representations may be formed by listeners who, according to the model, will tend to choose the most parsimonious. The acquisition of a representation is an ongoing process of generation and testing of multiple hypothesized structural representations [26.12, 13].

26.1.3 Markov Modeling

Another early approach to modeling musical structure was based on statistical learning and information theory, in particular using *information content* and *entropy* to measure the complexity of a musical style. It is interesting to note that information theory was applied to music as early as 1955 [26.14–17] just a few years after *Shannon*'s foundational work was published [26.18]. Typically, this approach involves representing a musical work as a sequence of symbols drawn from an alphabet (e.g., a melody might be represented as a sequence of pitch symbols, harmonic movement as a sequence of chord symbols). The learning task is to estimate a conditional probability governing the next symbol in the musical sequence, given the preceding symbols. Such models (in various guises) have been highly influential in terms of understanding predictive processing in human cognitive processing of music [26.19, 20].

It is possible to vary the length of the context used in estimating the probability, known as the *order* of the model (a zeroth-order model has no context, a first-order model a context of one symbol and so on). Usually the probabilities (i. e., the parameters of the model) are estimated through statistical analysis of a corpus of musical works. These models are also known as Markov models because the probability of an event is only dependent on the immediately previous context (the Markov property), or, in other words, the model does not take into account any nonlocal dependencies. Once a probability distribution has been generated in a particular context, the *entropy* of that distribution reflects the model's uncertainty about the following musical event before it arrives while the *information content* (the negative log probability) reflects how unexpected the next note is, once it has arrived – given a local model [26.21]. Note that the entropy and information content of a musical event or sequence are not properties of the music per se, but properties of the music from the perspective of an underlying model.

It is interesting that one of the first nonmilitary applications of early computers was software to generate music incorporating grammatical representations of musical styles corresponding to probabilistic Markov models [26.22]. Early work using these models tended to focus on fixed, low-order models with simple representational building blocks (e.g., chromatic pitch) to estimate and compare the average entropy of different musical works and corpora [26.16, 23–26] rather than dynamic prediction of ongoing sequential musical structure.

More recent research addressed these limitations by using variable-order Markov models, where the order varies depending on the context to generate accurate predictions dynamically throughout pieces of music [26.27, 28]. Prediction performance may also be improved by combining predictions from a long-term model (trained on a large corpus of music in the style) and a short-term model (which starts with an empty model and learns incrementally from the current musical work) [26.27, 28]. The long-term model represents the effects of long-term schematic exposure to a musical style while the short-term model reflects more local learning of repeated structure within a musical work. The probability distribution generated by the two models may be combined using arithmetic or geometric averaging, weighted by the entropy of the distributions [26.29].

Improved prediction performance can also be achieved by allowing the model to estimate and combine probabilities based on multiple features of the musical surface. Multiple viewpoint frameworks were originally developed by *Darrell Conklin* [26.27, 30, 31] to allow the integration of information from models of different features (inspired in part by *Ebcioğlu* [26.32]). The framework assumes a symbolic representation in which music is represented as a sequence of discrete events composed of a finite number of attributes each of which may assume a value from a finite alphabet. For example, a melody is often represented as a sequence of notes, each of which is composed of a pitch, onset time, duration and loudness.

A viewpoint is a mapping from a sequence of musical events to an element from the alphabet associated with the viewpoint. Basic viewpoints are projection functions associated with the attributes of the events (i. e., pitch, onset, duration and loudness in the example above). The framework also allows the specification of derived viewpoints (e.g., pitch interval, contour, scale

degree), which are derived from a basic viewpoint (e.g., pitch in this case). Note that some viewpoints may be undefined at particular locations (e.g., pitch interval for the first event in a melody). Test viewpoints are viewpoints that return a Boolean value (e.g., whether a note falls on a tactus beat or not) and threaded viewpoints represent a base viewpoint (e.g., pitch interval) at points where a test viewpoint is true (e.g., pitch interval between notes falling on tactus beats) thereby allowing for sequences of nonsequential events. Finally, linked viewpoints represent the Cartesian product of two or more primitive viewpoints – for example, a link between pitch interval and scale degree will have an alphabet composed of pairs, whose first element is a pitch interval and whose second is a scale degree.

When modeling music with a multiple viewpoint system, separate models are constructed for each viewpoint included in the system and the resulting distributions are combined in much the same way as the long-term and short-term model outputs are combined. The distributions are first mapped back into distributions over the alphabet associated with the basic feature from which they are derived (e.g., pitch in the above examples) so that they can be combined. Typically, the viewpoint models are combined in a first stage separately for the long- and short-term models, which are then combined [26.27, 28, 33]. Multiple-viewpoint models have been developed for the domains of melody, harmony and voice leading [26.28, 33, 34].

When configured appropriately and trained on relevant corpora, these methods both improve prediction performance of the models [26.27, 28] and also account accurately for listeners' pitch expectations in melody [26.28, 35–40]. In some cases, the model parameters that optimize prediction performance do not improve fit to human perception and vice versa [26.28] suggesting that human expectation may be subject to constraints (such as memory or representational limitations) that prevent optimal prediction.

26.1.4 Beyond Simple Markov Models: Hidden Markov Models and Dynamic Bayesian Networks

The models of the Markov and *n*-gram family discussed in Sect. 26.1.3 are essentially equivalent to probabilistic versions of strictly local grammars (Chap. 25). One important feature of Markov models is that they can only model local sequential dependencies and do not assume any underlying deep structure (hidden variables). Although they operate directly on surface symbols, it is possible to use multiple viewpoint frameworks to allow such models to operate on nonsequential events (e.g., notes on tactus beats or phrase-final events, [26.27, 28]) and on representations of higher-order structure (e.g., phrase classes). However, although some of these formal limitations in expressive power may be addressed in part with the multiple-viewpoint approach, more expressive models of sequential structure have been developed in machine learning research, many of which have been applied to music. In the Chomsky hierarchy, the different model classes (from finite-state to finite-context) assume an underlying deep structure (represented using nonterminal symbols) that predicts the surface terminal symbols (Chap. 25); in an analogous way, many modeling approaches take advantage of an explicit representation of deep structure in music.

One well-known example is hidden Markov models (HMMs, e.g., [26.41, 42]; for an introduction see the comprehensive review by [26.43]), which correspond to probabilistic extensions of finite-state automata. As an extension of Markov models, the hidden Markov model assumes the Markov transition matrix not as a model characterizing transitions between surface symbols (as in the *visible* Markov models described above), but as transitions between deep structural (hidden) states that themselves emit surface symbols from associated emission distributions over the terminal alphabet. In other words, it assumes a Markov model as the underlying deep structure of states that govern the symbol distribution over subsections of the sequence.

HMMs have been employed to model various aspects of music, including, for instance, melodic structure [26.44, 45], meter and rhythm [26.45, 46], text setting [26.47] and harmonic structure [26.34, 48, 49]. Modeling harmony in a corpus of jazz standards, *Rohrmeier* and *Graepel* [26.34] found that a simple HMM modeling chord sequences exhibited barely any overfitting of its training data. Dynamic Bayesian networks (DBNs, [26.50]) generalize HMMs and constitute a family of graphical models that model the dependency structure of different (temporal) deep-structural features. DBNs were applied to modeling music by *Paiement* et al. [26.51] as well as *Raczynski* et al. [26.52] to model polyphonic pitch structure. Furthermore, DBNs can be straightforwardly adapted to modeling the interaction of different parallel feature-streams in the framework of HMMs. *Rohrmeier* and *Graepel* [26.34] implemented a DBN modeling Jazz harmony using features of duration and mode to improve predictive power though the approach does not extend to derived viewpoints.

Most of the models of sequence processing discussed so far drew their motivation from the modeling of musical expectation (see also [26.53], for an extensive account of the role of expectation in music perception). Models with a rich deep structure, however, may be used to understand other aspects of music

processing such as inferring information encoded in the deep structure of the sequence. For instance *Raphael* and *Stoddard* [26.48] used a type of DBN for the inference of harmonic structure from the surface sequence of events. *Mavromatis* [26.45] employed a model selection procedure to find the optimal topology for a HMM, using this to draw theoretical and cognitive conclusions regarding representation of deep structure. He applied the model to two cases, the statistical learning and segmentation paradigm used by Saffran and colleagues (e.g., [26.54]) and metrical induction from rhythmical patterns in a corpus of Palestrina's vocal music. *Mavromatis* [26.46] extended this approach and drew computationally informed conclusions regarding a discussion surrounding Renaissance meter.

In summary, while the application of deep structure models and DBNs in cognitive music research is growing, there is a great potential to employ these models (in combination with model selection) to understand the role of deep structure in the representation and processing of musical syntax.

26.1.5 Hierarchical Models

While computational implementations of Markovian approaches (and derived approaches such as HMMs and DBNs) have largely addressed the problem of modeling effects of expectancy and prediction, many computational approaches have sought explicitly to implement hierarchical generative models of music. The motivation derives from theoretical insights that tonal music is organized in a hierarchical fashion and, accordingly, cognitive models of music processing should be able to account for such structural complexity. Moreover, theoretical hierarchical accounts of music stress that human music cognition involves substantially more than computation of sequential predictions including, in particular, the perception of large-scale processes (e.g., musical form), reductive listening, experience of hierarchical tension and recognizing similarity (see Chap. 25 for a discussion of various ways to motivate and explore the understanding of musical syntax). Accordingly, computational models of hierarchical structure have been inspired by a diverse array of modeling goals.

The hierarchical and generative branches of music theory mainly trace back to Schenkerian theory [26.55, 56]. Apart from Schenkerian theory itself, there have been three major lines of hierarchical modeling research: the generative theory of tonal music (GTTM) and tonal pitch space [26.57, 58], which originated from the goal of framing Schenkerian analysis in terms of formal linguistic approaches; *Narmour*'s theory of melodic expectancy and complexity [26.13, 59]; and approaches to tonal harmony that employ methods from generative linguistics and formal language theory [26.5, 6, 60, 61]. Each of these formal approaches have inspired efforts to build computational models.

Schenkerian Analysis and Derivative Models

Schenkerian theory constitutes one of the earliest and most comprehensive formal approaches to syntax in tonal music and remains dominant in music-theoretical teaching. It has been the object of several computational approaches from the early days of computing to the present day.

Early work by *Kassler* focused on understanding Schenkerian-like operations of analysis from the perspective of formal languages [26.62]. He formalized a subset of Schenkerian derivations with primitive-recursive functions [26.63] and described a basic implementation of a (presumably two-voice) model of Schenkerian analysis in terms of primitive operations on a matrix representation of pitch sequence, such as an Ursatz axiom, arpeggiation, neighbor note prolongation, simplified interruption (termed *articulation*), octave adjustment, bass ascent, mixture, etc. [26.62, 64, 65]. In a separate modeling attempt based on functional programming, Smoliar and colleagues used a recursive list structure of musical events that encoded a set of Schenkerian elaboration operations in terms of Lisp function calls. Drawing a direct formal analogy between (generative) linguistic parse trees and musical structure, they developed a tree-based structural representation of a Schenkerian reductive analysis [26.66–68]. In an approach similar to Schenkerian analysis techniques, *Baroni* and colleagues applied formal grammars to melodic structure [26.69, 70]. However, none of these early approaches resulted in a complete, fully automatic functioning model of Schenkerian analysis. *Marsden* [26.71] suggested that this may be due to to the massive explosion of combinatorial complexity that arises when encoding a formal account of Schenkerian-style reductive analysis at the level of notes.

With greater computational resources available, a number of researchers have recently returned to the problem of implementing Schenkerian analysis. *Mavromatis* and *Brown* [26.72] proposed a theoretical approach to modeling Schenkerian analysis with probabilistic context-free grammars. However, their model did not reach the stage of a full implementation due to issues of complexity (see [26.71]). *Marsden* [26.73] proposed a graph-theoretical representation of reductive structures (termed *E-Graph*) suitable for computational implementation. Subsequently, *Marsden* [26.74] proposed an expanded representation of Schenkerian reduction in a generative framework. Using the limited case of short musical phrases, Marsden

developed a preliminary implementation of Schenkerian analysis [26.74] and building on this theoretical framework, a first proof-of-concept prototype was developed [26.71].

Recent research on implementing Schenkerian theory includes *Yust* [26.75], who proposed the structure of *maximal outerplanar graphs* (MOP) as representations for the structure of Schenkerian prolongation, and the work of *Kirlin* (e.g., [26.76]) who has developed a corpus of 41 pieces annotated with corresponding Schenkerian analyses in machine-readable format [26.77]. *Kirlin* and *Jensen* [26.78] use supervised probabilistic learning to uncover deep hierarchical musical structure, including an algorithm for deriving the most probable analysis for a given piece of music.

The Implication–Realization Model

Breaking with Schenkerian tradition (see [26.79]), *Eugene Narmour* developed a distinct theoretical approach to modeling melodic expectancy [26.13] as well as melodic complexity and reduction [26.59]. The *implication-realisation model* is based on the implications that a melodic interval (the implicative interval) has for the following interval (the realized interval) and presents a detailed classification of interval pairs based on the size and direction of the component intervals. The model consists of two independent perceptual systems – the bottom-up and top-down systems of melodic implication. While the principles of the former are held to be automatic, unconscious and universal, the principles of the latter are held to be learned and hence culture dependent. Although the model is presented in a highly analytic manner, it has psychological relevance because it proposes hypotheses about general perceptual principles that are precisely and quantitatively specified and therefore amenable to empirical investigation [26.80, 81]. Research has also compared the implication-realization model with variable-order Markov models in terms of how well they account for listeners' pitch expectations [26.28, 36, 37]. Furthermore, Grachten and colleagues have examined the use of the implication-realization model in a computational model of similarity for use in music information retrieval [26.82].

Furthermore, *Narmour* [26.59] presents detailed proposals for the ways in which basic musical structures (pairs of intervals) may form larger sequential units (chains) and larger hierarchical units (transformations) based on the idea of closure, which occurs when a structure does not generate a strong implication. However, these hierarchical aspects of the theory have been somewhat neglected in terms of computational modeling and empirical evaluation.

The Generative Theory of Tonal Music (GTTM)

The GTTM [26.57] constitutes probably the most influential model in empirical musicology to date. It models the perception of musical structure in term of the interaction of four levels of structure: grouping structure, metrical structure, time-span reduction and prolongational reduction (see [26.58, Ch. 1] for a review); these levels are defined in terms of *well-formedness rules* exhaustively defining the (very large) set of well-formed candidate analyses and *preference rules* that define rules to select the best analysis from the well-formed candidates. Although it is partly inspired by generative grammars for language, it is important to note that the GTTM is not a formal grammar because no (tree-defining) context-free rules are specified, and it is unclear whether it could be developed into one. Furthermore, the GTTM is a model of the final representation of a piece of Western tonal music as it might appear in the mind of an idealized listener, enculturated in Western music. It does not account for any effects of stylistic enculturation nor make any predictions about the dynamic nature of structure perception during listening (this was later addressed by *Jackendoff* [26.83]).

While the GTTM is specified to a much higher degree of detail and specificity than Schenkerian theory, it is still highly imprecise from a computational point of view. For instance, it does not specify a ranking for the preference rules and in some cases assumes human musical intuition for making analytical decisions. These factors pose challenges for computational implementations of the GTTM. Nonetheless, *Lerdahl* [26.58] has devised a model of musical tension that is based on a complete GTTM analysis and predicts musical tension based on local and global factors [26.58]. The model has been found to predict participants' continuous ratings of musical tension for a small number of musical pieces [26.84]. To date, the most extensive progress towards an implementation of the GTTM has been made by *Hamanaka* and colleagues [26.85, 86].

Generative Grammars for Music

Many authors have proposed recursive generative grammars for modeling the hierarchical organization of harmonic sequences, an idea whose essence goes back to *Riemann* [26.87]. Following the formalization of context-free rewrite grammars and the Chomsky hierarchy, a number of earlier approaches applied these techniques to music (e.g., [26.69, 70, 88, 89]). More recent approaches include *Steedman* [26.5, 6] and *Rohrmeier* [26.60, 61]. Based on Steedman's categorical grammar formalism [26.90], *Granroth-Wilding* and *Steedman* [26.91] implemented a model of jazz harmony that extends Steedman's earlier theoretical gram-

mars for blues harmony [26.6]. Similarly, *De Haas* et al. developed an implementation of Rohrmeier's grammar for the purposes of music information retrieval, such as harmonic similarity and transcription [26.92–94]. *Tidhar* [26.95] implemented adapted versions of a context-free grammar formalism for the parsing of Couperin's unmeasured preludes.

Other approaches have attempted to combine context-free grammars with probabilistic learning. *Gilbert* and *Conklin* [26.96], for example, have employed a probabilistic context-free grammar for modeling melodic reduction. *Bod* [26.97] argues for a memory-based approach to modeling melodic grouping structure as an alternative to the Gestalt approach based on rules. He used grammar learning techniques to induce the annotated phrase structure of the Essen Folk Song Collection [26.98, 99]. Three grammar induction algorithms were examined: first, the treebank grammar learning technique, which reads all possible context-free rewrite rules from the training set and assigns each a probability proportional to its relative frequency in the training set; second, the Markov grammar technique, which assigns probabilities to context-free rules by decomposing the rule and its probability by an n-th-order Markov process, allowing the model to estimate the probability of rules that have not occurred in the training set; and third, a Markov grammar augmented with a data-oriented parsing (DOP) method for conditioning the probability of a rule over the rule occurring higher in the parse tree. A best-first parsing algorithm based on Viterbi optimization was used to generate the most probable parse for each melody in the test set given each of the three models. The results demonstrated that the treebank technique yielded moderately high precision but very low recall ($F = 0.065$), the Markov grammar yielded slightly lower precision but much higher recall ($F = 0.706$), while the Markov-DOP technique yielded the highest precision and recall ($F = 0.810$). A qualitative examination of the folk song data reveals a number of cases (15% of the phrase boundaries in the test set) where the annotated phrase boundary cannot be accounted for by Gestalt principles but that are predicted by the Markov-DOP parser.

26.1.6 Neural Networks

Neural networks represent a different class of models that have been used to understand the representation and processing of musical syntax. Rather than basing the model on a specific representation of musical structure (e.g., n-grams, production rules, music-theoretic principles), neural networks are biologically inspired in the sense that they take their motivation from basic properties of the brain (e.g., parallel processing across simple units, distributed representations, synaptic connectivity, graceful degradation and Hebbian learning). Note, however, that while they are *biologically inspired*, most neural network models (especially so-called connectionist models) are not actually *biologically plausible* models of neural processing. Thus they remain at *Marr*'s [26.100] algorithmic/representational level rather than at the implementational/physical level of description. At this level, one practical difficulty with neural networks as simulations of cognitive processing is that the nature of the learned representations are often not easily interpretable.

Mozer [26.101], for instance, developed a model based on a recurrent artificial neural network (RANN, [26.102]) and used psychoacoustic constraints in the representation of pitch and duration. In particular, the networks operated over predefined, theoretically motivated multidimensional spatial representations of pitch (which emphasized a number of pitch relations including pitch height, pitch chroma and fifth relatedness, [26.103]) and duration (emphasizing such relations as relative duration and tuplet class). These neural networks are trained within a supervised regime in which the discrepancy between the activation of the output units (the expected next event) and the desired activation (the actual next event) is used to adjust the network weights at each stage of training. When trained and tested on sets of simple artificial pitch sequences with a split-sample experimental paradigm, the RANN model outperformed simple digram models. In particular, the use of cognitively motivated multidimensional spatial representations led to significant benefits (over a local pitch representation) in the training of the networks. However, the results were less than satisfactory when the model was trained on a set of melodic lines from ten compositions by J.S. Bach and used to generate new melodies; the neural network architecture appeared unable to capture the higher-level structure in these longer pieces of music.

The question arises of whether these neural network models provide good simulations of human processing. In an artificial grammar learning paradigm of melodic structure [26.40, 104], a simple recurrent network model [26.102] was compared with an n-gram model [26.28] and a chunking model [26.105]. Results indicate that the n-gram model achieved by far the best performance, yet the simple recurrent network exhibited characteristic patterns of performance (including errors) that were closer to the human level [26.106, 107].

One approach to addressing the apparent inability of RANNs to represent recursive constituent structure in music involves what is called auto-association. An auto-associative network is simply one that is trained

to reproduce on its output layer a pattern presented to its input layer, generally forming a compressed representation of the input on its hidden layer. For example, training a network with eight-unit input and output layers separated by a three-unit hidden layer with the eight one-bit-in-eight patterns typically results in a three-bit binary code on the hidden units [26.108]. *Pollack* [26.109] introduced an extension of auto-association called recursive auto-associative memory (RAAM), which is capable of learning fixed-width representations for compositional tree structures through repeated compression. The RAAM architecture consists of two separate networks: first, an encoder network that constructs a fixed-dimensional code by recursively processing the nodes of a symbolic tree from the bottom up; and second, a decoder network that recursively decompresses this code into its component parts until it terminates in symbols, thus reconstructing the tree from the top down. The two networks are trained in tandem as a single auto-associator. *Large* et al. [26.110] examined the ability of RAAM to acquire reduced representations of Western children's melodies represented as tree structures according to music-theoretic predictions [26.57]. It was found that the trained models acquired compressed representations of the melodies in which structurally salient events are represented more efficiently (and reproduced more accurately) than other events. Furthermore, the trained network showed some ability to generalize beyond the training examples to variant and novel melodies although, in general, performance was affected by the depth of the tree structure used to represent the input melodies with greater degrees of hierarchical nesting leading to impaired reproduction of input melodies. However, the certainty with which the trained network reconstructed events correlated well with music-theoretic predictions of structural importance [26.57] and cognitive representations of structural importance as assessed by empirical data on the events retained by trained pianists across improvised variations on the melodies.

Recent developments in neural networks have led to successful modeling of musical structure using restricted Boltzmann machines (RBMs) [26.111, 112]. RBMs appear to be approaching the prediction performance of the best-performing variable-order Markov models described in Sect. 26.1.3 [26.112]. However, these neural network models are difficult to analyze and it remains to be seen how successful they will be in modeling cognitive processing of musical syntax. Nonetheless, they do at least demonstrate that such processing can be implemented using parallel distributed (and potentially nonsymbolic) representations.

26.2 Psychological Research

Computational models shed light on the plausibility of specific assumptions about, and constraints on, the nature of the cognitive representations and algorithms underlying music perception and serve as analytical tools to explore the types of syntactic structures present in music. However, to understand the kinds of syntactic structures that are perceived and represented by listeners, it is important to compare empirically the output of a model with the responses of listeners. Empirical research on musical listening can help identify the power, the limits and constraints of human perception and cognition of musical syntax. In this context, computational models are useful for generating hypotheses and selecting stimuli that differ quantitatively in terms of their syntactic properties (e.g., grammaticality, uniformity, see Chap. 25). Although there are notable differences between language and music (e.g., in terms of lexical categories and the nature of semantic content), cognitive scientific research on music follows an analogous approach, exploring processing via a combination of theoretical inquiry, computational modeling and psychological/neuroscientific testing [26.113]. Furthermore it has been suggested that music and language, which are both sequential auditory forms of human communication, may share similarities at more abstract levels of cognitive and neural representation [26.114–116], a point to which we return below.

Psychological research has established that enculturated listeners are sensitive to sequential structure in music; however, research paradigms have sometimes been driven by (overly) simple representational assumptions. We will examine the evidence relating to harmonic movement, melody and high-level form. One topic that continues to attract debate is the extent to which listeners are capable of representing hierarchical, nonsequential relationships in music. We examine research on this question below.

26.2.1 Perception of Local Dependencies

As discussed in Chap. 25, harmony constitutes one of the core building blocks in Western tonal music and it has been studied in the context of different theories of syntax. Accordingly, a large number of experimental studies have focused on the perception of harmonic structure from a variety of perspectives.

At a local level, listeners are able to detect harmonic relations between successive chords as evidenced by longer reaction times to in-key than out-of-key harmonic transitions in the paradigm of musical *priming* studies ([26.117, 118], see [26.119] for a review). These findings have been replicated with longer sequences of chords and more complex harmonic relations [26.120, 121]. Furthermore, there is evidence that these harmonic priming effects are affected both by the global context of the prevailing key (which determines the overall stability of a chord), by the local harmonic context and by the rhythmic organization of the progression [26.120]. Research has also examined whether these effects are best explained by learned properties of harmonic movement or by sensory properties of the target chords [26.121–123]. The results consistently show that the structural properties of harmonic movement have a stronger effect than sensory influences such as repetition priming [26.124, 125], even when efforts are made to make these sensory influences very strong. Furthermore, *Tillmann* et al. [26.126] used self-organizing maps (a variety of neural network) to argue that these priming effects can be explained by learning of tonal organization through musical exposure.

We can ask similar questions about the perception of structure in melody. Early work on predictive processing of musical structure focused on rules of melodic organization and how they influence pitch expectations [26.13, 80, 81]. Empirical studies of these rules have found that listeners' melodic expectations do generally exhibit influences of pitch proximity (smaller intervals are more expected) and pitch reversal (large pitch intervals are expected to be followed by smaller ones in the opposite registral direction) [26.53, 127]. Actual melodies also exhibit these properties [26.128], possibly reflecting physical constraints of performance – the difficulty of producing large intervals accurately and tessitura constraints producing regression to the mean after large intervals [26.129–131]. Therefore, it remains possible that listeners learn these regularities (or variants of them, subject to cognitive constraints, [26.53, 132]), which are then reflected in their pitch expectations.

In fact, there is empirical evidence of implicit learning of regularities in musical melody and other sequences of pitched events ([26.40, 54, 104, 133]; see [26.8] for a review on implicit learning of music). Consistent with an approach based on statistical learning, melodic pitch expectations vary between musical styles [26.134] and cultures [26.135–140], throughout development [26.141] and across degrees of musical training and familiarity [26.36, 37, 134]. Furthermore, pitch expectations appear to be informed both by long-term exposure to music [26.142] and by the encoding of regularities in the immediate context [26.133]. As with harmonic expectations, computational models relying on nonhierarchical statistical learning of regularities have proved highly successful in predicting listeners' melodic expectations [26.28, 37, 38].

All of these effects can be accounted for by local models corresponding to strictly local grammars. Therefore, it is crucial to ask which aspects of hierarchical and nonlocal structures affect music listening and processing.

26.2.2 Perception of Nonlocal Dependencies

There are several ways in which nonlocal dependencies can be expressed in music, ranging from short chord sequences (such as "I IV V/ii ii V I", in which the IV chord implies the V chord and not V/ii, and the second I chord prolongs the initial one, rendering the whole sequence a constituent of a prolonged tonic) to dependencies between (sub)phrases (e.g., antecedent-consequent patterns) and between larger formal units like the parts of a minuet or a sonata [26.143]. It may even be possible to understand these dependencies at different timescales (chords, parts of phrases, phrases, movements) as recursive instantiations of similar structures [26.57, 143].

Deutsch [26.144] describes experiments in which subjects were presented with sequences of 12 notes, which they recalled in musical notation. Half the sequences were structured in accordance with the model of *Deutsch* and *Feroe* [26.10], described above, such that a higher-level subsequence of four elements acted on a lower-level subsequence of three or four elements while the remaining sequences were unstructured. Sequences were presented in one of three conditions: first, with no temporal segmentation; second, temporally segmented in accordance with tonal structure; and third, temporally segmented in violation of tonal structure. The results demonstrated that recall was high in the first and second conditions and low in the third condition and for unstructured sequences, suggesting that hierarchically structured sequences are better encoded in memory. However, on a similar task, *Boltz* and *Jones* [26.145] have found that rule recursion has only a modest effect on memory for melodies and only in certain conditions.

Regarding large-scale form, researchers have studied the aesthetic judgments of musicians and nonmusicians for pieces of music in which the large-scale tonal form has been disrupted by rearranging or rewriting certain parts. These studies have been conducted by rearranging the various movements of Beethoven sonatas and string quartets [26.146], reordering the variations making up Bach's Goldberg variations [26.147],

altering the endings of excerpts from Romantic and Classical works (1–6 min) such that they start and end in a different key ([26.148]; methodologically criticized by *Gjerdingen* [26.149]) and using rearranged versions of the opening movement of Mozart's Symphony in G Minor [26.150]. The results consistently suggest that listeners gain as much pleasure from the altered versions as the unaltered versions and musicians are no more affected than nonmusicians by these disruptions [26.148, 150]. In a study of musicians listening to original and rearranged versions of six keyboard works by Handel [26.151] found that both versions were deemed equally conforming to stylistic expectations and accuracy in judging whether the starting and ending key were the same was at chance. These findings suggest that there are severe constraints on the ability to build cognitive representations of large-scale formal relationships in music, even for musicians (though none of these studies examined professional musicians with a very high level of expertise).

However, there is crucial evidence that listeners are sensitive to long-distance hierarchical dependencies at the phrase-level and midrange timescales. *Koelsch* et al. [26.152], examined responses to the final chord in a pair of chorale melodies composed of two subphrases, together with modified versions in which the first phrase was altered to break tonal closure with the final phrase. The results showed strong characteristic differences between the two versions in the neural response (the early right anterior negativity (ERAN), Sect. 26.3.2) to the final chord. In addition, behavioral measures showed that there were no differences in emotional response (valence and arousal) between the versions, suggesting that nonlocal harmonic dependencies are independent of emotional expression. Further behavioral measures showed that the final chord of the original version was judged to close the sequence better than the final chord of the altered versions (although the difference was relatively small, pointing towards the use of implicit rather than explicit knowledge). Fundamentally, because the harmonic dependencies in this study exceed ten chords, it is impossible that nonveridical *n*-gram or Markov models could explicitly represent the difference. Accordingly, this evidence for nonlocal dependencies affecting the perception of phrase closure falsifies the assumption that simple local models of harmony can adequately model human perception of musical syntax.

26.3 Neuroscientific Research

26.3.1 Introduction

It is pertinent to ask what cognitive neuroscience can tell us over and above research in experimental psychology and cognitive science [26.153–156]. Clearly, it can tell us something about the neural basis of psychological processes; less obviously, but perhaps more importantly, it can also tell us something about those psychological processes themselves. Once a neural response has been linked to a specific psychological process (and this itself may require great effort to establish), then we can use it as an additional measure of that process, alongside behavioral measures. Such neural responses have the potential advantages that they may be more sensitive and less prone to various kinds of bias than behavioral measures. Potential disadvantages include the difficulty of establishing direct, specific relationships between features of the neural response and particular psychological processes.

26.3.2 Neural Basis of Syntactic Processing in Music

Neuroscientific research has used electroencephalography (EEG) to investigate event-related-potential (ERP) responses to violations of harmonic syntax [26.157–164]. Two characteristic brain responses have been reported: an early anterior negativity (EAN) with a latency of 150–280 ms (sometimes right lateralized and referred to as the ERAN – early right anterior negativity), and a later bilateral or right-lateralized negativity (N5) with a latency of 500 ms [26.157, 161]. The EAN is thought to reflect the violation of harmonic expectation, while the N5 is thought to reflect the higher processing effort needed to integrate unexpected harmonies into the ongoing context [26.163]. The amplitude of the EAN is related to the long-term transition probability of the chord [26.165, 166] suggesting that it reflects implicit learning of harmonic movement through experience (see Sect. 26.1.3 above). Consistent with this proposal are findings that the EAN is attenuated (though still present) in five to six-year-old children compared to adults and accentuated in adult musicians, relative to adult nonmusicians [26.159].

To date, less is known about the neural correlates of structural processing in other aspects of music such as pitch, rhythm and timbre. Early studies [26.167–171] identified a late positive component (LPC) peaking between 300–600 ms at central and posterior sites in response to stylistically unexpected notes in a melody.

The amplitude and latency of the LPC are sensitive to musical expertise, the familiarity of the melody, the degree of expectancy violation [26.167], and also to the timing of the unexpected note [26.168].

There is an important distinction between schematic representations of the syntactic rules governing a musical style and veridical memory for the structure of a familiar piece of music [26.172]. *Miranda* and *Ullman* [26.173] describe a functional dissociation between two ERP components: an early (150–270 ms) anterior-central negativity (i. e., the EAN) associated with out-of-key violations in both familiar and unfamiliar melodies, and a subsequent (220–380 ms) posterior negativity elicited by both in-key and out-of-key violations of familiar melodies only. They suggested that these two components are driven by violations of musical rules (of tonality/harmony) and of veridical memory representations of familiar melodies respectively. In order to focus exclusively on schematic acquisition of syntax, *Loui* et al. [26.166] examined neural responses to deviant melodic endings in pitch sequences generated according to an artificial (strictly local) grammar, using pitches taken from the unfamiliar Bohlen–Pierce scale. They report an EAN, whose amplitude increased with greater learning of the grammar (measured by degree of exposure and performance in a grammaticality decision task). Again, this suggests that the EAN reflects a process of implicit statistical learning of sequential dependencies in the auditory environment (Sect. 26.1.3).

The EAN to violations of melodic syntax tends to occur earlier (around 100 ms) than to violations of harmonic syntax (circa 180 ms) [26.37, 38, 174], perhaps indicating that single notes are processed more quickly than chords. Using a computational model of auditory expectation [26.28], which makes probabilistic pitch predictions based on statistical learning, to identify notes in melodies that varied systematically in information content (IC), *Omigie* et al. [26.39] showed a linear relationship between the amplitude of the EAN and IC. *Pearce* et al. [26.37] reported an increase in the amplitude of beta oscillations for high information content (low probability) notes at a latency of around 500 ms in centroparietal regions and in phase locking in the same time window between electrodes located over centroparietal and occipital regions. Again these results are consistent with the proposal that these neural indicators of syntactic processing reflect probabilistic inference based on implicit statistical learning.

26.3.3 Syntax in Music and Language

EEG research has also addressed the question of whether there exists an overlap in neural processing of musical and linguistic syntax. Violations of syntax in language often generate two characteristic ERPs: a left anterior negativity (LAN) peaking around 300–450 ms at frontal scalp locations, and a positive-going deflection termed the P600 at a latency of about 600 ms with a posterior distribution. An early study suggested that violations of harmonic syntax generate an increased P600, which is very similar to that induced by violations of linguistic syntax [26.162]. More recent research has presented music synchronously with visually presented sentences where each word coincides temporally with a note or chord in the music. Introducing syntactic violations in the music and language allows the investigation of conditions where the violations are congruent or incongruent in the two domains. This research suggests that the LAN to syntactic violations in language is reduced by unusual harmonic movement (e.g., a Neapolitan chord, [26.175]) and also by low-probability notes in melodic phrases [26.176]. There is also evidence that the ERAN is reduced when presented concurrently with a linguistic violation [26.177]. Interestingly, these interactive effects are not in evidence when musical violations are paired with semantic incongruities in language [26.176, 177].

Analogies have been drawn between the ERAN and the early left anterior negativity (ELAN) often observed in response to violations of syntax in language [26.178]. Interestingly, children with specific language impairment not only show characteristic changes in the ELAN to language [26.179] but also a reduced ERAN to harmonic violations [26.180]. Broca's aphasics also show reduced neural responses to harmonic violations in music [26.181] and there is evidence from magnetoencephalography (MEG) research that the ERAN to violations of harmonic syntax originates in Broca's area and its right hemisphere homologue [26.182]. Furthermore, functional magnetic resonance imaging (fMRI) studies [26.183–186] suggest that violations of harmonic syntax induce activation in the inferior frontal cortex, which is also suggestive of a relationship between the neural processing of syntax in music and language [26.114].

26.3.4 Grouping Structure

One other aspect of musical structure that has been studied from the perspective of cognitive neuroscience is grouping structure. Using EEG and MEG, *Knösche* et al. [26.187] found that phrase boundaries in melodies generated a late (500–600 ms for EEG, 400–700 ms for MEG) positive deflection, which they termed the *closure positive shift (CPS)*. Source localization suggested that the CPS was generated by structures in the limbic system, including anterior/posterior cingulate and pos-

terior mediotemporal cortex. Similar neural responses have been observed at syntactic phrase boundaries in language [26.188], which seem to be related to prosodic cues [26.189, 190]. A subsequent study showed that the CPS to musical phrase boundaries is stronger in musicians than in nonmusicians [26.191], suggesting that strategies for segmenting music are influenced by musical training.

26.4 Implications and Issues

We have reviewed empirical research on musical syntax from computational, psychological and neuroscientific perspectives. We close with a discussion of two issues that naturally arise during this discussion: first, the extent to which the different perspectives on musical syntax converge; and second, the relationship between syntax and semantics in music, which naturally invokes the question of affective responses to music.

26.4.1 Convergence Between Approaches

One of the key challenges facing future research is to integrate insights from these different methodological and epistemological approaches. Computational modeling allows us to implement theories of musical structure and syntax and examine the behavior of the implemented algorithm when supplied with musical examples. This can provide insights into optimal syntactic representations and absolute constraints on the syntactic structure of a musical style [26.192]. Psychological research, on the other hand, can indicate the kinds of syntactic forms and relationships that listeners are capable of perceiving, representing and learning. Implementing a psychological theory as a computational model requires the theory to be precisely expressed and also allows the theory to be tested and refined through quantitative comparison of human and model responses. It also allows comparison of human performance at different levels of experience and expertise with optimal syntactic parsing. Finally, neuroscientific research provides information about the neural basis of syntactic processing, which can impose constraints on human syntactic processing and also provide data that is more sensitive to implicit knowledge than behavioral methods (e.g., [26.193, 194]. Therefore, future research should triangulate more explicitly between computational modeling, psychological experimentation and cognitive-neuroscientific investigation in further developing our understanding of musical syntax [26.37, 195].

26.4.2 Syntax, Semantics and Emotion

Musical styles can be said to possess syntax in an analogous way to that in which natural languages (or programming languages) do. However, music is different from natural language in that musical elements do not usually carry clear referential and propositional semantics in the way that linguistic atoms do. It is sometimes possible to establish indexical references for appropriately enculturated listeners – the old castle, the sea, a storm, the spring, love, James Bond's theme or Brunhilde's leitmotif being good examples taken from various pieces of music. However, it is impossible to communicate complex statements like *had he not gone to sea last spring and then returned to the old castle, James Bond would not have fallen in love with Brunhilde* (see [26.113, 196, 197] for further discussion). Therefore, meaning in musical communication is borne to a large extent by syntactic structure and the listener's perception of structural relations between musical elements [26.198]. In particular, the syntactic structure of music affords the communication of patterns of tension and resolution, through the systematic manipulation of the listener's structural expectations, based in turn on their internalized syntactic representations of the style.

There is one prominent theoretical perspective on affective responses to music that is relevant here since it relates the expression and perception of meaning to predictive processing of musical structure, using information-theoretic principles [26.53]. Building on arguments made by *Hanslick* [26.199], *Meyer* [26.198, 200] examines from a theoretical perspective the dynamic cognitive processes in operation when we listen to music and how these processes not only underlie the listener's understanding of musical structure but also give rise to the communication of affect and the perception of meaning in music. Meyer proposes that meaning arises through the manner in which musical structures activate, inhibit and resolve expectations in the listener about forthcoming musical structures. He notes that these expectations may differ independently in terms of the degree to which they are passive or active, their strength and their specificity. He contends, in particular, that affect is aroused when a passive expectation induced by antecedent musical structures is made active by it being temporarily inhibited or permanently blocked by consequent musical structures. Meyer discusses three ways in which the listener's expectations may be violated. The first occurs when the

expected consequent event is delayed, the second when the antecedent context generates ambiguous expectations about consequent events, and the third when the consequent is unexpected. While the particular effect of music is clearly dependent on the strength of the expectation, Meyer argues that it is also conditioned by the specificity of the expectation.

Meyer [26.200] discusses the relationship between his theory of musical expectancy and concepts in information theory. He starts with the suggestion that [26.200, p. 414]:

> *once a musical style has become part of the habit responses of composers, performers and practiced listeners it may be regarded as a complex system of probabilities*

and that expectations arise out of these internalized probability systems. In particular, he suggests that a musical style may be conceived as a Markov process (Sect. 26.1.3) and that experienced listeners possess internalized models of that process. The degree to which hypothetical meanings provide ambiguous expectations about consequent structures can be measured by entropy (or uncertainty) [26.200, p. 416]:

> *The lower the probability of a particular consequent* [...] *the greater the uncertainty (and information) involved in the antecedent-consequent relation.*

An unexpected consequent conveys a maximum of information. The process of revaluation corresponds to the feedback of information such that future behavior is conditioned by the results of past events.

Meyer notes that uncertainty may arise from different sources. Thus, systemic uncertainty decreases throughout the experience of a piece of music as the listener's model develops and the composer may deliberately introduce designed uncertainty to combat this effect. Furthermore, the redundancy (lack of uncertainty) inherent in a style serves to combat noise, be it cultural (resulting from discrepancies between the habit responses of a given listener and those operating in the style) or acoustical. *Witten* et al. [26.201, p.71] make a similar distinction between perceptual uncertainty (that which is relative to a particular listener's model) and stylistic uncertainty (that which is inherent in the musical style).

As noted above, the most obvious emotional response to expectancy violation and uncertainty in music is tension but according to *Meyer* [26.198], these processes may also give rise to a range of specific emotional experiences including apprehension/anxiety (p.27), hope (p.29), and disappointment (p.182) (see also [26.202]). Convergingly, empirical research has found that stylistically unusual chord progressions do stimulate increases in physiological arousal [26.177] while notes that have a low probability of occurrence in performed melodic music [26.28] have been found to be associated with increased arousal, reduced valence, increased skin conductance and reduction in heart rate [26.35].

References

26.1 P.N. Johnson-Laird: *Mental Models* (Harvard Univ. Press, Cambridge 1983)

26.2 H.C. Longuet-Higgins: Artificial intelligence – A new theoretical psychology?, Cognition **10**(1–3), 197–200 (1981)

26.3 H.A. Simon, C.A. Kaplan: Foundations of cognitive science. In: *Foundations of Cognitive Science*, ed. by M.I. Posner (MIT Press, Cambridge 1989) pp. 1–47

26.4 A. Newell, H.A. Simon: Computer science as empirical enquiry: Symbols and search, Commun. ACM **19**(3), 113–126 (1976)

26.5 M. Steedman: A generative grammar for jazz chord sequences, Music Percept. **2**(1), 52–77 (1984)

26.6 M. Steedman: The blues and the abstract truth: Music and mental models. In: *Mental Models in Cognitive Science*, ed. by A. Garnham, J. Oakhill (Erlbaum, Mahwah 1996) pp. 305–318

26.7 A. Clark: Learning trees from strings: A strong learning algorithm for some context-free grammars, J. Mach. Learn. Res. **14**, 3537–3559 (2013)

26.8 M. Rohrmeier, P. Rebuschat: Implicit learning and acquisition of music, Top. Cogn. Sci. **4**(4), 525–553 (2012)

26.9 N. Chomsky: *Aspects of the Theory of Syntax* (MIT Press, Cambridge 1965)

26.10 D. Deutsch, J. Feroe: The internal representation of pitch sequences in tonal music, Psychol. Rev. **88**(6), 503–522 (1981)

26.11 H.A. Simon, R.K. Sumner: Pattern in music. In: *Formal Representation of Human Judgement*, ed. by B. Kleinmuntz (Wiley, New York 1968) pp. 219–250

26.12 L.B. Meyer: *Explaining Music: Essays and Explorations* (University of Chicago Press, Chicago 1973)

26.13 E. Narmour: *The Analysis and Cognition of Basic Melodic Structures: The Implication-realisation Model* (University of Chicago Press, Chicago 1990)

26.14 C. Ames: Automated composition in retrospect: 1956–1986, Leonardo **20**(2), 169–185 (1987)

26.15 C. Ames: The Markov process as a compositional model: A survey and tutorial, Leonardo **22**(2), 175–187 (1989)

26.16 J.E. Cohen: Information theory and music, Behav. Sci. **7**(2), 137–163 (1962)

26.17 L. Hiller: Music composed with computers – A historical survey. In: *The Computer and Music*, ed. by H.B. Lincoln (Cornell Univ. Press, Cornell 1970) pp. 42–96

26.18 C.E. Shannon: A mathematical theory of communication, Bell Syst. Tech. J. **27**(3), 379–423 (1948), and 623–656

26.19 M.A. Rohrmeier, S. Koelsch: Predictive information processing in music cognition. A critical review, Int. J. Psychophysiol. **83**(2), 164–175 (2012)

26.20 M.T. Pearce, G.A. Wiggins: Auditory expectation: The information dynamics of music perception and cognition, TopiCS Cogn. Sci. **4**, 625–652 (2012)

26.21 D.J.C. MacKay: *Information Theory, Inference, and Learning Algorithms* (Cambridge Univ. Press, Cambridge 2003)

26.22 L. Hiller, L. Isaacson: *Experimental Music* (McGraw–Hill, New York 1959)

26.23 L. Hiller, C. Bean: Information theory analyses of four sonata expositions, J. Music Theory **10**(1), 96–137 (1966)

26.24 L. Hiller, R. Fuller: Structure and information in Webern's Symphonie, Op. 21, J. Music Theory **11**(1), 60–115 (1967)

26.25 R.C. Pinkerton: Information theory and melody, Sci. Am. **194**(2), 77–86 (1956)

26.26 J.E. Youngblood: Style as information, J. Music Theory **2**, 24–35 (1958)

26.27 D. Conklin, I.H. Witten: Multiple viewpoint systems for music prediction, J. New Music Res. **24**(1), 51–73 (1995)

26.28 M.T. Pearce: *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition*, Ph.D. Thesis (London City University, London 2005)

26.29 M.T. Pearce, D. Conklin, G.A. Wiggins: Methods for combining statistical models of music. In: *Computer Music Modelling and Retrieval*, ed. by U.K. Wiil (Springer, Berlin, Heidelberg 2005) pp. 295–312

26.30 D. Conklin: *Prediction and Entropy of Music*, Master's dissertation (University of Calgary, Calgary 1990)

26.31 D. Conklin, J.G. Cleary: Modelling and generating music using multiple viewpoints. In: *Proc. 1st Workshop AI Music, Menlo Park* (1988) pp. 125–137

26.32 K. Ebcioğlu: An expert system for harmonising four–part chorales, Comput. Music J. **12**(3), 43–51 (1988)

26.33 R. Whorley, G. Wiggins, C. Rhodes, M.T. Pearce: Multiple viewpoint systems: Time complexity and the construction of domains for complex musical viewpoints in the harmonisation problem, J. New Music Res. **42**, 237–266 (2013)

26.34 M. Rohrmeier, T. Graepel: Comparing feature-based models of harmony. In: *Proc. 9th Int. Symp. Comput. Music Model. Retr.* (Springer, London 2012) pp. 357–370

26.35 H. Egermann, M.T. Pearce, G.A. Wiggins: McAdams: Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music, Cogn. Affect. Behav. Neurosci. **13**, 533–553 (2013)

26.36 N.C. Hansen, M.T. Pearce: Predictive uncertainty in auditory sequence processing, Front. Psychol. **5**, 1052 (2014), https://doi.org/10.3389/fpsyg.2014.01052

26.37 M.T. Pearce, M.H. Ruiz, S. Kapasi, G.A. Wiggins, J. Bhattacharya: Unsupervised statistical learning underpins computational, behavioural and neural manifestations of musical expectation, NeuroImage **50**, 302–313 (2010)

26.38 D. Omigie, M.T. Pearce, L. Stewart: Tracking of pitch probabilities in congenital amusia, Neuropsychologia **50**, 1483–1493 (2012)

26.39 D. Omigie, M.T. Pearce, V. Williamson, L. Stewart: Electrophysiological correlates of melodic processing in congenital amusia, Neuropsychologia **51**, 1749–1762 (2013)

26.40 M. Rohrmeier, I. Cross: Artificial grammar learning of melody is constrained by melodic inconsistency: Narmour's principles affect melodic learning, PLOS ONE **8**(7), e66174 (2013)

26.41 L.E. Baum, T. Petrie: Statistical inference for probabilistic functions of finite state Markov chains, Ann. Math. Stat. **37**(6), 1554–1563 (1966)

26.42 L.E. Baum, T. Petrie, G. Soules, N. Weiss: A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, Ann. Math. Stat. **41**(1), 164–171 (1970)

26.43 L.R. Rabiner: A tutorial on hidden Markov models and selected applications in speech recognition, Proc. IEEE **77**(2), 257–286 (1989)

26.44 P. Mavromatis: A hidden Markov model of melody production in greek church chant, Comput. Musicol. **14**, 93–112 (2005)

26.45 P. Mavromatis: HMM analysis of musical structure: Identification of hidden variables through topology-sensitive model selection. In: *Mathematics and Computation in Music*, Communications in Computer and Information Science, (Springer, Berlin, Heidelberg 2009) pp. 205–217

26.46 P. Mavromatis: Exploring the rhythm of the Palestrina style. A case study in probabilistic grammar induction, J. Music Theory **56**(2), 169–223 (2012)

26.47 P. Mavromatis: Minimum description length modeling of musical structure, J. Math. Music **3**(3), 117–136 (2009)

26.48 C. Raphael, J. Stoddard: Functional harmonic analysis using probabilistic models, Comput. Music J. **28**(3), 45–52 (2004)

26.49 J.P. Bello, J. Pickens: A robust mid-level representation for harmonic content in music signals. In: *ISMIR*, Vol. 5 (2005) pp. 304–311

26.50 K.P. Murphy: *Dynamic Bayesian Networks: Representation, Inference and Learning*, Doctoral Dissertation (Univ. of California, Berkeley 2002)

26.51 J.-F. Paiement, Y. Grandvalet, S. Bengio: Predictive models for music, Connect. Sci. **21**, 253–272 (2009)

26.52 S.A. Raczyński, S. Fukayama, E. Vincent: Melody harmonization with interpolated probabilistic models, J. New Music Res. **42**(3), 223–235 (2013)

26.53 D. Huron: *Sweet Anticipation: Music and the Psychology of Expectation* (MIT Press, Cambridge 2006)

26.54 J.R. Saffran, E.K. Johnson, R.N. Aslin, E.L. Newport: Statistical learning of tone sequences by human infants and adults, Cognition **70**(1), 27–52 (1999)

26.55 H. Schenker: *Der Freie Satz. Neue Musikalische Theorien und Phantasien* (Margada, Liège 1935)

26.56 A. Forte, S.E. Gilbert: *Introduction to Schenkerian Analysis* (Norton, New York 1982)

26.57 F. Lerdahl, R. Jackendoff: *A Generative Theory of Tonal Music* (MIT Press, Cambridge 1983)

26.58 F. Lerdahl: *Tonal Pitch Space* (Oxford Univ. Press, Oxford 2001)

26.59 E. Narmour: *The Analysis and Cognition of Melodic Complexity: The Implication-realisation Model* (University of Chicago Press, Chicago 1992)

26.60 M. Rohrmeier: A generative grammar approach to diatonic harmonic structure. In: *Proc. 4th Sound Music Comput. Conf. (SMC07)*, ed. by C. Spyridis, A. Georgaki, G. Kouroupetroglou, C. Anagnostopoulou (2007) pp. 97–100

26.61 M. Rohrmeier: Towards a generative syntax of tonal harmony, J. Math. Music **5**(1), 35–53 (2011)

26.62 M. Kassler: Proving musical theorems I: The middleground of Hienrich Schenker's theory of tonality, Technical Report no. 103, (University of Sydney, Sydney 1975)

26.63 M. Kassler: *A Trinity of Essays*, Ph.D. Thesis (Princeton University, Princeton 1967)

26.64 M. Kassler: Explication of the middleground of Schenker's theory of tonality, Misc. Musicol. Adel. Stud. Music. **9**, 72–81 (1977)

26.65 M. Kassler: APL applied in music theory, APL Quote Quad **18**, 209–214 (1988)

26.66 R.E. Frankel, S.J. Rosenschein, S.W. Smoliar: Schenker's theory of tonal music – its explication through computational processes, Int. J. Man–Mach. Stud. **10**, 121–138 (1976)

26.67 R.E. Frankel, S.J. Rosenschein, S.W. Smoliar: A LISP-based system for the study of Schenkerian analysis, Comput. Humanit. **10**, 21–32 (1978)

26.68 S.W. Smoliar: A computer aid for Schenkerian analysis, Comput. Music J. **4**, 41–59 (1980)

26.69 M. Baroni: The concept of musical grammar (translated by S. Maguire with the assistance of W. Drabkin), Music Anal. **2**, 175–208 (1983)

26.70 M. Baroni, R. Dalmonte, C. Jacobini: Theory and analysis of European melody. In: *Computer Representations and Models in Music*, ed. by A. Marsden, A. Pople (Academic Press, London 1992) pp. 187–206

26.71 A. Marsden: Schenkerian analysis by computer: A proof of concept, J. New Music Res. **39**, 269–289 (2010)

26.72 P. Mavromatis, M. Brown: Parsing context-free grammars for music: A computational model of Schenkerian analysis. In: *Proc. 8th Int. Conf. Music Percept. Cogn., Evanston* (2004) pp. 414–415

26.73 A. Marsden: Representing melodic patterns as networks of elaborations, Comput. Humanit. **35**, 37–54 (2001)

26.74 A. Marsden: Generative structural representation of tonal music, J. New Music Res. **34**, 409–428 (2005)

26.75 J.D. Yust: *Formal Models of Prolongation*, Ph.D. Thesis (University of Washington, Washington 2006)

26.76 P.B. Kirlin: Using harmonic and melodic analyses to automate the initial stages of Schenkerian analysis. In: *Proc. Int. Conf. Music Inf. Retr. (ISMIR), Kobe* (2009) pp. 423–428

26.77 P.B. Kirlin: A data set for computational studies of Schenkerian analysis. In: *Proc. 15th Int. Soc. Music Inf. Retr. Conf* (2014) pp. 213–218

26.78 P.B. Kirlin, D.D. Jensen: Using supervised learning to uncover deep musical structure. In: *Proc. 29th AAAI Conf. Artif. Intell.* (2015) pp. 1770–1776

26.79 E. Narmour: *Beyond Schenkerism: The Need for Alternatives in Music Analysis* (University of Chicago Press, Chicago 1975)

26.80 C.L. Krumhansl: Music psychology and music theory: Problems and prospects, Music Theory Spectr. **17**, 53–90 (1995)

26.81 E.G. Schellenberg: Expectancy in melody: Tests of the implication-realisation model, Cognition **58**(1), 75–125 (1996)

26.82 M. Grachten, J.L. Arcos, R.L. de Mántaras: Melody retrieval using the Implication/Realization model. In: *Proc. 6th Int. Conf. Music Inf. Retr* (Queen Mary University of London, London 2005)

26.83 R. Jackendoff: Musical parsing and musical affect, Music Percept. **9**(2), 199–229 (1991)

26.84 F. Lerdahl, C.L. Krumhansl: Modeling tonal tension, Music Percept. **24**, 329–366 (2007)

26.85 M. Hamanaka, K. Hirata, S. Tojo: Implementing a generative theory of tonal music, J. New Music Res. **35**, 249–277 (2006)

26.86 M. Hamanaka, K. Hirata, S. Tojo: FATTA: Full automatic time-span tree analyzer. In: *Proc. Int. Comput. Music Conf. (ICMC), Copenhagen* (2007) pp. 153–156

26.87 H. Riemann: *Musikalische Syntaxis. Grundriss einer harmonischen Satzbildungslehre* (Breitkopf Härtel, Leipzig 1877)

26.88 T. Winograd: Linguistics and computer analysis of tonal harmony, J. Music Theory **12**, 3–49 (1968)

26.89 J. Sundberg, B. Lindblom: Generative theories in language and music descriptions, Cognition **4**, 99–122 (1976)

26.90 M. Steedman: *The Syntactic Process* (MIT Press, Cambridge 2000)

26.91 M. Granroth-Wilding, M. Steedman: A robust parser-interpreter for jazz chord sequences, J. New Music Res. **43**, 355–374 (2014)

26.92 W.B. De Haas: *Music Information Retrieval Based on Tonal Harmony*, Doctoral Dissertation (Utrecht Univ., Utrecht 2012)

26.93 B. De Haas, M. Rohrmeier, R. Veltkamp, F. Wiering: Modeling harmonic similarity using a generative

grammar of tonal harmony. In: *Proc. 10th Int. Soc. Music Inf. Retr. Conf. (ISMIR 2009)*, ed. by M. Goto (2009) pp. 549–554
26.94 W.B. De Haas, J.P. Magalhães, F. Wiering: Improving audio chord transcription by exploiting harmonic and metric knowledge. In: *Proc. 13th Int. Soc. Music Inf. Retr. Conf. (ISMIR 2012)* (2012) pp. 295–300
26.95 D. Tidhar: *A Hierarchical and Deterministic Approach to Music Grammars and its Application to Unmeasured Preludes*, Ph.D. Thesis (Technische Universität Berlin, Berlin 2005)
26.96 E. Gilbert, D. Conklin: A probabilistic context-free grammar for melodic reduction. In: *Proc. Int. Workshop Artif. Intell. Music, 20th Int. Jt. Conf. Artif. Intell. (IJCAI), Hyderabad* (2007) pp. 83–94
26.97 R. Bod: Memory-based models of melodic analysis: Challenging the Gestalt principles, J. New Music Res. **30**, 27–37 (2001)
26.98 H. Schaffrath: The ESAC databases and MAPPET software, Comput. Musicol. **8**, 66 (1992)
26.99 H. Schaffrath: The ESAC electronic songbooks, Comput. Musicol. **9**, 78 (1994)
26.100 D. Marr: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (W.H. Freeman, San Francisco 1982)
26.101 M.C. Mozer: Neural network music composition by prediction: Exploring the benefits of psychoacoustic constraints and multi-scale processing, Connect. Sci. **6**(2/3), 247–280 (1994)
26.102 J.L. Elman: Finding structure in time, Cogn. Sci. **14**, 179–211 (1990)
26.103 R.N. Shepard: Structural representations of musical pitch. In: *Psychology of Music*, ed. by D. Deutsch (Academic Press, New York 1982) pp. 343–390
26.104 M. Rohrmeier, P. Rebuschat, I. Cross: Incidental and online learning of melodic structure, Conscious. Cogn. **20**(2), 214–222 (2011)
26.105 E. Servan-Schreiber, J.R. Anderson: Chunking as a mechanism of implicit learning, J. Exp. Psychol.: Learn. Mem. Cogn. **16**, 592–608 (1990)
26.106 M. Rohrmeier: *Implicit Learning of Musical Structure: Experimental and Computational Modelling Approaches*, Doctoral Dissertation (University of Cambridge, Cambridge 2010)
26.107 M. Rohrmeier, I. Cross: Tacit tonality: Implicit learning of context-free harmonic structure. In: *Proc. 7th Trienn. Conf. Eur. Soc. Cogn. Sci. Music*, ed. by J. Louhivuori, T. Eerola, S. Saarikallio, T. Himberg, P.-S. Eerola (Univ. of Jyväskylä, Jyväskylä 2009) pp. 443–452
26.108 D.E. Rumelhart, G. Hinton, R. Williams: Learning internal representations through error propagation. In: *Parallel Distributed Processing: Experiments in the Microstructure of Cognition*, Vol. 1, ed. by D.E. Rumelhart, J.L. McClelland, PDP Research Group (MIT Press, Cambridge 1986) pp. 25–40
26.109 J.B. Pollack: Recursive distributed representations, Artif. Intell. **46**(1), 77–105 (1990)
26.110 E.W. Large, C. Palmer, J.B. Pollack: Reduced memory representations for music, Cogn. Sci. **19**(1), 53–96 (1995)
26.111 N. Boulanger-Lewandowski, Y. Bengio, P. Vincent: Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription. In: *Proc. 29th Int. Conf. Mach. Learn. (ICML), Edinburgh* (2012)
26.112 S. Cherla, T. Weyde, A. d'Avila Garcez, M.T. Pearce: Learning distributed representations for multiple-viewpoint melodic prediction. In: *Proc. 14th Int. Soc. Music Inf. Retriev. Conf. (ISMIR 2013), Curitiba* (2013)
26.113 M.T. Pearce, M. Rohrmeier: Music cognition and the cognitive sciences, TopiCS Cogn. Sci. **4**, 468–484 (2012)
26.114 A. Patel: *Music, Language and the Brain* (OUP, Oxford 2008)
26.115 A.D. Patel: Why would musical training benefit the neural encoding of speech? The OPERA hypothesis, Front. Psychol. **2**, 142 (2011)
26.116 S. Koelsch: *Brain and Music* (Wiley, Chichester 2012)
26.117 J.J. Bharucha, K. Stoeckig: Reaction time and musical expectancy: Priming of chords, J. Exp. Psychol. Hum. Percept. Perform. **12**(4), 403–410 (1986)
26.118 J.J. Bharucha, K. Stoeckig: Priming of chords: Spreading activation or overlapping frequency spectra?, Percept. Psychophys. **41**(6), 519–524 (1987)
26.119 B. Tillmann: Implicit investigations of tonal knowledge in nonmusicians listeners, Ann. N. Y. Acad. Sci. **1060**, 1–11 (2005)
26.120 E. Bigand, F. Madurell, B. Tillmann, M. Pineau: Effect of global structure and temporal organization on chord processing, J. Exp. Psychol. Hum. Percept. Perform. **25**(1), 184–197 (1999)
26.121 F. Marmel, F. Perrin, B. Tillmann: Tonal expectations influence early pitch processing: Evidence from ERPs, J. Cogn. Neurosci. **23**, 3095–3104 (2011)
26.122 E. Bigand, B. Tillmann, B. Poulin, D.A. D'Adamo: The effect of harmonic context on phoneme monitoring in vocal music, Cognition **81**, B11–B20 (2001)
26.123 E. Bigand, B. Poulin, B. Tillmann, D. D'Adamo: Cognitive versus sensory components in harmonic priming effects, J. Exp. Psychol. Hum. Percept. Perform. **29**(1), 159–171 (2003)
26.124 E. Bigand, B. Tillmann, B. Poulin-Charronnat, D. Manderlier: Repetition priming: Is music special?, Q. J. Exp. Psychol. Sect. A **58**(8), 1347–1375 (2005)
26.125 E. Bigand, B. Poulin, B. Tillmann, F. Madurell, D.A. D'Adamo: Sensory versus cognitive components in harmonic priming, J. Exp. Psychol. Hum. Percept. Perform. **29**(1), 159–171 (2003)
26.126 B. Tillmann, J.J. Bharucha, E. Bigand: Implicit learning of music: A self-organizing approach, Psychol. Rev. **107**, 885–913 (2000)
26.127 E.G. Schellenberg: Simplifying the implication-realisation model of melodic expectancy, Music Percept. **14**(3), 295–318 (1997)

26.128 W.F. Thompson, M. Stainton: Expectancy in Bohemian folk song melodies: Evaluation of implicative principles for implicative and closural intervals, Music Percept. **15**(3), 231–252 (1998)

26.129 P.T. von Hippel, D. Huron: Why do skips precede reversals? The effects of tessitura on melodic structure, Music Percept. **18**(1), 59–85 (2000)

26.130 F.A. Russo, L.L. Cuddy: A common origin for vocal accuracy and melodic expectancy: Vocal constraints, J. Acoust. Soc. Am. **105**, 1217 (1999)

26.131 A. Tierney, F.A. Russo, A.D. Patel: The motor origins of human and avian song structure, Proc. Natl. Acad. Sci. (PNAS) USA **108**, 15510–15515 (2011)

26.132 P.T. von Hippel: Melodic-expectation rules as learned heuristics. In: *Proc. 7th Int. Conf. Music Percept. Cogn.*, ed. by C. Stevens, D. Burnham, E. Schubert, J. Renwick (Causal Productions, Adelaide 2002) pp. 315–317

26.133 N. Oram, L.L. Cuddy: Responsiveness of Western adults to pitch-distributional information in melodic sequences, Psychol. Res. **57**(2), 103–118 (1995)

26.134 C.L. Krumhansl, P. Toivanen, T. Eerola, P. Toiviainen, T. Järvinen, J. Louhivuori: Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks, Cognition **76**(1), 13–58 (2000)

26.135 T. Eerola: Data-driven influences on melodic expectancy: Continuations in North Sami Yoiks rated by South African traditional healers. In: *Proc. 8th Int. Conf. Music Percept. Cogn*, ed. by S.D. Lipscomb, R. Ashley, R.O. Gjerdingen, P. Webster (Causal Productions, Adelaide 2004) pp. 83–87

26.136 J.C. Carlsen: Some factors which influence melodic expectancy, Psychomusicology **1**(1), 12–29 (1981)

26.137 M.A. Castellano, J.J. Bharucha, C.L. Krumhansl: Tonal hierarchies in the music of North India, J. Exp. Psychol. Gen. **113**(3), 394–412 (1984)

26.138 E.J. Kessler, C. Hansen, R.N. Shepard: Tonal schemata in the perception of music in Bali and the West, Music Percept. **2**(2), 131–165 (1984)

26.139 C.L. Krumhansl, J. Louhivuori, P. Toiviainen, T. Järvinen, T. Eerola: Melodic expectation in Finnish spiritual hymns: Convergence of statistical, behavioural and computational approaches, Music Percept. **17**(2), 151–195 (1999)

26.140 M. Rohrmeier, R. Widdess: Incidental learning of melodic structure of North Indian music, Cogn. Sci. (2016), https://doi.org/10.1111/cogs.12404

26.141 E.G. Schellenberg, M. Adachi, K.T. Purdy, M.C. McKinnon: Expectancy in melody: Tests of children and adults, J. Exp. Psychol.: Gen. **131**(4), 511 (2002)

26.142 C.L. Krumhansl: *Cognitive Foundations of Musical Pitch* (Oxford Univ. Press, Oxford 1990)

26.143 W.E. Caplin: *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven* (Oxford Univ. Press, Oxford 1998)

26.144 D. Deutsch: The processing of structured and unstructured tonal sequences, Percept. Psychophys. **28**(5), 381–389 (1980)

26.145 M.G. Boltz, M.R. Jones: Does rule recursion make melodies easier to reproduce? If not, what does?, Cogn. Psychol. **18**(4), 389–431 (1986)

26.146 V.J. Konečni: Elusive effects of artists' "messages". In: *Cognitive Processes in the Perception of Art*, ed. by W.R. Crozier, A.J. Chapman (North Holland, Amsterdam 1984) pp. 71–93

26.147 H. Gotlieb, V.J. Konečni: The effects of instrumentation, playing style, and structure in the Goldberg Variations by Johann Sebastian Bach, Music Percept. **3**, 87–102 (1985)

26.148 N. Cook: The perception of large-scale tonal closure, Music Percept. **5**(2), 197–206 (1987)

26.149 R.O. Gjerdingen: An experimental music theory? In: *Rethinking Music*, ed. by M. Everist, N. Cook (Oxford Univ. Press, Oxford 1999) pp. 161–170

26.150 M. Karno, V.J. Konečni: The effects of structural interventions in the First Movement of Mozart's Symphony in G-Minor K. 550 on aesthetic preference, Music Percept. **10**, 63–72 (1992)

26.151 E.W. Marvin, A. Brinkman: The effect of modulation and formal manipulation on perception of tonic closure by expert listeners, Music Percept. **16**, 389–408 (1999)

26.152 S. Koelsch, M. Rohrmeier, R. Torrecuso, S. Jentschke: Processing of hierarchical syntactic structure in music, Proc. Natl. Acad. Sci. Am. **110**, 15443–15448 (2013)

26.153 S.M. Kosslyn: If neuroimaging is the answer, what is the question?, Proc. R. Soc. B **354**, 1283–1294 (1999)

26.154 M. Mather, J.T. Cacioppo, N. Kanwisher: How fMRI can inform cognitive theories, Perspect. Psychol. Sci. **8**, 108–113 (2013)

26.155 R.A. Poldrack: Can cognitive processes be inferred from neuroimaging data?, Trends Cogn. Sci. **10**, 59–63 (2006)

26.156 C.N. White, R.A. Poldrack: Using fMRI to constrain theories of cognition, Perspect. Psychol. Sci. **8**, 79–83 (2013)

26.157 S. Koelsch, T. Gunter, A.D. Friederici: Brain indices of music processing: "Nonmusicians" are musical, J. Cogn. Neurosci. **12**(3), 520–541 (2000)

26.158 S. Koelsch, S. Kilches, N. Steinbeis, S. Schelinski: Effects of unexpected chords and of performer's expression on brain responses and electrodermal activity, PLoS One **3**(7), e2631 (2008)

26.159 S. Koelsch, B.-H. Schmidt, J. Kansok: Effects of musical expertise on the early right anterior negativity: An event-related brain potential study, Psychophysiology **39**(5), 657–663 (2002)

26.160 S. Leino, E. Brattico, M. Tervaniemi, P. Vuust: Representation of harmony rules in the human brain: Further evidence from event-related potentials, Brain Res. **1142**, 169–177 (2007)

26.161 P. Loui, T. Grent, D. Torpey, M.G. Woldorff: Effects of attention on the neural processing of harmonic syntax in western music, Cogn. Brain Res. **25**, 678–687 (2005)

26.162 A.D. Patel, E. Gibson, J. Ratner, M. Besson, P.J. Holcomb: Processing syntactic relations in language and music: An event-related potential

study, J. Cogn. Neurosci. **10**, 717–733 (1998)
26.163 N. Steinbeis, S. Koelsch, J.A. Sloboda: The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological and neural responses, J. Cogn. Neurosci. **18**(8), 1380–1393 (2006)
26.164 S. Koelsch, E. Schroger, T.C. Gunter: Music matters: Preattentive musicality of the human brain, Psychophysiology **39**, 38–48 (2002)
26.165 S.G. Kim, J.S. Kim, C.K. Chung: The effect of conditional probability of chord progression on brain response: An MEG study, PLoS ONE **6**, e17337 (2011)
26.166 P. Loui, E.H. Wu, D.L. Wessel, R.T. Knight: A generalized mechanism for perception of pitch patterns, J. Neurosci. **29**(2), 454–459 (2009)
26.167 M. Besson, F. Faïta: An event-related potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians, J. Exp. Psychol. Hum. Percept. Perform. **21**, 1278–1296 (1995)
26.168 H. Nittono, T. Bito, M. Hayashi, S. Sakata, T. Hori: Event-related potentials elicited by wrong terminal notes: Effects of temporal disruption, Biol. Psychol. **52**, 1–16 (2000)
26.169 K.A. Paller, G. McCarthy, C.C. Wood: Event-related potentials elicited by deviant endings to melodies, Psychophysiology **29**(2), 202–206 (1992)
26.170 D. Schön, M. Besson: Visually induced auditory expectancy in music reading: A behavioral and electrophysiological study, J. Cogn. Neurosci. **17**, 694–705 (2005)
26.171 R. Verleger: P3-evoking wrong notes: Unexpected, awaited, or arousing?, Int. J. Neurosci. **55**(2–4), 171–179 (1990)
26.172 J.J. Bharucha: Music cognition and perceptual facilitation: A connectionist framework, Music Percept. **5**, 1–30 (1987)
26.173 R.A. Miranda, M.T. Ullman: Double dissociation between rules and memory in music: An event-related potential study, NeuroImage **38**(2), 331–345 (2007)
26.174 S. Koelsch, S. Jentschke: Differences in electric brain responses to melodies and chords, J. Cogn. Neurosci. **22**(10), 2251–2262 (2010)
26.175 S. Koelsch, T.C. Gunter, M. Wittfoth, D. Sammler: Interaction between syntax processing in language and in music: An ERP study, J. Cogn. Neurosci. **17**(10), 1565–1577 (2005)
26.176 E. Carrus, M.T. Pearce, J. Bhattacharya: Melodic pitch expectation interacts with neural responses to syntactic but not semantic violations, Cortex **49**, 2186–2200 (2012)
26.177 N. Steinbeis, S. Koelsch: Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns, Cereb. Cortex **18**(5), 1169–1178 (2008)
26.178 A.D. Friederici: Towards a neural basis of auditory sentence processing, Trends Cogn. Sci. **6**, 78–84 (2002)
26.179 E. Fonteneau, H.K.J. van der Lely: Electrical brain responses in language-impaired children reveal grammar-specific deficits, PLoS One **3**(3), e1832 (2008)
26.180 S. Jentschke, S. Koelsch, S. Sallat, A.D. Friederici: Children with specific language impairment also show impairment of music-syntactic processing, J. Cogn. Neurosci. **20**(11), 1940–1951 (2008)
26.181 A.D. Patel, J.R. Iversen, M. Wassenaar, P. Hagoort: Musical syntax processing in agrammatic Broca's aphasia, Aphasiology **22**, 776–789 (2008)
26.182 B. Maess, S. Koelsch, T.C. Gunter, A.D. Friederici: 'Musical syntax' is processed in Broca's area: An MEG-study, Nat. Neurosci. **4**, 540–545 (2001)
26.183 S. Koelsch, T.C. Gunter, D.Y. von Cramon, S. Zysset, G. Lohmann, A.D. Friederici: Bach speaks: A cortical 'language-network' serves the processing of music, Neuroimage **17**, 956–966 (2002)
26.184 B. Tillmann, P. Janata, J.J. Bharucha: Activation of the inferior frontal cortex in musical priming, Cogn. Brain Res. **16**, 145–161 (2003)
26.185 D. Levitin, V. Menon: Musical structure is processed in "language" areas of the brain: A possible role for Brodmann area 47 in temporal coherence, NeuroImage **20**, 2142–2152 (2003)
26.186 S. Brown, M.J. Martinez, L.M. Parsons: Music and language side by side in the brain: A PET study of the generation of melodies and sentences, Eur. J. Neurosci. **23**, 2791–2803 (2006)
26.187 T.R. Knösche, C. Neuhaus, J. Haueisen, K. Alter, B. Maess, A.D. Friederici, O. Witte: The perception of phrase structure in music, Hum. Brain Mapp. **24**, 259–273 (2005)
26.188 K. Steinhauer, K. Alter, A.D. Friederici: Brain potentials indicate immediate use of prosodic cues in natural speech processing, Nat. Neurosci. **2**, 191–196 (1999)
26.189 K. Steinhauer, A.D. Friederici: Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers, J. Psycholinguist. Res. **30**, 267–295 (2001)
26.190 A. Pannekamp, U. Toepel, K. Alter, A. Hahne, A.D. Friederici: Prosody-driven sentence processing: An event-related brain potential study, J. Cogn. Neurosci. **17**, 407–421 (2005)
26.191 C. Neuhaus, T. Knösche, A. Friederici: Effects of musical expertise and boundary markers on phrase perception in music, J. Cogn. Neurosci. **18**, 472–493 (2006)
26.192 P.N. Johnson-Laird: Jazz improvisation: A theory at the computational level. In: *Representing Musical Structure*, ed. by P. Howell, R. West, I. Cross (Academic Press, London 1991) pp. 291–325
26.193 C. Francois, D. Schön: Musical expertise boosts implicit learning of both musical and linguistic structures, Cereb. Cortex **21**, 2357–2365 (2011)
26.194 I. Peretz, E. Brattico, M. Järvenpää, M. Tervaniemi: The amusic brain: In tune, out of key, and unaware, Brain **132**, 1277–1286 (2009)
26.195 M. Rohrmeier, S. Koelsch: Predictive information processing in music cognition. A critical review, Int. J. Psychophysiol. **83**(2), 164–175 (2012)
26.196 S. Koelsch: Towards a neural basis of processing musical semantics, Phys. Life Rev. **8**(2), 89–105 (2011)

26.197 U. Reich: The meanings of semantics: Comment on "Towards a neural basis of processing musical semantics" by Stefan Koelsch, Phys. Life Rev. **8**(2), 120–121 (2011)
26.198 L.B. Meyer: *Emotion and Meaning in Music* (University of Chicago Press, Chicago 1956)
26.199 E. Hanslick: *Vom Musikalisch-Schönen* (R. Weigel, Leipzig 1854), Reprint: Darmstadt 1965
26.200 L.B. Meyer: Meaning in music and information theory, J. Aesthet. Art Crit. **15**(4), 412–424 (1957)
26.201 I.H. Witten, L.C. Manzara, D. Conklin: Comparing human and computational models of music prediction, Comput. Music J. **18**(1), 70–80 (1994)
26.202 P.N. Juslin, D. Västfjäll: Emotional responses to music: The need to consider underlying mechanisms, Behav. Brain Sci. **31**(5), 559–575 (2008), discussion pp. 575–621

27. Rhythm and Beat Perception

Tram Nguyen, Aaron Gibbings, Jessica Grahn

From established musicians to musical novices, humans perceive temporal patterns in music and respond to them. There is much that we still do not understand, however, about how the temporal patterns of music are processed in the brain. Understanding the neural mechanisms that underlie processing of temporal sequences will help us learn why humans perceive the temporal regularities, or periodicities, in musical rhythms. Therefore, in this chapter, we discuss the latest findings in beat perception research, touching on both behavioral and neuroimaging findings from studies that have used electroencephalography (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and transcranial magnetic stimulation (TMS). Overall, the findings establish the importance of both auditory and motor brain areas in rhythm and beat processing. The authors also discuss the implications of beat perception research and highlight the challenges currently facing the field.

27.1 Temporal Regularity and Beat Perception

Picture yourself at a concert listening to your favorite band. The music is booming over the speakers and it engulfs the entire room. Take a look around you. From the established musicians on stage to the musical novices in the crowd, people are bobbing their heads, tapping their feet, clapping their hands, or swaying their bodies along to the music. It is evident from this picture that all humans, to some extent, have the ability to perceive the temporal patterns in music and respond to them. Although music seems to universally and automatically affect humans, there is still much that we do not know about how the temporal patterns of music are processed in the brain. Understanding the mechanisms in the brain that underlie time-keeping will help us learn why humans perceive the temporal regularities, or periodicities, in musical rhythms. The ability to perceive these temporal periodicities has been termed *beat perception* or *beat-based timing* [27.1–3]. Research on beat perception has been tackled with a variety of methodologies in both humans and animals, including behavioral work and brain imaging techniques such as electroencephalography (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and transcranial magnetic stimulation (TMS). In this chapter, we will consider the latest findings in beat perception research, touching on both behavioral and neuroimaging findings. We will also discuss the implications of this research and highlight the challenges facing the field today.

Understanding the structure of temporal patterns in our environment is crucial for many aspects of perceptual, cognitive, and motor function. For example, humans rely on accurate temporal perception and production for normal hearing, speech, motor control, and music [27.4, 5]. In fact, the ability to perceive temporal patterns in our environment is evident even

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_27

in infancy [27.6–8]. Some characteristics of temporal pattern perception appear to be unique to humans, including our receptiveness to musical rhythm and more specifically, our ability to perceive the beat [27.9, 10]. As many of the words associated with rhythm and beat have different meanings in different fields, we will define how they are used in this chapter. A musical *rhythm* is a pattern of temporal intervals in a stimulus sequence (Fig. 27.1). The rhythmic pattern is indicated by the onsets of a stimulus, usually a tone, a click, or some other sound. The length of the temporal intervals in the sequence is defined by the time between onsets, termed interonset intervals. Listening to a musical rhythm often gives rise to a sense of *pulse*, or sometimes referred to as a *beat*. The beat is a series of regularly recurring, salient psychological events [27.1, 11]. The beat is a psychological event because it is not stimulus driven, even though it usually arises in response to a rhythmic stimulus [27.12–15]. The psychological internalizing of the beat is why we can sense it even when music is rhythmically complex or has notes occurring off the beat.

Individual beats are frequently perceived to possess different degrees of accents, or stress, which gives rise to meter. *Meter* (or metrical hierarchy) is the grouping or temporal organization of beats, in which some beats are perceived as more salient than others [27.13, 16]. For example, in a march rhythm, every other beat is accented (**1** 2 **1** 2), whereas in a waltz it is the first of every three beats (**1** 2 3 **1** 2 3). These patterns therefore differ in their perceived meter: the 1's are *strong* beats and the others are *weak* beats. Depending on the metrical structure, feeling the beat is harder in some musical rhythms than others [27.17, 18]. Different researchers also define structural complexity differently, but one common definition involves the presence of simple integer ratios between intervals in a sequence [27.19, 20]. For example, a sequence containing intervals of 250, 500, 750, and 1000 ms has a 1 : 2 : 3 : 4 simple integer ratio relationship between its intervals. Simple integer ratios are not the only factor used to defined complexity. Another factor is whether there is an onset of a tone occurring at regularly spaced intervals, generally between 300–900 ms [27.21]. Using these two factors, researchers have created rhythms of different complexity [27.22, 23]. For example, *metric simple* rhythms have intervals that are related by simple integer ratios. However, integer ratios alone are not sufficient to induce a beat. Regular occurring *perceptual accents* or cues may also be necessary [27.24]. Therefore, the intervals are also arranged so that an interval onset occurs every four units (e.g., **4 3** 1 **1** 1 2), creating a perceptual accent that induces perception of a regular beat at those times [27.25]. For *metric complex* rhythms, the intervals are arranged so that they are not reliably grouped into having an onset every four units (e.g., 2 1 3 2 1 4 1). Therefore, metric simple rhythms induce a stronger percept of a beat than metric complex rhythms [27.22]. In music, the beat is often emphasized by nontemporal accents or cues such as pitch, timbre, or loudness [27.26–28]. For example, changes in loudness (or intensity) cause louder notes to be perceived as more prominent. Perceiving the beat through changes of loudness requires very little effort. In fact, the beat can be perceived even if one is barely attending to it. However, even rhythmic patterns without changes in nontemporal accents, can induce listeners to feel the beat. In this case, any perceptual accents that occur are due to the temporal accents rather than the nontemporal accents [27.29].

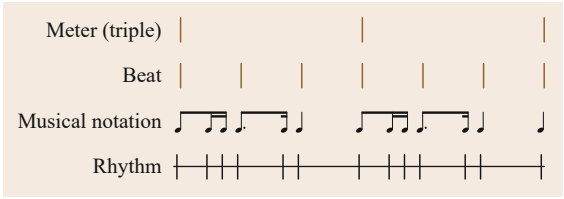


Fig. 27.1 Illustration of a musical rhythm and its associated metrical hierarchy. Rhythm (*bottom*) shows the waveform of a rhythmic sequence. The *vertical lines* show where notes occur and the *horizontal line* indicates time between note onsets. Musical notation (*second from the bottom*) of the same rhythmic pattern is shown. Beat locations (*third from the bottom*) in the rhythm are shown. Sounds under the *light brown beat lines* are perceptually accented and are perceived as louder, or more salient, than the other sounds. Meter (*top*) shows a metrical organization of the beats in the rhythm. Beats under the *light brown metrical lines* are perceived as more salient than other beats

27.2 Behavioral Investigations

Beat perception in humans can be studied using *sensorimotor synchronization* (SMS), a technique where a movement (motor), such as the tap of a foot, is temporally synchronized with a predictable external event (sensory), such as a beat [27.30]. Humans have the unique ability to move in (and out of) phase with a beat using different body parts and it is this coupling of the sensory and motor systems that has made SMS a popular technique to study beat perception [27.31]. Although SMS is a skill that is refined in many musicians, even

musical novices are able to move in synchrony with the beat [27.32], yet another reason why SMS is a practical technique. Beat-finding studies using SMS often focus on synchronizing finger tapping to auditory sequences that consist of tones or clicks. In many beat perception studies, synchronization is less accurate when the beat is more difficult to perceive [27.31]. For example, synchronizing to rhythms with a structured timing pattern is easier than synchronizing to rhythms with an unstructured timing pattern [27.33], suggesting that metrical structure influences the precise timing of an individual's movement [27.2]. Likewise, rhythms with complex interval ratios are more difficult to synchronize to than rhythms with simple interval ratios [27.34]. Metrical levels, which are regular time points at integer multiples or fractions of the basic (or referent) beat level, can also influence an individual's ability to synchronize with the beat [27.26]. For example, although synchronization is better with musical pieces that are temporally regular compared to musical pieces that are temporally less regular and thus more expressive, tapping to expressive pieces occurred at a higher metrical level than the less regular pieces. Therefore, the timing variation in expressive music reduces synchronization accuracy, but seems to aid the recovery of the music's metrical structure [27.35].

Synchronizing to auditory sequences of tones or clicks differs from synchronizing to music. Unlike the auditory sequences created in the laboratory, music is usually more variable and contains information about tempo, harmony, pitch, etc. There is evidence that melodic, harmonic, and pitch structure influence beat perception [27.36–38]. For example, synchronizing to expressively timed music is easier than synchronizing to metronomic sequences with the same timing pattern [27.33]. However, one study found that removing melodic and harmonic information had little effect on pianists' ability to find the beat in ragtime music [27.39]. However, when participants heard only the right-hand part, not the left-hand part, tapping the beat was more variable and less accurate. The authors found the following factors were important for accurate tapping to the beat: a predictable alternating pattern in the left-hand part and a majority of notes on metrically strong positions (on the beat) in both the right-hand and left-hand parts. Thus, metrical structure has a strong influence on beat perception. Rhythms with strong metrical structures often correspond with high temporal expectations. Therefore, synchronization accuracy is high [27.40, 41].

For the most part, beat perception is believed to be universal in humans. However, there are factors that give rise to large individual differences. Musical training or experience is one such factor. Many laboratory studies of SMS have shown that musical experts are better at synchronizing to the beat (producing smaller asynchronies between taps, smaller variability and tapped to a wider range of metrical levels) than musical novices [27.32, 35, 42]. Performance on other rhythm tasks is also improved by musical training [27.43, 44]. For example, musically trained individuals are better at detecting changes in rhythms than musically untrained individuals [27.35, 45, 46], regardless of whether the rhythms are metrically simple or complex [27.35]. This improvement may be because formal lessons emphasize explicit learning of many complex rhythmic structures, allowing musicians to better parse both metrically simple and complex rhythms [27.43]. Another possibility, however, is that musical training enhances sensitivity to underlying temporal structures that make the beat more salient.

SMS has allowed researchers to examine many aspects of beat-finding behavior within a single task. These aspects include, but are not limited to, the variability of tapping and the deviations of taps from the beat [27.30]. Another advantage of the technique is that it provides researchers with the opportunity to study the interaction between the perceptual systems and the motor systems. However, this advantage is also a drawback for many studies using SMS because it can be challenging to separate the perceptual processes from the motor processes [27.47]. Thus, it is still unclear what cognitive principles govern how humans perceive the beat and what mechanisms might underlie these principles. The use of brain imaging techniques attempts to resolve some of these questions.

27.3 Electrophysiological Investigations

Electroencephalography (EEG) or magnetoencephalography (MEG) are examples of nonbehavioral techniques used to study beat perception. An advantage of electrophysiological techniques is that they can provide temporal resolution at the millisecond level [27.48, 49]. Electrophysiological methods work because communicating neurons send and receive signals through the movement of either positive or negative ions (depending on whether the signal is excitatory or inhibitory). The movement of ions creates a current flow, which results in an electrical potential (i. e., an unequal distribution of electrical charges). Individually, these po-

tentials are too small to be detected, but when numbers of neurons in the outer cortical layers all function in synchrony there is a summative effect of individual electrical potentials that gives rise to the signal detected by surface EEG readings at the scalp.

27.3.1 Electroencephalography

A common use of EEG is to record the brain responses to particular events (e.g., on-beat and off-beat notes in a rhythm) and compare them to look for differences. The brain response is recorded as a waveform that shows the changes in strength of electrical potentials over time (Fig. 27.2). Responses that are triggered by a specific event are called event-related potentials (ERPs). To obtain ERPs, waveforms from each electrode are averaged over many presentations of the same stimulus. Many presentations are required to overcome the large amount of electrical activity in the brain that is unrelated to the task of interest. The electrical *interference* from this superfluous activity is also captured in the EEG signal as noise, but the noise is averaged out when the large number of presentations is combined (Fig. 27.3). This time-locked, averaged response is called an *evoked response* because it is the average response evoked by a stimulus.

One question in regard to beat perception has been whether the perception of metrical structure requires attention, or is perceived even without attention. Two EEG studies have been successful in using evoked responses to test for distinguishable neural components (positive or negative *peaks* in the waveform of the evoked response) related to rhythm and meter processing. In the first, *Geiser* et al. [27.50] measured changes in ERP responses to alterations of metrical organization (shifting the metric structure from 3/4 to 5/8 or 7/8 by adding or omitting an 8th note in a repeating rhythm) and rhythmic patterns (replacing one long note with two faster ones). Participants either attended to the rhythm and meter directly (pressing a key when rhythmic/metrical changes were detected) or attended and responded to unrelated pitch changes. In line with previous work, other researchers found that changes to the rhythm elicited a negative ERP component between 100 and 150 ms after the perturbation, regardless of attention [27.51, 52]. Perturbations to the metric structure, however, only elicited a negative ERP when attended. The fact that negative ERPs were observed for rhythmic violations regardless of attention, but to metrical violations only when attended, suggests that processing of meter is a more complex, attention-demanding process than processing of rhythm. One caveat, however, is that the metrical violations used in these studies were more difficult to detect than the rhythmic violations. Therefore, the lack of negative deflection to meter changes in the pitch-detection condition may indicate that these particular metric violations were difficult and required attention to detect, rather than *all* metric encoding requiring attention.

On the contrary to previous findings attention may not be necessary to encode metrical structure [27.53]. Participants heard repeating rhythms, but rather than changing the rhythms by adding or subtracting notes similar to the manipulation in the *Geiser* et al. [27.50] study, certain notes in the rhythm were omitted entirely and there was silence where the note should have been. The omissions occurred on notes that were either metrically strong (e.g., downbeats), or metrically weak. The beauty of comparing omissions on strong and weak metric positions is that both omissions are acoustically identical (silence is silence), but if participants had a metrical representation of the rhythm then omission of metrically strong notes should produce a greater response than omissions of metrically weak notes. To manipulate attention, participants were asked to either monitor the rhythm and indicate when an omission occurred, or to monitor a stream of white noise presented at the same time as the rhythm and indicate when the noise intensity changed. Metrically strong omissions elicited a larger brain response compared to metrically weak omissions, as measured by an ERP component called the mismatch negativity (MMN). The MMN was also larger for strong omissions in the white noise condition, when participants were not attending to the rhythm but rather to the white noise stream, suggesting that, encoding of metrical structure does not require attention.

27.3.2 Magnetoencephalography

Another method for investigating the neural components of beat perception is to use MEG rather than EEG. Both EEG and MEG measure signals from the movement of ions into and out of neurons, but they do so in different ways. Rather than detecting the electrical changes like EEG, MEG detects the changes in magnetic fields [27.54] caused by the current flow of the ions. The magnetic fields produced by even thousands of synchronously firing neurons are very small, so an array of hypersensitive sensors is required to detect the changes (Fig. 27.4). MEG measures the average magnetic field strength at each sensor positioned around the head, over time. Similar to event-related potentials in EEG, average stimulus-locked waveforms can be generated from the MEG output. These waveforms, called event-related fields (ERFs), can be analyzed and interpreted much the same as ERPs. ERF components that are equivalent to ERP components are even named accordingly, with ERF components given an “m” suffix

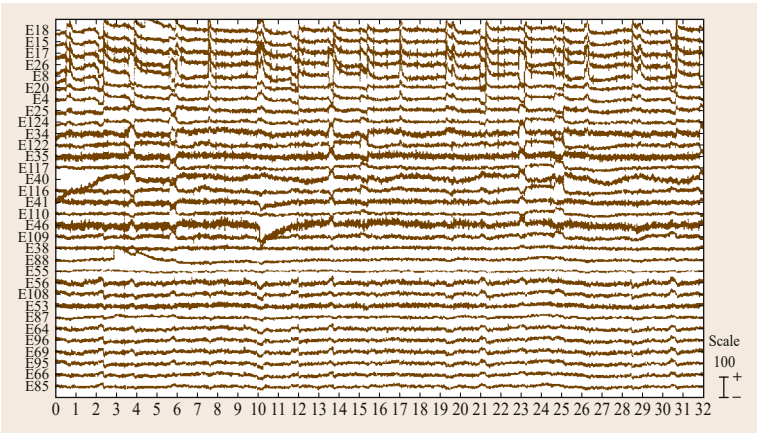


Fig. 27.2 Sample of electroencephalography (EEG) waveforms. Each line, called a trace, shows the changes in scalp-recorded electrical activity at a particular electrode (*y*-axis) recorded over time (*x*-axis)

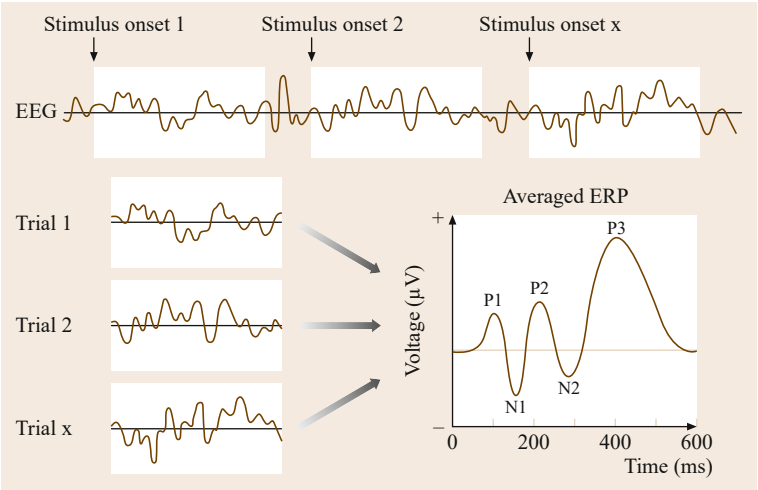


Fig. 27.3 Simplified illustration of averaging electroencephalography (EEG) signal over many presentations to reduce noise in the signal. (*top*) An example of an EEG trace over time from a single electrode, with stimulus onsets and trial epochs marked by the *white boxes*. The data outside the *white boxes* are removed leaving only data collected during the trial epoch (time-locked to the stimulus onset) in the analysis. (*bottom*) Trial epochs are then averaged together to remove noise, leaving a smoothed event-related potential (ERP)

to differentiate that they are magnetic field peaks in MEG (e.g., a mismatch negativity (MMN) is called an MMNm when recorded with MEG).

A recent study used MEG to study changes in ERF responses to increasingly salient violations of rhythmic expectancy [27.55]. One of the key consequences of listening to rhythm is the setting up of expectancy. We make predictions about how the rhythm will continue based on what we have already heard and we expect the events occurring *on-beat* (e.g., the *downbeat*) to be more salient than events occurring *off-beat*. *Vuust* et al. [27.55] had participants listen to repeating rhythms in which there were occasional violations of rhythmic expectancy. In line with the researchers' predictions, when the rhythmic violations occurred, two neural components were observed in the ERFs: an MMNm, associated with expectancy violation detection and a P3am, associated with the internal evaluation of the rhythmic sequence. The components were strongest for the condition with the largest violation of expectancy. Moreover, consistent with behavioral [27.15, 56] and ERP work [27.57, 58], musicians showed a greater sensitivity than nonmusicians to violations, with larger, earlier MMNm peaks. The authors suggest that musicians have a better internal representation of the metrical structure, which allows them to make more precise predictions about the incoming stimuli. Therefore, even small violations elicit stronger, more rapid responses when those precise predictions are violated.

27.3.3 Oscillatory Activity and Auditory Steady-State Response

When perceiving a beat, we form an expectation about when the beat will occur. ERPs and ERFs have proven useful in detecting violations of beat expectations, but one does not have to use violations to measure expectancy based on beat perception. Another method is to study the oscillations of populations of neurons in response to rhythmic stimuli. Oscillatory activity in pop-

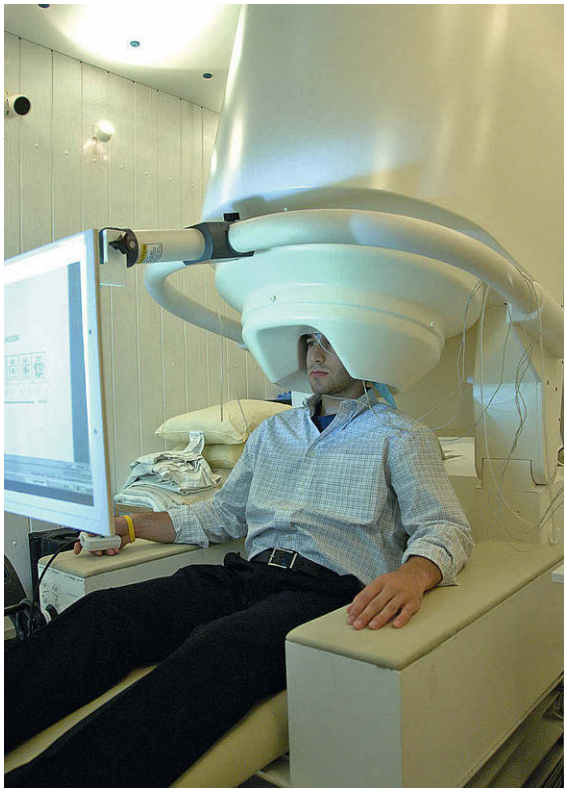


Fig. 27.4 A magnetoencephalography (MEG) machine. Participants sit comfortably in the chair while the large array of sensors, housed in the large white cylinder on the top of the machine, are lowered onto their head. The sensors detect very small changes in magnetic fields caused by neural firing (Photo: National Institute of Mental Health, National Institutes of Health, Department of Health and Human Service)

ulations of neurons arises from feedback connections between the neurons that lead to the synchronization of their firing patterns. Neural oscillations can occur at different rates, or frequency bands: one well-known example is the alpha band ($\approx 8-12$ Hz), associated with relaxation. Neural oscillatory responses form the basis of an emerging theory of beat perception called resonance theory [27.11, 59, 60]. The basic idea of resonance theory is that subpopulations of neurons entrain to (or fire in synchrony with) an incoming auditory stimulus, for example, a rhythm. As more neurons fire in synchrony with the rhythm, the notes of the rhythm that occur at the times of peak firing may be perceived as more salient and are processed more accurately. The increased neural firing that is in synchrony with the incoming stimuli could explain how feelings of pulse and meter arise, which is something that other theories struggle to explain [27.61].

Investigations of oscillatory processes related to rhythm and timing have found changes in both the gamma (> 30 Hz) and beta ($13-30$ Hz) frequency bands. Gamma band oscillations have been associated with anticipation [27.62] and attention [27.63]. Oscillations in the beta band have been strongly associated with motor tasks and have been observed in the sensory and motor cortices [27.64]. In studies of rhythm, beta and gamma band activity have been linked to the anticipation of tones [27.65, 66].

A study by *Snyder* and *Large* [27.66] measured gamma band oscillatory brain responses to examine rhythmic expectancy and metric encoding when listening to an isochronous tone sequence. The researchers accented (made louder) every other tone in the sequence to induce perception of a duple meter (**1** 2 **1** 2, etc.). The researchers found a greater evoked neural oscillatory response for strong beat (accented) tones compared to weak beat tones. This finding was consistent with findings in ERP studies [27.50, 53]. Moreover, in line with neural resonance theories of timing, they found increases in induced oscillatory activity when the omitted strong beat tone *should* have occurred. The higher induced gamma activity (even in the absence of a tone) was interpreted as evidence of the participant's expectancy of a strong beat. Thus, their meter perception could be measured by the neural oscillations. As metrical perception and attentional processes are both linked to gamma band activity, the findings may support the hypothesis that detection of metrical violations [27.50] requires attention.

Extending the findings of *Snyder* and *Large* [27.66], *Iversen* et al. [27.65] used MEG to examine oscillatory neural responses evoked by a repeating rhythmic pattern consisting of two tones followed by a rest. Listeners were instructed to impose different metrical interpretations on the rhythm by placing mental accents on either the first or second tone. The early evoked response, specifically in the upper beta range (20–30 Hz), was stronger for the beat that was mentally accented. A second experiment established that the beta band increase was very similar to the increase seen when tones were physically (rather than mentally) accented by being made louder. Increased beta band activity (which is typically linked to motor control/movement) may parallel fMRI findings that activation in motor networks correlates with metrical interpretation of rhythm, even in the absence of movement [27.22, 67], although strong conclusions cannot be drawn because the relationship between neural synchrony and fMRI activation is only beginning to be investigated [27.68, 69].

In addition to increases in beta and gamma band oscillatory activity, recent studies have found increases in

a lower frequency band called the delta band (< 4 Hz). Delta band activity is generated by populations of neurons firing at the same rate as (i. e., resonating with) a periodic stimulus [27.70, 71], in this case: the beat of a rhythm. This low-frequency resonance can be captured with EEG by measuring steady-state-evoked potentials (SSEPs). A pair of studies by *Nozaradan* et al. [27.70, 71] has found that the SSEP response is larger at the frequency of the beat rate when participants listen to rhythm, suggesting a greater number of neurons are firing at the time the beat occurs. In the first study, participants were asked to mentally accent either every other (**1** 2 **1** 2, binary meter), or every third (**1** 2 3 **1** 2 3, ternary meter) tone in an isochronous tone sequence. The researchers found that the SSEP was larger at the frequency of the imagined-to-be-accented tones for each condition (Fig. 27.5). These findings indicate two things: that participants can impose a perceptual metric interpretation on isochronous tone sequences and that entrained neural responses occur at the rate of the imposed interpretation. A second study extended these findings with short, repeating, syncopated rhythms instead of isochronous tones. The rhythms induced the perception of a regular beat but didn't necessarily have a note occurring on every beat. The participants still *felt* a beat, even when the beat occurred when no note was being played. Therefore, if an enhancement of the SSEP response was seen at the beat rate, the researchers could conclude that this enhancement was, at least in part, driven by perception of a beat, rather than being only driven by the notes in the rhythmic stimulus. This is indeed what the researchers found. Taken together, the findings of these different studies provide compelling evidence that enhancements of neural synchrony measured in the delta, beta, and gamma bands could be neural markers of our perception of pulse and meter.

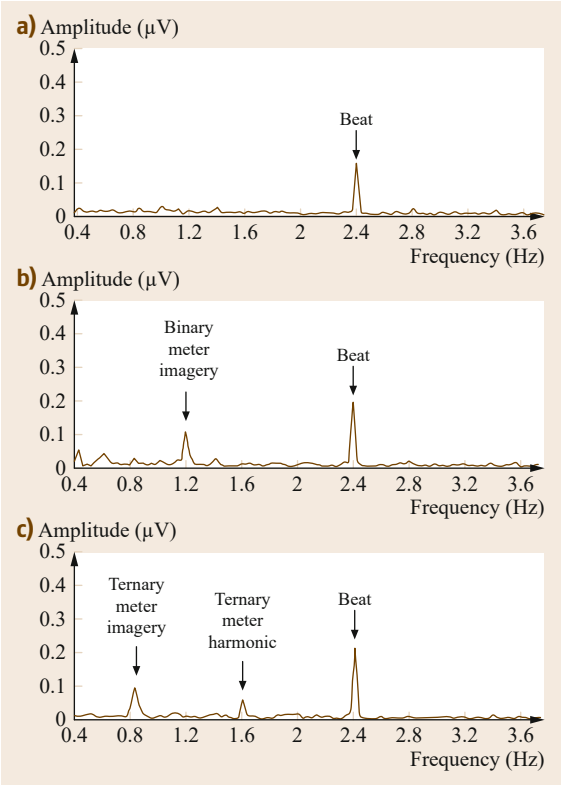


Fig. 27.5a–c Perceived meter is evident in neural response. In the control condition **(a)**, participants heard a stimulus with an evident 2.4 Hz beat. In the binary meter imagery condition **(b)**, participants heard the same stimulus as the control condition, but imagined an accent on every second beat. The *peak* at 1.2 Hz in the binary meter condition reflects the perceived beat rate (half as fast as the control). In the ternary meter imagery condition **(c)**, participants heard the same stimulus, but imagined an accent on every third beat. The *peak* at 0.8 Hz (and its harmonic at 1.6 Hz) reflects the perceived ternary meter (a third as fast as the control). As the stimulus was always the same, the different EEG responses between conditions reflect changes in the imagined beat rate

27.3.4 Limitations of Electrophysiological Methods

Although EEG and MEG are very useful techniques for investigating the neural response to beat and rhythm, these methodologies are not without their limitations. For example, it is difficult with EEG to tell *where* in the brain the activity is occurring. The skull, membranes covering the brain, cerebrospinal fluid, and even other neurons can *blur* the signal, obscuring the true intracranial source of the signal. The simplest model of activity is that the signal is coming from a single source, called a *dipole*. The goal of *source localization* is to determine, for each point in time, where in the head the dipole is, what its orientation is, and how strong it is (the dipole moment). The difficulty in determining these things is that an infinite number of combinations of intracranial sources could give rise to a particular signal distribution on the scalp; this is what's known as the *inverse problem* [27.72]. In essence, trying to determine the source of an EEG signal based on its distribution on the scalp is akin to trying to reconstruct an unknown object (or set of objects) using its shadow as your only reference. To narrow down the number of possibilities it is necessary to use complicated modeling algorithms and information from other neuroimaging techniques (e.g., PET or fMRI) about possible activation centers (or *seeds*) for dipole locations [27.73, 74].

MEG is slightly better for source localization because magnetic fields are not distorted as much by bone and tissue. However, magnetic fields decay more quickly as a function of distance than electric fields so they can only be recorded from the surface of the cortex. This means that MEG signals have even less contribution from deeper sources than EEG signals [27.49]. To see deeper into the brain and to obtain better spatial resolution, we need to use alternative neuroimaging techniques.

27.4 Hemodynamic (fMRI/PET) Investigations

Techniques such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) are better at indicating where in the brain the activity is occurring. The hemodynamic approach offers fMRI superior spatial resolution that cannot be accomplished with EEG or MEG. The approach relies on the principles of neurovascular coupling. Simply put, when a brain region is more active, it consumes oxygen, thus requiring increased blood flow to the region to replenish the depleted oxygen. The increase in oxygenated blood leads to a magnetic signal variation, known as the blood-oxygen-level-dependent (BOLD) response, which can be detected using an MRI scanner [27.75–77] (Fig. 27.6). Therefore, fMRI works by detecting changes in blood oxygenation related to increased regional blood flow [27.78]. The change is detectable around 2–3 s after the start of neuronal activity. In comparison to electrophysiological methods (EEG and MEG), hemodynamic methods (fMRI and PET) have poorer temporal resolution, but the trade-off is better spatial resolutions.

Fig. 27.6 A functional magnetic resonance imaging (fMRI) machine. Participants lie on the bed, which is moved until their head is in strong magnetic field in the bore, the small hole in the middle of the round structure. The coil (small round cage-like structure on the bed, close to the magnet) is used to detect changes in the blood-oxygen-level-dependent (BOLD) signal, which denotes brain activity (Photo: Siemens 3-Tesla Prisma MRI scanner with 64-channel head coil, housed at the Centre for Functional and Metabolic Mapping at The University of Western Ontario, London, Ontario, Canada)

Several studies have used fMRI to investigate how the brain responds during beat perception and rhythm perception and production. These studies generally show activation in several brain areas traditionally associated with movement: the premotor cortex (PMC), the supplementary motor area (SMA), the cerebellum, and the basal ganglia [27.22, 79–82] (Fig. 27.7). These motor areas are active during rhythm listening even when no movement is involved, suggesting that the brain areas associated with the control of movement are involved in the perception of rhythms [27.22]. To isolate the brain areas that respond specifically to the beat, the neural activity to beat-based rhythms (e.g., metric simple rhythms, or other rhythms that induce beat perception) and nonbeat-based rhythms (e.g., metric complex rhythms, or rhythms that do not induce beat perception) are compared. The basal ganglia, the SMA, and the superior temporal gyri (STG) generally respond more to beat-based than nonbeat-based rhythms, whereas other motor areas (the PMC and the cerebellum) respond similarly to both rhythm types [27.22, 67, 83]. Similar findings are shown for simpler regular (isochronous) and irregular auditory sequences. Regular tone sequences generate greater basal ganglia activity than irregular tone sequences [27.84].

Activity in the motor areas during beat perception has led to an emerging perspective that beat perception relies on the interaction between the auditory and the motor systems. That is, the basal ganglia exhibits greater communication, or coupling, with other motor areas during beat perception [27.67]. Specifically, beat perception leads to increased coupling between the basal ganglia and the SMA and the PMC [27.67]. A network including the basal ganglia, SMA, and PMC responds when participants need to internally predict or generate a beat in a rhythm. Therefore, the ability to tap along to the beat may be facilitated by this network, which enables predictions to be formed about when the beat will occur [27.67]. Greater connectivity is also shown between the putamen and the ventrolateral prefrontal cortex (VLPFC) when synchronizing finger taps to the beat, suggesting that the basal ganglia play an important role in beat perception by interacting with auditory work-

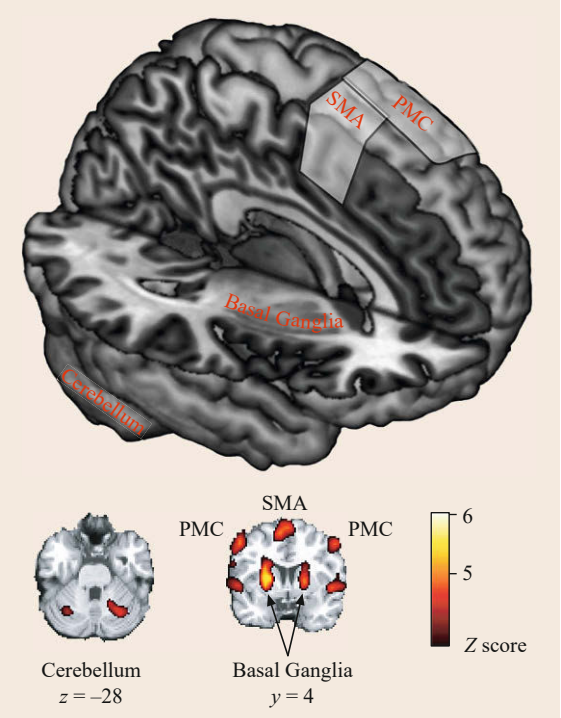


Fig. 27.7 Functional magnetic resonance images of motor areas that are active when listening to rhythms. (*top*) A *cut away* anatomical image with labels for the motor regions involved in beat perception: the premotor cortex (PMC), the supplementary motor area (SMA), the cerebellum, and the basal ganglia. (*bottom*) Functional images showing increased activation in motor areas when participants listened to rhythms, but did not move to them. The *bottom left image* is a horizontal slice (nose would be at the top) showing increased activation in the cerebellum. The *bottom right image* is a coronal slice showing increased activation in the SMA, as well as bilaterally in the PMC and basal ganglia

ing memory mechanisms in the inferior frontal cortex [27.85].

fMRI has also been used to investigate individual differences in rhythm and beat perception. These differences might correlate with the communication between auditory and motor areas. Musical training or experience appears to influence auditory-motor communication. Although both musicians and nonmusicians show auditory-motor coupling between left auditory and premotor cortex, only musicians show significant coupling in the right hemisphere [27.44]. Other work has found greater auditory-motor coupling for musicians compared to nonmusicians, but this time in both hemispheres [27.67]. Increased coupling between the STG and the PMC might be important for integrating auditory perception with motor production [27.44, 67]. Musical training may lead to greater integration because of the extensive practice at using auditory feedback to alter motor production.

Interestingly, basal ganglia activity does not appear to correlate with the speed of the beat, instead showing maximal activity around 500–700 ms, the beat rate that humans find ideal [27.86, 87]. Findings from behavioral work that used both musical and nonmusical stimuli suggest that beat perception is maximal at a preferred beat period near 500 ms [27.21, 60, 88]. Therefore, the basal ganglia are not simply responding to any perceived temporal regularity in the stimuli, but are most responsive to regularity at the frequency that best induces a sense of the beat.

Further evidence for the basal ganglia role in beat perception comes from cross-modality research of rhythms. Beat perception is not necessarily restricted to the auditory modality. There is evidence that it can also exist in the visual modality [27.89], but in many beat-based timing tasks the auditory system consistently outperforms the visual system [27.90–92]. However, when visual rhythms are presented after beat-based auditory rhythms with the same pattern, the beat can subsequently be induced in the visual rhythm [27.89]. One fMRI study showed that activity in the basal ganglia during visual rhythm presentation significantly predicted whether the visual rhythms induced a beat or not [27.93], suggesting that the basal ganglia are not only responsible for beat perception in the auditory modality, but may be responsible for beat perception in the visual modality as well.

27.5 Patient and Brain Stimulation Investigations

The disadvantage of fMRI is that it cannot be used to tell us whether an activated brain area is *necessary* for a given process to occur, only that it responds when that process is being carried out. Data from patients with impaired brain function can establish causal relationships. Thus, patients with impaired basal ganglia function can clarify whether the basal ganglia play a causal role in beat perception. Parkinson's disease (PD) is one example of a neurological disorder that affects basal ganglia function. PD results from death of dopaminergic neurons that project to the basal ganglia, therefore affecting normal basal ganglia functions,

leading to complications with movements. Therefore, if the basal ganglia are essential for normal beat perception, healthy controls should be better at discriminating beat-based rhythms than individuals with PD. This prediction is supported; in PD patients, the ability to detect changes in beat-based rhythms is significantly impaired compared to healthy controls [27.94]. The controls were significantly better at discriminating beat-based rhythms than nonbeat-based rhythms, but PD patients did not benefit to the same degree from the presence of the beat. These findings support the fact that the basal ganglia have a pivotal role in beat perception.

Brain stimulation techniques can also establish whether a brain area is necessary for a particular cognitive function. Transcranial magnetic stimulation (TMS) is a noninvasive stimulation technique that excites neurons in the cortex with a rapidly changing magnetic field placed at the surface of the scalp. A rapidly changing current is run through the TMS coil (Fig. 27.8), which generates rapidly changing magnetic fields around the coil. These magnetic fields subsequently generate weak electric currents in the underlying tissue. TMS is a technique that disrupts neuronal processing; making it ideal for determining whether a brain area is necessary for a particular task or function. However, it is only ideal for stimulating the cortical surface. Therefore, deeper structures such as the basal ganglia may be out of reach, but other motor areas, such as the cerebellum or premotor cortex can be tested. TMS work in the domain of beat perception is emerging [27.95–99]. A study on healthy controls found that after the application of TMS over the medial cerebellum, duration-based timing was significantly impaired [27.96]. A follow-up study using fMRI was able to support these findings, suggesting that motor areas are implicated in beat perception [27.83]. Although beat perception research has advanced quickly, there are still many unanswered questions and unresolved issues, but with new techniques (or novel combinations of current techniques), the future of beat perception research is looking upbeat.

Fig. 27.8 Example of a transcranial magnetic stimulation (TMS) experiment. The experimenter uses a TMS coil to stimulate the participant's premotor cortex (PMC). TMS is a non-invasive stimulation technique used to excite neurons in the cortex by creating a rapidly changing magnetic field at the surface of the scalp

27.6 Discussion

A wide variety of computational approaches have been used to model timing and rhythm perception. Although the findings of recent neuroscientific investigations support some of the predictions of neural resonance theory there is still no consensus about which model is best. Both EEG and MEG have found neurological markers of anticipation and representations of metric structure, especially in beta and gamma bands and SSEPs appear to mark beat and meter perception in the delta band. Beat perception may rely on automatic, preattentive processes, in contrast to metrical encoding which may require attention (although this is not certain). Neuroimaging studies have found evidence of specific auditory-motor networks of neural areas that support beat perception, arguably a crucial process for musical rhythm perception.

While it is true that a great deal has been learned from neuroscientific research techniques, the methodologies have drawbacks as well. Two major criticisms of neuroscientific research are the unnecessary expense of the methodologies and the current inability to distinguish between psychological models [27.100, 101]. Many of the techniques also require participants to remain virtually motionless for long periods, sometimes in very noisy environments (e.g., the bore of an fMRI). In addition, the relative novelty of the investigative techniques and increasing complexity of preprocessing and statistical analysis of the data may lead to flaws in the analysis being overlooked or unrecognized. For example, it is difficult for a researcher to remain familiar with all the details of the variety of statistical procedures that correct for multiple comparisons across brain regions in fMRI, or time windows in EEG/MEG.

Despite these issues, neuroscientific techniques have some undisputable advantages over behavioral techniques. One advantage is that it provides another

independent variable to test: the brain's response can provide valuable information even if an overt behavioral response cannot be measured. A study by *Winkler* et al. [27.102] illustrates this advantage nicely. The researchers investigated whether the ability to *feel the beat* was present in newborns. To test this, infants (2–3 days old) listened to rhythms while their brain activity was recorded using EEG. Occasionally part of the rhythm was omitted; sometimes this omission was on a beat, other times it was off the beat. If newborns have beat perception, the omissions of beats should be more salient than the omissions of off-beats. Obviously, the researchers could not ask infants to make a behavioral response, but changes observed in the EEG recordings could be used instead. A clear difference in the EEG responses to omissions and off the beat was detected, which the experimenters interpreted to mean that beat perception might be innate. As a caveat, the interpretation of these findings may or may not prove to be true and depend heavily on one's specific definition of the word *innate*; infants can hear in the womb from as early as six months after conception and this prenatal learning could potentially contribute to the effect. Even prenatal dancing by the mother may need to be taken into account, as there is evidence that infants' rhythm perception is influenced by being *bounced* in time with music [27.103]. At the very least, findings that rhythm perception can be influenced by culture suggests that if the innate ability does exist it is also shaped by subsequent experience [27.8, 65, 104, 105]. Interpretation of the findings aside, the *Winkler* et al. [27.102] study does nicely illustrate how neuroscientific methods can provide meaningful information in the absence of behavioral response.

Another important use of neuroscientific methods is when behavioral results do not allow for two competing models to be distinguished. One example of this is how electrophysical research has led to entrainment models (timing in relation to a steady pulse) gaining evidence, in addition to the more traditional interval-based models (timing of discrete events) of timing [27.3, 106–108]. Interval-based timing theories have the advantage of parsimony: many things that we time have no regular beat, so if a beat-based timing system were to exist, it would be in addition to an interval-based system. The existence of an additional beat-based system would be justified, however, if it provided more accurate timing. While some studies show increased temporal accuracy in beat-based timing tasks [27.3, 109] others have failed to show this advantage [27.108]. *Grahn* and *Brett* [27.22] designed a study to reconcile these conflicting findings. In a two-part behavioral and fMRI study, there was an advantage for reproducing beat-based rhythms over nonbeat rhythms. In the fMRI portion, using the same stimuli, but in a discrimination task that equated performance between the two rhythm types, the researchers found increased activation in a specific network of areas (including the basal ganglia) when perceiving beat-based rhythms, even though no behavioral advantage was present. This showed that a beat-based mechanism could be active, even when performance improvements were not observed. Thus, even when behavioral results do not distinguish between the predictions of beat-based and interval-based models, neuroscientific methods can indicate whether or not the brain is using the same mechanisms. Adding to the support for entrainment theories are recent EEG findings showing entrainment of neural populations to the beat [27.70, 71] and oscillatory synchronization in the beta and gamma bands [27.65, 66] discussed earlier.

Ultimately, for the field to move beyond measuring neural correlates of rhythm and beat perception and simple localization of cognitive processes to brain areas, the delineation of specific neurobiological timing components and mechanisms is required. This will rely on greater integration between neurosciences and modeling is crucial. As techniques for data acquisition and analysis advance, so does the richness of neuroscientific data. Continued interdisciplinary communication and collaboration promises to yield further advances in our understanding of how rhythm and beat perception comes about.

References

27.1 G. Cooper, L.B. Meyer: *The Rhythmic Structure of Music* (Univ. Chicago Press, Chicago 1960)

27.2 P.E. Keller, B.H. Repp: Staying offbeat: Sensorimotor syncopation with structured and unstructured auditory sequences, Psychol. Res. **69**(4), 292–309 (2005)

27.3 H.H. Schulze: The detectability of local and global displacements in regular rhythmic patterns, Psychol. Res. **40**, 172–181 (1978)

27.4 F. Apoux, E.W. Healy: A glimpsing account of the role of temporal fine structure information in speech recognition, Adv. Exp. Med. Biol. **787**, 119–126 (2013)

27.5 R.B. Ivry: The representation of temporal information in perception and motor control, Curr. Opin. Neurobiol. **6**(6), 851–857 (1996)

27.6 E.E. Hannon, S. Trehub: Metrical categories in infancy and adulthood, Psychol. Sci. **16**(1), 48–55 (2005)

27.7 E.E. Hannon, L.J. Trainor: Music acquisition: Effects of enculturation and formal training on development, Trends Cogn. Sci. **11**(11), 466–472 (2007)

27.8 G. Soley, E.E. Hannon: Infants prefer the musical meter of their own culture: A cross-cultural comparison, Dev. Psychol. **46**(1), 286–292 (2010)

27.9 H. Honing, H. Merchant, G.P. Háden, L. Prado, R. Bartolo: Rhesus monkey (Macaca mulatta) detect rhythmic groups in music, but not the beat, PLoS ONE **12**, e51369 (2012)

27.10 H. Merchant, H. Honing: Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis, Front. Neurosci. **7**, 1–8 (2014)

27.11 E.W. Large: Resonating to musical rhythm: Theory and experiment. In: *Psychology of Time*, ed. by S. Grondin (Emerald, Bingley 2008) pp. 189–231

27.12 W.E. Benjamin: A theory of musical meter, Music Percept. **1**(4), 355–413 (1984)

27.13 F. Lerdahl, R. Jackendoff: *A Generative Theory of Tonal Music* (MIT Press, Cambridge 1983)

27.14 J. London: *Hearing in Time: Psychological Aspects of Musical Meter* (Oxford Univ. Press, New York 2004)

27.15 C. Palmer, C.L. Krumhansl: Mental representations for musical meter, J. Exp. Psychol. Hum. Percept. Perform. **16**(4), 728–741 (1990)

27.16 D. Epstein: *Shaping Time: Music, The Brain and Performance* (MacMillan, New York 1995)

27.17 M. Grube, T.D. Griffiths: Metrically-enhanced temporal encoding and the subjective perception of rhythmic sequences, Context **45**(1), 72–79 (2009)

27.18 E.W. Large, P. Fink, J.A. Kelso: Tracking simple and complex sequences, Psychol. Res. **66**(1), 3–17 (2002)

27.19 P.J. Essens: Hierarchical organization of temporal patterns, Percept. Psychophys. **40**(2), 69–73 (1986)

27.20 K. Sakai, O. Hikosaka, S. Miyauchi, R. Takino, T. Tamada, N.K. Iwata, M. Nielsen: Neural representation of a rhythm depends on its interval ratio, J. Neurosci. **19**(22), 10074–10081 (1999)

27.21 R. Parncutt: A perceptual model of pulse salience and metrical accent in musical rhythms, Music Percept. **11**(4), 409–464 (1994)

27.22 J.A. Grahn, M. Brett: Rhythm and beat perception in motor areas of the brain, J. Cogn. Neurosci. **19**(5), 893–906 (2007)

27.23 J.A. Grahn: Neural mechanisms of rhythm perception: Current findings and future perspectives, Top. Cogn. Sci. **4**, 585–606 (2012)

27.24 P.J. Essens, D.J. Povel: Metrical and nonmetrical representations of temporal patterns, Percept. Psychophys. **37**(1), 1–7 (1985)

27.25 D.J. Povel: Internal representation of simple temporal patterns, J. Exp. Psychol. Hum. Percept. Perform. **7**(1), 3–18 (1981)

27.26 C. Drake, M.R. Jones, C. Baruch: The development of rhythmic attending in auditory sequences: Attunement, referent period, focal attending, Cognition **77**(3), 251–288 (2000)

27.27 M.R. Jones, M. Boltz: Dynamic attending and responses to time, Psychol. Rev. **96**(3), 459–491 (1989)

27.28 M.R. Jones, P.Q. Pfordresher: Tracking melodic events using joint accent structure, Can. J. Exp. Psychol. **51**(4), 271–291 (1997)

27.29 D.J. Povel, H. Okkerman: Accents in equitone sequences, Percept. Psychophys. **30**(6), 565–572 (1981)

27.30 B.H. Repp: Sensorimotor synchronization: A review of the tapping literature, Psychon. Bull. Rev. **12**(6), 969–992 (2005)

27.31 A.D. Patel, J.R. Iversen, Y. Chen, B.H. Repp: The influence of metricality and modality on synchronization with a beat, Exp. Brain Res. **163**(2), 226–238 (2005)

27.32 B.H. Repp: Sensorimotor synchronization and perception of timing: Effects of music training and task experience, Hum. Mov. Sci. **29**(2), 200–213 (2010)

27.33 B.H. Repp: The embodiment of musical structure: Effects of musical context on sensorimotor synchronization with complex timing patterns. In: *Common Mechanisms in Perception and Action: Attention and Performance XIX*, ed. by W. Prinz, B. Hommel (Oxford Univ. Press, Oxford 2002) pp. 245–265

27.34 M. Franěk, T. Radil, M. Indra: Tracking irregular acoustic patterns by finger tapping, Int. J. Psychophysiol. **6**(4), 327–330 (1988)

27.35 C. Drake, A. Penel, E. Bigand: Tapping in time with mechanically and expressively performed music, Music Percept. **18**(1), 1–13 (2000)

27.36 N. Oram, L.L. Cuddy: Responsiveness of Western adults to pitch-distributional information in melodic sequences, Psychol. Res. **57**(2), 103–118 (1995)

27.37 M.A. Schmuckler, M.G. Boltz: Harmonic and rhythmic influences on musical expectancy, Percept. Psychophys. **56**(3), 313–325 (1994)

27.38 P. Toivanen, J.S. Snyder: Tapping to Bach: Resonance-based modeling of pulse, Music Percept. **21**(1), 43–80 (2003)

27.39 J. Synder, C.L. Krumhansl: Tapping to ragtime: Cues to pulse finding, Music Percept. **18**(4), 455–489 (2001)

27.40 P. Desain: A (de)composable theory of rhythm perception, Music Percept. **9**(4), 439–454 (1992)

27.41 E.W. Large, M.R. Jones: The dynamic of attending how people track time-varying events, Psychol. Rev. **106**(1), 119–159 (1999)

27.42 B.H. Repp: R. Doggett: Tapping to a very slow beat: A comparison of musicians and non-musicians, Music Percept. **24**(4), 367–376 (2007)

27.43 J.A. Bailey, V.B. Penhune: Rhythm synchronization performance and auditory working memory in early- and late-trained musicians, Exp. Brain Res. **204**(1), 91–101 (2010)

27.44 J.L. Chen, V.B. Penhune, R.J. Zatorre: Moving on time: Brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training, J. Cogn. Neurosci. **20**(2), 226–239 (2008)

27.45 M. Besson, F. Faita: An event-related potential (ERP) study of musical expectancy: Comparisons of musicians with non-musicians, J. Exp. Psychol. Hum. Percept. Perform. **21**(6), 1278–1296 (1995)

27.46 M.L.A. Jongsma, E. Meeuwissen, P.G. Vos, R. Maes: Rhythm perception: Speeding up or slowing down affects different subcomponents of the ERP P3 complex, Biol. Psychol. **75**(3), 219–228 (2007)

27.47 A.M. Wing, A.B. Kristofferson: Response delays and the timing of discrete motor responses, Percept. Psychophys. **14**(1), 5–12 (1973)

27.48 P.L. Nunez, R. Srinivasan: *Electric Fields of the Brain: The Neurophysics of EEG* (Oxford Univ. Press, London 1981)

27.49 M. Hämäläinen, R. Hari, R.J. Ilmoniemi, J. Knuutila, O.V. Lounasmaa: Magnetoencephalography-theory, instrumentation and applications to noninvasive studies of the working human brain, Rev. Modern Phys. **65**, 413–497 (1993)

27.50 E. Geiser, E. Ziegler, L. Jäncke, M. Meyer: Early electrophysiological correlates of meter and rhythm processing in music perception, Cortex **45**, 93–102 (2009)

27.51 M.L.A. Jongsma, T. Eichele, R.Q. Quiroga, K.M. Jenks, P.W.M. Desain, H.J. Honing: Expectancy effects on omission evoked potential in musicians and non-musicians, Psychophysiology **42**, 191–201 (2005)

27.52 P. Vuust, K.J. Pallesen, C. Bailey, T.L. van Zuijen, A. Gjedde, A. Roepstorff, L. Ostergaard: To musicians, the message is in the meter: Pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians, Neuroimage **24**, 560–564 (2005)

27.53 O. Ladinig, H. Honing, G.P. Haden, I. Winkler: Probing attentive and pre-attentive emergent meter in adult listeners with no extensive music training, Music Percept. **26**, 377–386 (2009)

27.54 D. Cohen: Magnetoencephalography: Evidence of magnetic fields produced by alpha-rhythm currents, Science **161**, 784–786 (1968)

27.55 P. Vuust, L. Ostergaard, K.J. Pallesen, C. Bailey, A. Roepstorff: Predictive of music–brain responses to rhythmic incongruity, Cortex **45**(1), 80–92 (2009)

27.56 S.J. Kung, O.J.L. Tzeng, D.L. Hung, D.H. Wu: Dynamic allocation of attention to metrical and grouping accents in rhythmic sequences, Exp. Brain Res. **210**(2), 269–282 (2011), https://doi.org/10.1007/s00221-011-2630-2

27.57 E. Geiser, P. Sandmann, L. Jäncke, M. Meyer: Refinement of metre perception – Training increases hierarchical metre processing, Eur. J. Neurosci. **32**(11), 1979–1985 (2010)

27.58 M.L.A. Jongsma, R.Q. Quiroga, C.M. van Rijn: Rhythmic training decreases latency-jitter of omission evoked potentials (OEPs) in humans, Neurosci. Lett. **355**(3), 189–192 (2004)

27.59 E.W. Large, J.F. Kolen: Resonance and the perception of musical meter, Connect. Sci. **6**, 177–208 (1994)

27.60 L. Van Noorden, D. Moelants: Resonance in the perception of musical pulse, J. New Music Res. **28**(1), 43–66 (1999)

27.61 E.W. Large, J.S. Snyder: Pulse and meter as neural resonance, Ann. N. Y. Acad. Sci. **1169**, 46–57 (2009)

27.62 J. Bhattacharya, H. Petsche, E. Pereda: Long-range synchrony in the gamma band: Role in music perception, J. Neurosci. **21**(16), 6329–6337 (2001)

27.63 O. Jensen, J. Kaiser, J.P. Lachaux: Human gamma-frequency oscillations associated with attention and memory, Trends Neurosci. **30**(7), 317–324 (2007)

27.64 R. Salmelin, M. Hämäläinen, M. Kajola, R. Hari: Functional segregation of movement-related rhythmic activity in the human brain, Neuroimage **2**(4), 237–243 (1995)

27.65 J.R. Iversen, A.D. Patel, K. Ohgushi: Perception of rhythmic grouping depends on auditory experience, J. Acoustical Soc. Am. **124**(4), 2263–2271 (2008)

27.66 J.S. Snyder, E.W. Large: Gamma-band activity reflects the metric structure of rhythmic tone sequences, Cogn. Brain Res. **24**(1), 117–126 (2005)

27.67 J.A. Grahn, J.B. Rowe: Feeling the beat: Premotor and striatal interactions in musicians and nonmusicians during beat processing, J. Neurosci. **29**(23), 7540–7548 (2009)

27.68 J.P. Lachaux, P. Fonlupt, P. Kahane, L. Minotti, D. Hoffmann, O. Bertrand, M. Baciu: Relationship between task-related gamma oscillations and BOLD signal: New insights from combined fMRI and intracranial EEG, Hum. Brain Mapp. **28**(12), 1368–1375 (2007)

27.69 J.M. Zumer, M.J. Brookes, C.M. Stevenson, S.T. Francis, P.G. Morris: Relating BOLD fMRI and neural oscillations through convolution and optimal linear weighting, NeuroImage **49**(2), 1479–1489 (2010)

27.70 S. Nozaradan, I. Peretz, M. Missal, A. Mouraux: Tagging the neuronal entrainment to beat and meter, J. Neurosci. **31**(28), 10234–10240 (2011)

27.71 S. Nozaradan, I. Peretz, A. Mouraux: Selective neuronal entrainment to the beat and meter embedded in a musical rhythm, J. Neurosci. **32**(49), 17572–17581 (2012)

27.72 R.D. Pascual-Marqui: Review of methods for solving the EEG inverse problem, Int. J. Bioelectromag. **1**, 75–86 (1999)

27.73 A.M. Dale, M.I. Sereno: Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach, J. Cogn. Neurosci. **5**, 162–176 (1993)

27.74 A.M. Dale, E. Halgren: Spatiotemporal mapping of brain activity by integration of multiple imaging modalities, Curr. Opin. Neurobiol. **11**, 202–208 (2001)

27.75 S. Ogawa, T.M. Lee, A.R. Kay, D.W. Tank: Brain magnetic resonance imaging with contrast dependent on blood oxygenation, Proc. Natl. Acad. Sci. USA **87**(24), 9868–9872 (1990)

27.76 R. Turner, D.L. Bihan, C.T.W. Moonen, D. Despres, J. Frank: Echo-planar time course MRI of cat brain oxygenation changes, Mag. Reson. Med. **22**(1), 156–166 (1991)

27.77 K.K. Kwong, J.W. Belliveau, D.A. Chesler, I.E. Goldberg, R.M. Weisskoff, B.P. Poncelet, D.N. Kennedy, B.E. Hoppel, M.S. Cohen, R. Turner, H.-M. Cheng, T.J. Brady, B.R. Rosen: Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation, Proc. Natl. Acad. Sci. USA **89**(12), 5675–5679 (1992)

27.78 H. Devlin: What is functional magnetic resonance imaging (fMRI)?, http://psychcentral.com/lib/2007/what-is-functional-magnetic-resonance-imaging-fmri (2012)

27.79 J.L. Chen, V.B. Penhune, R.J. Zatorre: Listening to musical rhythms recruits motor regions of the brain, Cereb. Cortex **18**(12), 2844–2854 (2008)

27.80 J.M. Mayville, K.J. Jantzen, A. Fuchs, F.L. Steinberg, J.A.S. Kelso: Cortical and subcortical networks underlying syncopated and synchronized coordination revealed using fMRI, Hum. Brain Mapp. **17**(4), 214–229 (2002)

27.81 R.I. Schubotz, D.Y. von Cramon: Interval and ordinal properties of sequences are associated with distinct premotor areas, Cereb. Cortex **11**(3), 210–222 (2001)

27.82 F. Ullén, H. Forssberg, H.H. Ehrsson: Neural networks for the coordination of the hands in time, J. Neurophysiol. **89**(2), 1126–1135 (2003)

27.83 S. Teki, M. Grube, S. Kumar, T.D. Griffiths: Distinct neural substrates of duration-based and beat-based auditory timing, J. Neurosci. **31**(10), 3805–3812 (2011)

27.84 E. Geiser, M. Notter, J.D. Gabrieli: A corticostriatal neural system enhances auditory perception through temporal context processing, J. Neurosci. **32**(18), 6177–6182 (2012)

27.85 S.J. Kung, J.L. Chen, R.J. Zatorre, V.B. Penhune: Interacting cortical and basal ganglia networks underlying finding and tapping to the musical beat, J. Cogn. Neurosci. **25**(3), 401–420 (2013)

27.86 A. Riecker, J. Kassubek, K. Groschel, W. Grodd, H. Ackermann: The cerebral control of speech tempo: Opposite relationship between speaking rate and BOLD signal changes at striatal and cerebellar structures, NeuroImage **29**(1), 46–53 (2006)

27.87 A. Riecker, D. Wildgruber, K. Mathiak, W. Grodd, H. Ackermann: Parametric analysis of rate-dependent hemodynamic response functions of cortical and subcortical brain structures during auditory cued finger tapping: A fMRI study, NeuroImage **18**(3), 731–739 (2003)

27.88 P. Fraisse: Rhythm and tempo. In: *Psychology of Music*, ed. by D. Deutsch (Academic, New York 1982) pp. 149–180

27.89 J.D. McAuley, M. Henry: Visual rhythms do not receive automatic auditory encoding, Atten. Percept. Psychophys. **72**(5), 1377–1389 (2010)

27.90 G. Collier, G. Logan: Modality differences in short-term memory for rhythms, Memory Cogn. **28**(4), 529–538 (2000)

27.91 R. Fendrich, P. Corballis: The temporal cross-capture of audition and vision, Atten. Percept. Psychophys. **63**(4), 719–725 (2001)

27.92 B.H. Repp, A. Penel: Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences, J. Exp. Psychol. Hum. Percept. Perform. **28**(5), 1085–1099 (2002)

27.93 J.A. Grahn, M.J. Henry, J.D. McAuley: fMRI investigation of cross-modal interactions in beat perception: Audition primes vision, but not vice versa, NeuroImage **54**(2), 1231–1243 (2011)

27.94 J.A. Grahn, M. Brett: Impairment of beat-based rhythm discrimination in Parkinson's disease, Cortex **45**(1), 54–61 (2009)

27.95 D. Bueti, E.V. van Dongen, V. Walsh: The role of superior temporal cortex in auditory timing, PLoS ONE **3**, e2481 (2008)

27.96 M. Grube, K.-H. Lee, T.D. Griffiths, A.T. Barker, P.W. Woodruff: Transcranial magnetic theta-burst stimulation of the human cerebellum distinguishes absolute, duration-based from relative, beat-based perception of subsecond time intervals, Front. Psychol. (2010), https://doi.org/10.3389/fpsyg.2010.00171

27.97 M.P. Malcolm, A. Lavine, G. Kenyon, C. Massie, M. Thaut: Repetitive transcranial magnetic stimulation interrupts phase synchronization during rhythmic motor entrainment, Neurosci. Lett. **435**(3), 240–245 (2008)

27.98 M.C. Ridding, B. Brouwer, M.A. Nordstrom: Reduced interhemispheric inhibition in musicians, Exp. Brain Res. **133**(2), 249–253 (2000)

27.99 E.M.F. Wilson, N.J. Davey: Musical beat influences corticospinal drive to ankle flexor and extensor muscles in man, Int. J. Psychophysiol. **44**(2), 177–184 (2002)

27.100 M. Coltheart: What has functional neuroimaging told us about the mind (so far), Cortex **42**, 323–331 (2006)

27.101 M.P.A. Page: What can't functional neuroimaging tell the cognitive psychologist, Cortex **42**(3), 428–443 (2006)

27.102 I. Winkler, G.P. Haden, O. Ladinig, I. Sziller, H. Honing: Newborn infants detect the beat in music, Proc. Natl. Acad. Sci. USA **106**(7), 2468–2471 (2009)

27.103 J. Phillips-Silver, L.J. Trainor: Feeling the beat: Movement influences infant rhythm perception, Science **308**(5727), 1430 (2005)

27.104 I. Cross: Music, cognition, culture and evolution. In: *The Biological Foundations of Music*, ed. by R. Zatorre, I. Peretz (The New York Academy of Sciences, New York 2001) pp. 28–42

27.105 D.W. Gerry, A.L. Faux, L.J. Trainor: Effects of Kindermusik training on infants' rhythmic enculturation, Dev. Sci. **13**(3), 545–551 (2010)

27.106 S.W. Keele, R. Nicoletti, R. Ivry, R.A. Pokorny: Mechanisms of perceptual timing: Beat-based or interval-based judgements?, Psychol. Res. **50**, 251–256 (1989)

27.107 J.D. McAuley, M.R. Jones: Modeling effects of rhythmic context on perceived duration: A comparison of interval and entrainment approaches to short-interval timing, J. Exp. Psychol. Hum. Percept. Perform. **29**(6), 1102–1125 (2003)
27.108 H. Pashler: Perception and production of brief durations: Beat-based versus interval-based timing, J. Exp. Psychol. Hum. Percept. Perform. **27**(2), 485–493 (2001)
27.109 J.D. McAuley, G.R. Kidd: Effect of deviations from temporal expectations on tempo discrimination of isochronous tone sequences, J. Exp. Psychol. Hum. Percept. Perform. **24**, 1786–1800 (1998)

28. Music and Action

Giacomo Novembre, Peter E. Keller

In this chapter, the relationship between music and action is examined from two perspectives: one where individuals learn to play an instrument, and another where music induces movement in a listener. For both perspectives, we review experimental research, mostly consisting of neuroscientific studies, as well as select behavioral investigations. We first review research examining how learning to play music induces functional coupling between motor and sensory neural processes, which ultimately changes the way in which music is perceived. Next, we review research examining how certain temporal properties of music (such as the rhythm or the beat) induce motor processes in a listener, depending on or irrespective of musical training. The coupling of perceptual and motor processes underpins predictive computations that facilitate the anticipation and adaptation of one's movement to music. Such skills in turn support the capacity to coordinate one's movements with another in the context of joint musical performance. This picture emphasizes how studying the relationship between music and action will ultimately lead us to understand music's powerful social and interpersonal potential.

When we consider music as a form of human action, it is easy to become astonished by the remarkable complexity and precision achieved by expert musicians during a performance. Yet it is under such exceptional conditions that the human brain discloses its capability for *orchestrating* multiple body parts in terms of their timing, force, and kinematics, which undergo dynamic transformations throughout ever-changing musical pieces. But this is only part of the puzzle. Music can be read from a score or recalled from memory. It can be rehearsed, but also improvised. It can be performed solo or in an ensemble, which necessitates coordination at a millisecond timescale between individuals who may be perfect strangers. When all this is considered together, it becomes clear that music is not just an artistic form of expression, but also a behavioral trace of the outstanding resources of the human brain and body. Somehow complementary with these observations, music is known to induce movement responses in a listener, a phenomenon that starts in the fetus during the final weeks of gestation [28.1], and develops through infancy [28.2, 3] into more complex forms of actions [28.4] including dance. The present chapter is intended to give an exposé of the significant advances that has been made towards understanding the cognitive and neural mechanisms that link music and action.

Several different psychological and physiological processes take place simultaneously during music production (sensory, motor, memory, emotional), and because all of these are intrinsically linked with one another, it is challenging for scientists to examine specific functions in isolation. Let us take a basic example: striking a piano key with a finger. The movement (striking the key) is intended to generate a goal (a piano tone). When this is observed from the *outside* perspective of another individual, this seems straightforward: the movement preceded its goal. However, when con-

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_28

sidering a *first person* perspective, it is the musician's intention (i. e., producing a piano tone) that leads the generation of a movement: moving the finger toward the piano key.

This distinction might seem trivial, but in fact it represents a fundamental step to understanding that movements and their ensuing effects are intrinsically *coupled* in the human brain and cognition. More specifically, a representation of a perceptual effect can trigger the movement necessary to produce the effect itself, particularly in musicians, who display special benefits of brain plasticity [28.5–7]. This notion has strongly influenced the study of music cognition and its underpinning neural correlates, and it has been inspired by converging evidence from the fields of cognitive psychology [28.8, 9] and neurophysiology [28.10, 11], as well as computational neuroscience [28.12, 13].

While focusing on the notion of action-perception coupling, we will examine two distinct perspectives on this issue. Firstly, we will examine the role of experience and musical training in binding sensory and motor neural processes. Secondly, we will consider certain temporal properties of music, such as its rhythm or the musical beat, that induce motor processes in a listener. For both perspectives, we will pay particular attention to the predictive character of these neural processes, and their crucial role in allowing individuals to physically respond to music, including interpersonal coordination with other musicians in the context of a joint musical action. We will focus our attention on neuroimaging studies that place special emphasis on cognitive mechanisms, mostly run with musically trained participants (or directly relevant for them), with only a few references to purely behavioral investigations.

28.1 Coupling Action and Perception Through Musical Experience

The highly plastic nature of the musician's brain has been emphasized in the literature in recent years (for reviews, see [28.6, 7]). In this section, we will focus on studies that specifically addressed action-perception coupling in musicians or in individuals who received musical training for experimental purposes. We will first list some behavioral investigations (Sect. 28.1.1), and then move on to neuroimaging evidence from studies using hemodynamic measures (fMRI) and electrophysiological techniques (EEG, MEG), as well as brain stimulation methods (TMS) (Sect. 28.1.2). Finally, special attention will be given to recent studies addressing the predictive character of action-perception coupling, and related mechanisms such as *internal models* [28.14–16] (Sect. 28.1.3).

28.1.1 Behavioral Evidence

Research in experimental psychology has explored action-perception links in music through the use of action-effect compatibility manipulations. *Drost* et al., have conducted a number of behavioral studies [28.17–19] comparing pianists and nonmusicians in the context of an interference paradigm where participants had to play a chord on a piano in response to visual imperative stimuli. These visual stimuli were accompanied by simultaneously presented task-irrelevant sounds, which could either match or not match the target chord. It was found that incongruent sounds delayed execution time in pianists but not nonpianists [28.17]. In addition, these incongruent sounds tended to induce false responses, i. e., production of the heard chord, instead of the imperative one [28.18]. Moreover, this interference could only be observed when the timbre of the musical sound matched the participant's instrument background [28.19]. The studies by Drost and colleagues demonstrate that auditory perception primes action if strong action-perception links have been established through instrument-specific training.

In related research, *Keller* and *Koch* [28.20] found that mental images of anticipated action effects can prime responses to a similar degree as is observed with congruent and incongruent sounds, highlighting the role of action-perception coupling in action preplanning (i. e., before sounds are actually perceived) [28.20, 21]. Subsequent studies investigated such preplanning in sequential actions – a definitive aspect of music performance – by requiring participants to respond to visual imperative stimuli by producing a series of finger taps on vertically aligned keys. Taps triggered tones in some conditions, where the key-to-tone mapping was manipulated (between blocks of trials) to be either congruent or incongruent in terms of pitch and spatial height. One version of this task [28.22] required participants to respond as quickly as possible to the imperative stimuli. Results indicated that reaction times were shorter in conditions where sequences of finger taps and tones were congruent in *height* than when they were incongruent. This effect was restricted to musicians and, furthermore, increased in size with years of musical experience. Therefore, action-perception coupling associated with musical training allowed participants to plan their actions by imagining the auditory sequences in an

anticipatory fashion, and the efficiency of such preplanning was greatest when movements and their auditory effects were congruent. Further studies using a version of the paradigm that required taps to be produced at a specific tempo (rather than rapidly) demonstrated that action-perception coupling not only enhances the efficiency of action planning, but also facilitates timing accuracy and economical force control by optimizing movement kinematics [28.21, 23].

Taken together, this behavioral evidence indicates that the perception or mental imagery of sounds – which would normally be associated with specific movements – trigger representations of those specific movements (see also [28.24, 25] for additional evidence supporting this notion).

28.1.2 Neuroimaging Evidence

Neuroimaging research has shed light on the neurophysiological mechanisms underpinning action-perception coupling in the musician's brain. *Haueisen* and *Knösche* designed a magnetoencephalography (MEG) study that allowed them to investigate brain responses to well-known piano pieces in musicians with or without piano experience [28.26]. In piano players, perception of these pieces led to an increase in neural activity over the motor cortex hand area. Most interestingly, the authors found a distinct spatial response to notes that would be preferably played with the thumb versus the little finger, which matched the homuncular organization of the primary motor cortex (M1). The finding that the acoustic perception of music within an individual's behavioral repertoire led to an increase of motor cortical activity in musicians has been replicated in other neuroimaging studies using different methods. For example, *D'Ausilio* et al., used transcranial magnetic stimulation (TMS) to trigger motor evoked potentials (MEP) in a forearm muscle normally used to play the piano [28.27]. Corticospinal excitability (which was indexed by the amplitude of the MEPs) was found to increase while pianists listened to a rehearsed piano piece compared to an unrehearsed one. In another study, *Bangert* et al., ran an fMRI study where professional pianists and nonmusicians heard novel piano sequences that were synthesized online (and therefore could not be familiar) [28.28]. Compared to nonmusicians, professional pianists showed a broad network of motor areas responding to the piano sequences, including both primary motor and premotor (BA 4/6) regions, inferior frontal cortex and Broca's area. Interestingly, to explore whether this auditory-to-motor transformation was bidirectional, the authors also examined the effect of producing piano tones in the absence of auditory feedback. This latter task led to the activation of auditory-related brain regions, such as the superior temporal gyrus (BA 22) in professional pianists.

Motor activations in the musician's brain can not only be elicited by the acoustic presentation of music, but also visual presentations of musical actions. In fMRI experiments, *Haslinger* et al., and *Hasegawa* et al., presented video recordings of hands playing a silent keyboard [28.29, 30]. Despite the fact that these videos were mute (i. e., no sounds were presented), the authors reported the activation of a frontoparietal brain network – including premotor cortex (BA6) and inferior parietal cortex – which was similar to the one revealed in the study by *Bangert* et al. [28.28] (who presented sounds, instead of silent videos).

Taken together, these data indicate that musical training leads to the emergence of cross-modal action-perception coupling, where the effects of musical actions (either the sounds produced or the visual presentation of the movement patterns) trigger a representation of the movements necessary to produce these effects. This idea has profound implications for human cognition more broadly, because it could be applied to other motor tasks and therefore generalize across individuals with different types of experience. For this reason, other research has investigated to what extent these results could be replicated in naïve participants who received musical training only shortly before taking part in an experiment. *Lahav* et al. [28.31] trained nonmusicians to play a piano piece by ear (without notation) over a period of five days. Following the training, participants were presented with either the trained pieces, untrained pieces (having the same notes, but in a different order) or familiar but motorically unknown pieces [28.31]. Remarkably, the frontoparietal motor-related network discussed above (here comprising Broca's area, the premotor region, the intraparietal sulcus, and the inferior parietal region) responded most strongly to the trained pieces, weakly to the untrained ones, and not at all to the motorically unknown pieces (relative to a rest baseline condition). Thus, a few days of training were sufficient to replicate the effects – previously described in studies with experienced musicians – in a group of nonmusicians (see also [28.32] for earlier EEG evidence supporting this).

A few recent studies have made noteworthy progress towards understanding the functioning of this action-perception coupling mechanism, and how it emerges through learning. *Engel* et al. [28.33] trained nonmusicians to play melodies either by ear (and without seeing their hands) or by imitating visual movement patterns (without auditory feedback) [28.33]. Following training, participants were able to recognize melodies

learned in one modality upon presentation in the other (i. e., untrained) modality. However, recognition accuracy and fMRI data indicated the cross-modal transfer was stronger when the melodies had been trained by ear. Moreover, in order to demonstrate that sensory-motor coupling emerges as a result of motor learning, and not visual familiarity, *Candidi* et al. [28.34] trained nonmusicians to recognize piano fingering errors during the visual presentation of silent musical sequences [28.34]. Expert pianists showed a somatotopic corticospinal facilitation – indexed by the amplitude of motor-evoked potentials triggered by transcranial magnetic stimulation (TMS) – of the finger that committed the error (consistent with the study by *Haueisen* and *Knösche* [28.26], who also reported finger-specific activations, but in response to acoustically presented music). Visually-trained nonmusicians, however, did not show the same facilitation effect, although they were equally able to recognize the errors. Thus, visual experience is not sufficient to recognize movement patterns if motor learning has not taken place, at least for low-level (i. e., fingering) motor representations.

Taking together the studies reviewed above, there is converging evidence from both behavioral and neurophysiological methods (including EEG, MEG, TMS and fMRI) that, given an association between movements and their ensuing effects, the perception of an effect can trigger a representation of the movement necessary to execute it. The musician's brain is an excellent example of action-perception coupling because movements and intended sounds become strongly associated after long-term musical training. On the behavioral level, it has been consistently shown that the representation of a musical sound and the motor resources necessary to perform the sound are represented by a comparable code and can interfere with each other [28.17–19, 22] see also [28.35]. On the neural level, listening to a trained musical sequence activates the motor brain areas necessary for executing it, as evidenced by measures such as corticospinal excitability [28.27], blood oxygen level dependent signal [28.28, 31], EEG potentials [28.32] and MEG fields [28.26]. Conversely, the visual perception of (silent) musical actions leads to similar brain co-activations [28.29, 30, 33, 34], demonstrating that action-perception coupling in the musicians' brain is multimodal (i. e., visual and auditory). Additional research has shown that these coupling effects can also result after a short period of musical training (with naïve participants), implying that such an action-perception matching system is not necessarily musician-specific, but rather stands as task-specific example of a cognitive mechanism with broader relevance [28.31–33, 36].

28.1.3 The Temporal Dynamics and Predictive Character of Action–Perception Coupling

The studies reviewed above were not designed to address the temporal dynamics of action-perception coupling in the brain, but this aspect is fundamental to understanding the cognitive and behavioral relevance of such coupling.

Let us come back to the example of the finger striking a piano key (movement) to generate a sound (goal). As we noted, from a first-person perspective, it is the musician's intention (i. e., to produce a piano tone) that leads to the execution of a movement. Given this, one would hypothesize that – in the musician's brain – the two processes associated with intending to perform a specific keystroke, and hearing the auditory feedback, are at least in part independent and have different priorities, i. e., the actual sound of the key should be predicted on the basis of its preceding *intended* neural representation.

Following this reasoning, *Maidhof* et al. [28.37] and *Ruiz* et al. [28.38] conducted similar EEG studies [28.37, 38] where they examined the event-related potentials (ERPs) preceding the execution of piano errors. Both these studies reported both a behavioral and an electrophysiological marker of performance errors. First, erroneous keystrokes were produced with less force, and therefore generated a softer sound. Second, an early negative deflection (or error-related negativity, ERN) of the EEG signal was found to anticipate the actual mistake by 100 ms [28.37] and 50–70 ms [28.38]. Source localization analysis revealed that these responses were generated by the anterior cingulate cortex (a brain region implicated in action monitoring [28.39, 40]) and, most interestingly, this effect was independent of whether or not auditory feedback was available [28.38]. Thus, errors were detected *prior* to their execution, and this occurred independently of whether the pianists could hear the actual feedback of the performance.

These findings are particularly important because they provide evidence that, during performance, internal forward models predict the outcome of an ongoing motor command by comparing it with an efference copy (i. e., a prediction of the perceptual effects of the motor command) [28.41]. This indicates that, during the execution of a musical sequence, images of the *intended* sounds are formed well ahead their generation, and compared in real time with the state of the body. Thus, the coupling of sensory and motor cortices is a dynamical process with a strong anticipatory character that, given the existence of an association between movements and their ensuing effects, permits the generation

of predictions about the state of one's own body and the sensory consequences of movements.

Further evidence has supported the proposal that internal models play a role in real-time prediction during online action planning. *Maidhof* et al. [28.42] compared EEG brain responses to expectancy violations in musical action (i. e., during piano performance) and perception (i. e., during listening) [28.42]. Both types of violation led to a negativity peaking at around 200 ms after tone presentation. However, the amplitude was larger for the action violation compared to the perceptual violation, indicating that the expectations associated with the intention to produce a tone override those based on perceptual processes alone. Furthermore, *Ruiz* et al. [28.43] explored EEG oscillatory markers predicting an error during musical performance [28.43]. It was shown that a burst of beta-band oscillations that originated from the posterior frontomedial cortex (pFMC, which includes the anterior cingulate cortex, see [28.38]) anticipated the error by 120 ms. Moreover, the efficiency of motor control correction mechanisms, i. e., the reduction of the force utilized to execute a wrong note, could be predicted based on the beta-band synchronization (note that beta oscillations are associated with sensorimotor and, especially, motor functions [28.44, 45]) between pFMC and brain regions implementing control adjustments (i. e., lateral prefrontal cortex) ([28.37, 38]). Taken together, these data indicate further that musical training leads to a tight coupling between sensory and motor cortices, which might underpin the generation of sensory predictions – based on internal models – *within* the musician's brain.

A remarkable property of this dynamical process is that it does not only permit predictions of our own movements, but can also be used to generate predictions about others' actions [28.14, 46]. *Lee* and *Noppeney* [28.47] designed an elegant combined psychophysical-fMRI study to investigate the temporal binding between sensory and motor processes in musicians and nonmusicians [28.47]. Participants were required to attend to musical or speech stimuli in which the synchrony between sounds and images (of either a speaking face or a hand playing the piano) was manipulated parametrically. As could be expected, the two groups were equally sensitive to the temporal asynchronies in the speech domain, but the musician group was superior in detecting temporal asynchronies in the musical domain. Dynamic causal modeling revealed that this superior performance was associated with greater effective connectivity within a network of brain regions including the superior temporal sulcus, the premotor cortex and the cerebellum. Thus, cross-modal plasticity due to musical training (as reviewed in Sect. 28.1.2) led to the fine-tuning of internal forward models (Sect. 28.1.3) that, critically, permit the generation of predictions of observed actions with high temporal resolution. Accordingly, coupling emerges within an individual brain, but can also be used to generate predictions about others' actions.

These fine-tuned internal forward models might allow musicians to predict not only *when* an event will occur, but in some cases also *what* event will occur. Through training, the musician's brain does not only bind specific events across sensory and motor modalities. In addition, the brain learns which successions of tones are most likely to occur according to statistical regularities associated with the rules that govern harmony (i. e., sequential chord progressions) in a given musical tradition. This phenomenon has been studied for some time in the context of purely auditory perceptual experiments. Participants with and without musical backgrounds were presented with sequences of chords containing or not a violation of their harmonic structure, while event-related potentials (ERP) were quantified using EEG. In a series of experiments [28.48, 49], it was shown that the perception of a harmonic violation led to an early right anterior negativity (ERAN) peaking at around 200 ms after chord presentation. By comparing expert musicians with novices, it was further shown that the amplitude of this negativity was larger for expert musicians [28.50]. This finding indicates that musical experience led to the generation of stronger expectancies in the perceptual domain. It should be noted, however, that expert musicians acquire these rules not only by means of perceptual exposure (as a naïve listener), but also by means of intensive practice. Therefore, given the tight functional link between sensory and motor cortices highlighted by previous studies, it remained to be explored to what extent expectancies in the auditory domain extend to the motor domain.

A few recent studies examined this issue. First, in a behavioral study, *Novembre* and *Keller* [28.51] presented expert pianists with silent videos displaying a musician's hand performing mute sequences, including occasional chords that were harmonically incongruent with the preceding musical context [28.51]. The pianists were asked to imitate the chords as quickly and accurately as possible. It was shown that, despite the absence of auditory feedback, imitation was faster and more accurate for chords that were preceded by a congruent context. This result suggests that the harmonic rules implied by the observed actions induced strong expectancies that influenced action execution. Thus, this study provided first behavioral evidence in favor of harmonic structures regulating not only perceptual processes (as shown by the previous studies), but

also the motor processes involved in producing these structures.

This finding was replicated in a subsequent study [28.52] where EEG was recorded during task performance. ERP data from this investigation revealed a negativity following the presentation of the final incongruent sequential chord [28.52]. The negativity resembled the ERAN that follows auditory presentation of a harmonically incongruent chord [28.48, 50], or possibly the ERN that anticipates keystroke errors in piano performance [28.37, 38]. Additionally, a second negativity with more posterior distribution was observed, which anticipated the time during which the chord was imitated. This ERP was taken as evidence of response reprogramming.

A follow-up study examined the specificity of this latter ERP, where pianists were presented with photos of a model hand performing the chords, hence addressing more directly the motor – as opposed to the perceptual – component of the task [28.53]. Besides replicating the later negativity observed by *Sammler* et al. [28.52], *Bianco* et al. [28.53] demonstrated that such neural signatures were specifically linked to movements that had to be reprogrammed in terms of *what* to do (type of chord), and was different from the ERP elicited by the revision of *how* to perform the chord (e.g., which set of fingers to employ) [28.53]. Moreover, contrary to results for *how* to perform, imitation time relative to *what* to do was influenced by length of the sequence, indicating that the motor plan for the congruent chord was prepared ahead based on contextual information [28.51]. Most strikingly, recent fMRI results have shown that motor predictions based on implicitly acquired rule-based structures (harmonic relationships) are associated with a network comprising the right homologue of Broca's area (the dorsal part of BA44, bordering premotor cortex – PMC) interlinked with visuomotor parietal areas (bilateral superior parietal lobe) [28.54]. These findings, together with the established involvement of the right homologue of Broca's area in detection of auditory musical irregularities, support the hypothesis that this area is fundamentally involved in structural integration and prediction of complex musical material in both perception and production.

The above findings are particularly noteworthy in that they offer evidence that the well-known predictive character of the motor system is strongly based on a musician's knowledge of harmonic principles. This latter piece of evidence indicates that the motor system predicts not only when an action will occur, but also what kind of action will occur [28.55]. Rule-based predictions in the motor system are consistent with other accounts postulating a sensorimotor processing of syntax, including harmony [28.56, 57].

In conclusion, the studies reviewed in this section suggest that the coupling between sensory and motor cortices underpins predictive computations by means of internal models. The studies by *Maidhof* et al. [28.37, 42] and *Ruiz* et al. [28.38, 43] explored this notion within the musician's brain by looking at the relationship between intended sounds and executed movements. The studies by *Lee* and *Noppeney* [28.47], *Novembre* and *Keller* [28.51], *Sammler* et al. [28.52], and *Bianco* [28.54] examined the prediction of another musician's musical actions. Taken together, the results of this research suggest that musical training shapes and develops a sensorimotor system that generates predictions about the identity and timing of upcoming events. In the next section, we discuss how these processes affect the control of an individual's own movement timing and potentially support real-time interaction between ensemble musicians.

28.2 Responding to Music with Action and (Social) Interaction

Recent research has explored to what extent action–perception coupling functions as a resource for moving our bodies in response to music, which is likely to be a fundamental skill supporting interpersonal behavior and ensemble musical coordination. In this domain, three fundamental mechanisms have received particular attention. Firstly, perceptual investigations have examined the neural correlates mediating the processing of temporal properties of music (such as the rhythm or the musical beat) that induce motor processes in musically trained and untrained individuals. Secondly, sensorimotor tasks have examined the neural correlates that underpin the human ability to coordinate and adapt body movement with an external signal, with special focus on the social processes involved. Finally, we review recent studies that addressed how action representations of self and other emerge and are eventually integrated in the context of joint musical actions that require synchronization between multiple musical parts. We will focus on cognitive and neural mechanisms that are likely to support interpersonal coordination, with a special emphasis on studies conducted with musically trained participants.

28.2.1 Temporal Properties of Music that Induce Motor Processes

From a behavioral perspective, the most straightforward way to explore timing mechanism in the context of action-perception coupling is to employ sensorimotor synchronization tasks that require minimal movement (hence allowing movement artifacts to be controlled for during neuroimaging measurements). The most commonly used approach consists of finger-tapping tasks, where a participant taps his/her finger in time with a pacing signal (often an auditory sequence). Over the course of many decades, behavioral research using finger-tapping tasks has made significant advances in understanding the mechanisms governing sensorimotor synchronization (for a review see [28.58, 59]). Only recently, however, neuroimaging methods have extended this body of knowledge by allowing researchers to observe the brain response to musical properties, such as rhythm, independently of a participant's behavioral response, e.g., during a passive perception task.

In a series of neat studies, *Chen* et al. [28.60] investigated action-perception coupling by using fMRI to measure brain activity while nonmusician participants tapped along with isochronous auditory rhythms [28.60]. The study compared brain and behavioral (tapping) measures across conditions in which the rhythm had different levels of metrical saliency (i. e., sound intensity was manipulated in order to accent every third tone to varying degrees across conditions). It was shown that, as metric saliency increased, louder tones were associated with longer tap durations and, most interestingly, stronger coupling between auditory and motor (especially dorsal premotor) cortical regions. In order to dissociate activations related to rhythm perception from motor processes involved in the behavioral tapping task, a second study used a listening task, and demonstrated that a neural network of motor regions including the supplementary motor area (SMA), mid-PMC and cerebellum, was activated similarly as in the synchronization task [28.61]. Finally, and particularly germane to the theme of the present chapter, a third study compared the activation of this network of motor brain regions across musicians and nonmusicians [28.62]. It was found that regions within this network responded to rhythmic stimuli independently of musical training. Thus, the perception of rhythm evoked motor responses in the brains of musicians and nonmusicians alike.

Another musical property that has been associated with motor activations in the brain is the musical beat. The term *beat* refers to the psychological sense of a regular pulse underlying a rhythm pattern. In an fMRI study, *Grahn* and *Brett* [28.63] measured the response of motor brain regions to the beat of auditory rhythms in participants with and without musical training [28.63]. These authors reported that rhythms having perceptual accents at regular intervals (i. e., hence inducing a stronger sense of the beat) were better reproduced, and associated with increased activation of SMA and basal ganglia in both groups of participants. Another study compared rhythms in which beats were accentuated physically by increasing either the intensity or the duration of tones within cyclically presented patterns (as well as a condition in which the beat was imagined and *nonbeat* conditions without a regular underlying beat) [28.64]. It was reported that, relative to the nonbeat conditions, all three beat conditions were associated with putamen activation (one of the structures of the basal ganglia) and increased connectivity between putamen and premotor, SMA and auditory cortex. The strength of this coupling (particularly for patterns with duration-based accents) was higher in musicians than nonmusicians, indicating that – to some extent – these activations can be boosted through musical training.

A final study shed light on the nature of these activations. *Grahn* and *Rowe* [28.65] aimed at establishing whether the putamen response to beat perception reflected the process of finding the beat (i. e., extracting regularity from the pattern's temporal structure) or the process of generating temporal predictions based on the beat [28.65]. The authors compared conditions in which participants (with varying degrees of musical training) listened to series of rhythmic sequences in which a given sequence could be preceded by a sequence with the same beat, a different beat (faster or slower), or an irregular *nonbeat* sequence. These conditions required beat continuation, beat adjustment, and beat finding respectively. The authors reported that beat continuation was associated with an increase in putamen activity, and that this effect was independent of musical expertise. These data were taken to indicate that putamen activation reflects the generation of online predictions about when a future event will happen in the context of the beat-based musical rhythms. This finding, which is consistent with other evidence that the basal ganglia play a key role in associating auditory stimuli with motor responses (see e.g., [28.66]), has important implications for several musical skills underpinning the interpersonal coordination of multiple musicians (Sects. 28.2.2 and 28.2.3 will address this point directly). More generally, these studies shed light on the neurobiological structures that mediate movement responses to music perception (independently of musical training), such as when listeners tap their feet or sway their bodies to music [28.4, 67].

While the perception of rhythm and beat is a temporally dynamic process, the series of studies described

above utilized fMRI – a technique with very low temporal resolution – to explore these neural processes. This approach was optimal for identifying networks of brain regions responsive to these musical properties, but was not well suited to providing fine-grained information about the temporal dynamics of these activities. EEG and MEG are better equipped for this aim. *Fujioka* et al. [28.68] conducted an MEG study in which neuromagnetic beta-band oscillations (which are associated with sensorimotor and, especially, motor functions [28.44, 45]) were recorded while participants listened to isochronous auditory rhythms (at different rates) or randomly timed sequences [28.68]. While a decrease in beta amplitude occurred after each sound onset in both isochronous and random sequences, the subsequent beta rebound anticipated the next sound onset in a rate-dependent manner only for isochronous sequences. This finding was taken to indicate that the time course of beta modulations reflects the internalization of rhythmic periodicities and that this internalization facilitates predictive timing (a finding that is also consistent with animal studies [28.69]). These periodic beta modulations were localized to auditory and motor brain regions – including the pre- and postcentral gyrus, SMA and the cerebellum. Most importantly, corticocortical phase coherence between primary auditory (Heschl's gyrus) and motor regions (especially SMA) was observed to unfold as a function of stimulus rate. Therefore, these results illuminate the time course of auditory-motor coupling in response to rhythmic stimuli, and indicate directly their relevance for generating temporal predictions.

It is worth noting that auditory-motor coupling effects have also been observed at different temporal scales, such as those that match the signal's periodicities [28.70, 71], using EEG-recorded steady-state-evoked potentials (SSEPs). A recent advance in this field is to link individual differences in covert neural entrainment, indexed by SSEPs, with overt behavioral measures, as in a study showing that stronger endogenous neural entrainment to the rhythmic beat is associated with superior temporal prediction abilities during sensorimotor synchronization with auditory sequences [28.72]. Although it is currently not clear whether the generators of such effects are specifically auditory and motor regions, and the interaction among them (as in the study by [28.68]), it is undeniable that SSEPs are a promising index to tag neural entrainment or attention allocation to the musical beat.

In sum, the fMRI studies reviewed earlier were successful in demonstrating that temporal properties of music, i. e., rhythm and beat, led to activation of a corticosubcortical network of brain regions that underpin auditory-motor interactions. The fine-grained temporal resolution of these cross-modal activations highlighted in electrophysiological results supports the hypothesis that they are implicated in the prediction of sensory events (see also [28.73]). It is important to note that these neural processes could have not been discovered through behavioral methods alone, as in fact they can be generated during the passive perception of a rhythm (i. e., in the absence of behavioral responses). In this sense, the coupling of perception and action during perception of rhythms or beats reflects a spontaneous association of music and action that is independent of movement production (and independent of musical training as well). We argue that this is a fundamental mechanism that permits the binding of movements and percepts in real time (either to precisely time self-produced movements, or predict other-related ones). These computations might be a fundamental resource for mediating interpersonal coordination, i. e., the temporal coordination between multiple individuals, by triggering synchronized motor responses in one individual to perceived sounds produced by another. We will now turn to a body of literature that explored this issue directly.

28.2.2 Sensorimotor Coordination, Prediction and Social Processes

The motor responses to musical rhythm and beat may drive music-induced movement, such as when we tap our feet or sway our bodies to music [28.4, 67]. Likewise, when considering musically trained individuals, these responses may facilitate interpersonal coordination, as when performing music with another individual, where the shared goal is to maintain synchrony [28.74]. Synchronizing with another individual often comes easily, and has proven to be an unintentional phenomenon in many contexts. For instance, two people seated in rocking chairs tend to involuntarily synchronize their rocking frequencies [28.75] or audiences clapping tend to fall into unison [28.76]. Recent research has used paced finger tapping as a model for exploring the neural mechanisms that permit individuals not only to predict others' action-timing, but also to adapt their movements accordingly [28.59, 77]. Thus, this research explores the social functions of action-perception coupling, and its relevance for sensorimotor synchronization [28.74].

In an fMRI study, *Fairhurst* et al. [28.78] simulated the temporal dynamics of a joint action by having musicians tapping with an adaptive auditory pacing signal: a computer-controlled virtual partner [28.78]. The degree of cooperation of the virtual partner was manipulated to elicit different levels of interagent synchrony with the participant ([28.79]). The results in-

dicated that a small change in the cooperation of the virtual partner led to a large-scale shift in activated brain networks. Uncooperative virtual partners who were overly adaptive led to poor synchronization and the activation of lateral prefrontal areas associated with cognitive control. Cooperative virtual partners who adapted to the participant's tap timing to an optimal degree facilitated precise interagent synchrony, accompanied by activation of cortical midline structures associated with socio-affective processes. These findings are thus potentially informative about the neural underpinnings of phenomena whereby actions of groups of individuals that are well synchronized feel effortless and promote cohesion [28.74]. Indeed, other research using behavioral methods has revealed that interpersonal synchronization can lead to prosocial behavior [28.80], feelings of affiliation [28.81], in-group cooperation [28.82] or trust [28.83]. Moreover, the findings of *Fairhurst* et al. [28.78] are consistent with other neuroimaging evidence for an association between experiencing interpersonal synchronization and activity of brain areas implicated in reward (such as the caudate) [28.84].

Successful coordination, however, it not entirely an automatic process, but might vary depending on an individual's skills, such as the capacity to predict others' behavior. The anticipatory mechanisms that enable these predictions have been associated with motor-related brain regions (see Sects. 28.1.3 and 28.2.1) and, in part, might be independent of musical background. However, the way in which these mechanisms operate in the context of a sensorimotor synchronization task might vary across individuals according to personal tendencies, abilities and/or cognitive strategies adopted [28.85]. In a relevant study, *Pecenka* and *Keller* [28.86] measured individual differences in musicians' temporal prediction abilities in a task that required sensorimotor synchronization with auditory pacing sequences that contained tempo changes (accelerando and decelerando) resembling those found in expressively timed music [28.86]. Based on these estimates of temporal prediction, individuals were classified as *high predictors*, i. e., individuals whose tap time series indicated that they anticipated the tempo changes, and *low predictors*, i. e., those who tended to track the tempo changes at a lag [28.87]. These prediction/tracking tendencies were not related to participants' motivation and, importantly, were stable when tested after several months. Next, *Pecenka* and *Keller* [28.86] assessed the musicians' abilities to tap in synchrony with one another in dyads comprised of two high predictors, two low predictors, and mixed pairs. Consistent with the hypothesis that that prediction is a key requirement for achieving good coordination, dyads composed of high predictors were more precise than low predictors and mixed pairs.

In order to explore the brain networks underlying these temporal prediction abilities, a subsequent fMRI study [28.88] manipulated cognitive load (using a visual working memory task of variable complexity) while musicians tapped along with tempo-changing pacing sequences. These sequences resembled another human's actions to the extent that their temporal profiles matched those found in expressive music performance. Estimates of the degree to which musicians predicted the tempo changes indicated that prediction ability decreased gradually with increasing cognitive load, and the related reduction in sensorimotor synchronization accuracy was accompanied by a modulation of a network of brain areas including premotor and motor regions, SMA, inferior parietal cortex and the cerebellum. It is noteworthy that some of these areas are considered to be crucial for achieving audio-motor synchronization, as indicated by TMS studies showing that interfering with the ventral [28.89] or dorsal [28.90] premotor cortex causes a drop in synchronization accuracy, which may be compensated for by an increase in activity of the cerebellum [28.89].

Another important component of successful interpersonal coordination in music performance is the capacity to adapt to another's action timing. Even if one can predict the timing of another's actions accurately, synchronization may be less than perfect if these predictions are not used effectively to guide self-generated behavior or due to biological noise in central and peripheral nervous systems. It is therefore necessary to employ adaptive timing mechanisms that facilitate interpersonal coordination by correcting for synchronization errors in a reactive fashion (see [28.77]). A recent study [28.91] indicates that the capacity for adaptive timing may be a factor that influences musical ensemble coordination by affecting who adapts to whom and to what degree. *Fairhurst* et al. [28.91] conducted an fMRI study in which musicians were asked to tap along with a computer-controlled virtual partner that was programmed to adapt to participants' tapping either in an optimal and reliable manner or in a way that rendered its timing unreliable and prone to tempo drift. The musicians were classified into two groups – *leaders* and *followers* – based on whether they found the interaction easier when they were required to exert control to maintain a steady tempo during interactions with unreliable virtual partners (leaders) or when they did not during interactions with reliable virtual partners (followers). This measure was correlated with scores on a questionnaire that assessed locus of control, with leaders being more likely than followers to perceive life events to a consequence of their own actions. Be-

havioral results indicated that leaders and followers employed different synchronization strategies. Leaders employed less adaptive timing than followers, suggesting that the former focused on maximizing the stability of their own tap tempo at the expense of precise synchrony with the virtual partner. Followers, conversely, employed more adaptive timing and thereby prioritized synchrony with the partner over the stability of their own tempo. This difference between groups was reflected on the neural level by higher activation in leaders of right-lateralized areas classically implicated in self-initiated movements, such as midcingulate cortex and pre-SMA, which is consistent with the prioritization of self-generated actions. Results also suggested that the precuneus, a brain area relevant to the integration of external and self-generated information [28.92], is sensitive to the distinction between focusing on actions of the self versus interactions between self and other.

Taken together, the studies reviewed in this section represent an emerging field of research that holds potential for understanding not only how motor brain regions entrain to temporal properties of music (Sect. 28.2.1), but also how these activations become relevant in the context of a social interaction such as interpersonal coordination – including joint musical performance. The studies reviewed in this section have explored the behavioral relevance of audio-motor coupling processes by which the human brain entrains to properties of music and generates predictions about the timing of others' upcoming actions. It was discussed how this mechanism might facilitate social interactions by bringing about the effortless synchronization of movements among members of a dyad or larger group. In the case of musically trained individuals, this mechanism is likely to play a significant role in enabling the precise interpersonal coordination that characterizes musical ensemble performance. Yet, these processes are subject to a certain degree of interindividual variability, most likely depending both on basic sensorimotor skills (e.g., [28.86, 88]) and on cognitive strategies [28.91]. Predictive and adaptive abilities appear to be key capacities for achieving precise interpersonal coordination, but the manner in which these are employed in the context of joint actions vary from one person to another, implying that not every dyad interacts in the same way.

28.2.3 Action Representation Mechanisms in Musical Interaction

Predicting and adapting to the timing of another individual are necessary but not sufficient skills for achieving temporally precise interpersonal coordination, e.g., in a joint musical performance. Another important skill that deserves consideration is the capacity to form integrated representations of self- and other-related actions. Such corepresentations are crucial because, in real-life situations, the actions produced by interacting individuals are often *complementary* rather than identical. Therefore, the goal in complementary joint action is not just to produce a movement, but the *appropriate* movement, in synchrony with another. This presumably requires dedicated mechanisms underpinning the integration of self-initiated action, which are generated internally, with others' actions, which are exogenous and, as such, can only be accessed via perceptual input. Action–perception coupling has been proposed a mechanism that potentially fulfills this function in joint action generally [28.93–95], as well as in musical joint action in particular [28.15, 96–98].

Behavioral research on music performance suggests that expert musicians form representations of self- and other-related actions, and that these representations are influenced by properties of the individual's own motor system. Evidence for this comes from studies demonstrating that pianists synchronize better with recordings of themselves than with those of other musicians [28.99] and with pianists who are well matched in terms of preferred performance tempo than with pianists who are less well matched in preferred tempo [28.100]. Furthermore, it has been shown that practicing a coperformer's part can in fact be detrimental to interpersonal coordination because in this case predictions about microtiming in the other's part are based upon one's own playing style, which may differ from the coperformer's style [28.101]. These findings suggest that musicians form representations of others' parts [28.16] that allow one individual to simulate another's actions. Thus, the manner in which a performer would execute a given piece strongly influences the way in which the performer synchronizes with another's performance of the piece. While this suggests that representations of others' actions are generated by means of a simulation process [28.14, 98] within one's motor system, this was not tested directly in the above-mentioned studies as they did not employ brain measures.

To address this, in a single-pulse TMS study, *Novembre* et al. [28.102] investigated the representation of self- and other-related actions in the context of a musical joint action paradigm [28.102]. Pianists learned to perform several pieces bimanually prior to the experiment. During the experiment, they were asked to perform the right hand part of the piece, while the left hand part was either not performed, or believed to be played by a coperformer hidden behind a screen (while the pianists were actually listening to a recording). This paradigm was intended to lead to a corepresentation of the left-hand part, reflecting

either the self or the coperformer. The authors examined action representation processes related to the left-hand part by stimulating the right motor cortex (using single-pulse TMS), and observing changes in the motor evoked potentials (MEPs) recorded from the resting left arm [28.27]. Results indicated that MEP amplitude was larger when the participant believed that he/she was interacting with a (hidden) coperformer. Remarkably, this effect persisted in a subsequent session in which neither visual nor auditory feedback from the coperformer were provided (though the participants were led to believe that the hidden coperformer was nevertheless playing), and was larger in individuals possessing stronger empathic traits (i. e., perspective taking skills, see [28.103]).

Thus, the study by *Novembre* et al. [28.102] indicates directly that ensemble musicians form *motor* representations of their ensemble members in the context of joint action, which is consistent with the behavioral evidence reported above. Moreover, this study suggests that these representations have an intrinsic social component, as: 1) perceptual feedback is not a prerequisite for corepresentation; and 2) individuals who are more prone to take the perspective of others form stronger corepresentations.

A study by *Loehr* et al. [28.104] used a joint musical performance paradigm similar to that employed by *Novembre* et al. [28.102] to investigate self and other monitoring and integration while EEG was recorded from pairs of pianists simultaneously [28.104]. The pianists learned to play both the left- and right-hand parts of musical pieces, and were then asked to perform one part each (while hearing and seeing each other's actions). The experimenters manipulated the auditory feedback from either pianist by creating a mismatch between piano keystrokes and produced tones. The mismatch either did or did not affect the harmony between the players' parts, hence permitting the differentiation of processes related to monitoring the self's performance and the joint action outcome. Altered outcomes elicited a feedback-related negativity irrespective of whether it occurred in the pianist's own or the partner's part, and a P300 with higher amplitude when the alteration was related to the pianist's own part. Crucially, the P300 had higher amplitude if it affected the joint outcome compared to the individual outcome, indicating that this task led to the emergence of integrated representations of self- and other-related actions.

A further study by *Novembre* et al. [28.105] used another modification of the virtual piano duo paradigm [28.102] to explore the extent to which motor representations of ensemble members support efficient temporal coordination between musicians [28.105]. To this end, pianists were required to adapt with the right hand to tempo changes contained in a recording of the left-hand part. The left-hand part either had or had not been practiced before the experiment in order to manipulate whether or not a motor representation was formed. In order to interfere with the representation of the left-hand part (which was practiced, but not performed), repetitive (double-pulse) TMS was used to disrupt the neural processing in the right primary motor cortex [28.106, 107], and tempo adaptation accuracy was measured following it. It was shown that interfering with the motor representation of the left-hand part affected temporal adaptation only when the part had been practiced (and therefore could be motorically represented). Moreover, this interference was stronger in individuals with high perspective taking skills, which is noteworthy given that *Novembre* et al. [28.102] demonstrated that these individuals also form stronger representations of others' action. This finding is also consistent with other accounts that postulate the relevance of empathic and perspective taking skills in the context of interactions between musicians [28.88, 108]. Thus, the results of *Novembre* et al. [28.105] provided evidence that motor representation processes might be a means used by musicians to maintain synchrony with one another [28.109]. This conclusion was supported, and further extended, by a follow-up experiment in which the same paradigm was employed in the context of a turn-taking task between a pianist and a virtual partner [28.110]. By delivering paired-pulses around the turn switch, the authors of the study were able to transiently impair the pianists' entry accuracy. This effect was specifically observed when the dorsal premotor cortex was targeted (while stimulating SMA did not yield any effect), and only to the extent that the part executed by the virtual partner had previously been rehearsed.

It can be noted that the studies reviewed above [28.102, 104, 105, 110] used sensorimotor training tasks as a means to build representations of self- and other-related actions. This is an important detail in that it suggests the particular relevance of the body of studies reviewed earlier in Sect. 28.1.2, where it was shown that listening to or watching the performance of a trained musical piece leads initially to the formation, and later to the activation, of motor representations in the musician's brain. Considering this, it appears clear that this recent research extends previous work by showing that action-perception coupling processes might support interaction between ensemble musicians [28.15, 96, 98].

28.3 Conclusion and Perspectives

In conclusion, we have considered the relationship between music and action from two perspectives. Firstly, in terms of how musical training can change the way in which music is perceived. Secondly, in terms of music inducing motor processes in a listener. For both, we noticed that sensory and motor brain regions become coupled in the human brain in response to music, and that this coupling underpins predictive computations that scaffold the ability to synchronize and adapt movements in response to a musical stimulus. This mechanism is likely to support the human ability not only to dance, or move, in response to music, but also to coordinate with others if music has to be jointly performed. This suggests that the relationship between music and action can be seen as a profoundly social phenomenon, for which individuals interact with one another through sounds and body movements [28.111]. This view in a sense inverts the way in which the relationship between music and action has been traditionally conceived. Progressing from increasing focus on technical and cultural factors that influence music production, we might soon witness a growing interest in the natural, intuitive, and spontaneous skills that make musical action, interaction and participation such a universal phenomenon [28.112].

References

28.1 B.S. Kisilevsky, S.M.J. Hains, A.-Y. Jacquet, C. Granier-Deferre, J.P. Lecanuet: Maturation of fetal responses to music, Dev. Sci. **7**, 550–559 (2004)

28.2 J. Phillips-Silver, L.J. Trainor: Feeling the beat: Movement influences infant rhythm perception, Science **308**, 1430 (2005)

28.3 M. Zentner, T. Eerola: Rhythmic engagement with music in infancy, Proc. Natl. Acad. Sci. USA **107**, 5768–5773 (2010)

28.4 J. Stupacher, M.J. Hove, G. Novembre, S. Schütz-Bosbach, P.E. Keller: Musical groove modulates motor cortex excitability: A TMS investigation, Brain Cogn. **82**, 127–136 (2013)

28.5 A. Pascual-Leone: The brain that plays music and is changed by it, Ann. N.Y. Acad. Sci. **930**, 315–329 (2001)

28.6 R.J. Zatorre, J.L. Chen, V.B. Penhune: When the brain plays music: Auditory-motor interactions in music perception and production, Nat. Rev. Neurosci. **8**, 547–558 (2007)

28.7 S.C. Herholz, R.J. Zatorre: Musical training as a framework for brain plasticity: Behavior, function, and structure, Neuron **76**, 486–502 (2012)

28.8 W. Prinz: A common-coding approach to perception and action. In: *Relationships Between Perception and Action: Current Approaches*, ed. by O. Neumann, W. Prinz (Springer, Berlin, New York 1990) pp. 167–201

28.9 B. Hommel, J. Müsseler, G. Aschersleben, W. Prinz: The Theory of Event Coding (TEC): A framework for perception and action planning, Behav. Brain. Sci. **24**, 849–878 (2001), discussion 878–937

28.10 G. Rizzolatti, L. Craighero: The mirror-neuron system, Annu. Rev. Neurosci. **27**, 169–192 (2004)

28.11 G. Rizzolatti, C. Sinigaglia: The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations, Nat. Rev. Neurosci. **11**, 264–274 (2010)

28.12 D.M. Wolpert, M. Kawato: Multiple paired forward and inverse models for motor control, Neural Netw. **11**, 1317–1329 (1998)

28.13 D.M. Wolpert, Z. Ghahramani: Computational principles of movement neuroscience, Nat. Neurosci. **3**, 1212–1217 (2000)

28.14 D.M. Wolpert, K. Doya, M. Kawato: A unifying computational framework for motor control and social interaction, Philos. Trans. R. Soc. Lond. B **358**(1431), 593–602 (2003)

28.15 P.E. Keller: Mental imagery in music performance: Underlying mechanisms and potential benefits, Ann. N.Y. Acad. Sci. **1252**, 206–213 (2012)

28.16 P.E. Keller, G. Novembre, J. Loehr: Musical ensemble performance: Representing self, other, and joint action outcomes. In: *Hared Representations: Sensorimotor Foundations of Social Life*, ed. by E. Cross, S. Obhi (Cambridge Univ. Press, Cambridge 2016)

28.17 U.C. Drost, M. Rieger, M. Brass, T.C. Gunter, W. Prinz: When hearing turns into playing: Movement induction by auditory stimuli in pianists, Q. J. Exp. Psychol. Sect. A **58**, 1376–1389 (2005)

28.18 U.C. Drost, M. Rieger, M. Brass, T.C. Gunter, W. Prinz: Action-effect coupling in pianists, Psychol. Res. **69**, 233–241 (2005)

28.19 U.C. Drost, M. Rieger, W. Prinz: Instrument specificity in experienced musicians, Q. J. Exp. Psychol. (Hove) **60**, 527–533 (2007)

28.20 P.E. Keller, I. Koch: Exogenous and endogenous response priming with auditory stimuli, Adv. Cogn. Psychol. **2**, 269–276 (2006)

28.21 P.E. Keller, I. Koch: The planning and execution of short auditory sequences, Psychon. Bull. Rev. **13**, 711–716 (2006)

28.22 P.E. Keller, I. Koch: Action planning in sequential skills: Relations to music performance, Q. J. Exp. Psychol. (Hove) **61**, 275–291 (2008)

28.23 P.E. Keller, S.B. Dalla, I. Koch: Auditory imagery shapes movement timing and kinematics: Evi-

dence from a musical task, J. Exp. Psychol. Hum. Percept. Perform. **36**, 508–513 (2010)
28.24 P.Q. Pfordresher, C. Palmer: Effects of hearing the past, present, or future during music performance, Percept. Psychophys. **68**, 362–376 (2006)
28.25 P.Q. Pfordresher: Auditory feedback in music performance: The role of melodic structure and musical skill, J. Exp. Psychol. Hum. Percept. Perform. **31**, 1331–1345 (2005)
28.26 J. Haueisen, T.R. Knösche: Involuntary motor activity in pianists evoked by music perception, J. Cogn. Neurosci. **13**, 786–792 (2001)
28.27 A. D'Ausilio, E. Altenmüller, M. Olivetti Belardinelli, M. Lotze: Cross-modal plasticity of the motor cortex while listening to a rehearsed musical piece, Eur. J. Neurosci. **24**, 955–958 (2006)
28.28 M. Bangert, T. Peschel, G. Schlaug, M. Rotte, D. Drescher, H. Hinrichs, H.-J. Heinze, E. Altenmüller: Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction, Neuroimage **30**, 917–926 (2006)
28.29 B. Haslinger, P. Erhard, E. Altenmüller, U. Schroeder, H. Boecker, A.O. Ceballos-Baumann: Transmodal sensorimotor networks during action observation in professional pianists, J. Cogn. Neurosci. **17**, 282–293 (2005)
28.30 T. Hasegawa, K.-I. Matsuki, T. Ueno, Y. Maeda, Y. Matsue, Y. Konishi, N. Sadato: Learned audio-visual cross-modal associations in observed piano playing activate the left planum temporale. An fMRI study, Cogn. Brain Res. **20**, 510–518 (2004)
28.31 A. Lahav, E. Saltzman, G. Schlaug: Action representation of sound: Audiomotor recognition network while listening to newly acquired actions, J. Neurosci. **27**, 308–314 (2007)
28.32 M. Bangert, E. Altenmüller: Mapping perception to action in piano practice: A longitudinal DC-EEG study, BMC Neuroscience **4**, 26 (2003)
28.33 A. Engel, M. Bangert, D. Horbank, B.S. Hijmans, K. Wilkens, P.E. Keller, C. Keysers: Learning piano melodies in visuo-motor or audio-motor training conditions and the neural correlates of their cross-modal transfer, NeuroImage **63**(2), 966–978 (2012)
28.34 M. Candidi, L.M. Sacheli, I. Mega, S.M. Aglioti: Somatotopic mapping of piano fingering errors in sensorimotor experts: TMS studies in pianists and visually trained musically naives, Cereb. Cortex **24**, 435–443 (2014)
28.35 I. Koch, P. Keller, W. Prinz: The ideomotor approach to action control: Implications for skilled performance, Int. J. Sport Exerc. Psychol. **2**, 362–375 (2004)
28.36 J.L. Chen, C. Rae, K.E. Watkins: Learning to play a melody: An fMRI study examining the formation of auditory-motor associations, NeuroImage **59**, 1200–1208 (2012)
28.37 C. Maidhof, M. Rieger, W. Prinz, S. Koelsch: Nobody is perfect: ERP effects prior to performance errors in musicians indicate fast monitoring processes, PLoS One **4**, e5032 (2009)
28.38 M.H. Ruiz, H.-C. Jabusch, E. Altenmüller: Detecting wrong notes in advance: Neuronal correlates of error monitoring in pianists, Cereb. Cortex **19**, 2625–2639 (2009)
28.39 K.A. Kiehl, P.F. Liddle, J.B. Hopfinger: Error processing and the rostral anterior cingulate: An event-related fMRI study, Psychophysiology **37**, 216–223 (2000)
28.40 J.G. Kerns, J.D. Cohen, A.W. MacDonald, R.Y. Cho, V.A. Stenger, C.S. Carter: Anterior cingulate conflict monitoring and adjustments in control, Science **303**, 1023–1026 (2004)
28.41 D.M. Wolpert, Z. Ghahramani, M.I. Jordan: An internal model for sensorimotor integration, Science **269**, 1880–1882 (1995)
28.42 C. Maidhof, N. Vavatzanidis, W. Prinz, M. Rieger, S. Koelsch: Processing expectancy violations during music performance and perception: An ERP study, J. Cogn. Neurosci. **22**, 2401–2413 (2010)
28.43 M.H. Ruiz, F. Strübing, H.-C. Jabusch, E. Altenmüller: EEG oscillatory patterns are associated with error prediction during music performance and are altered in musician's dystonia, NeuroImage **55**, 1791–1803 (2011)
28.44 J.M. Kilner, S.N. Baker, S. Salenius, V. Jousmäki, R. Hari, R.N. Lemon: Task-dependent modulation of 15-30 Hz coherence between rectified EMGs from human hand and forearm muscles, J. Physiol. **516**, 559–570 (1999)
28.45 M. Feurra, G. Bianco, E. Santarnecchi, M. Del Testa, A. Rossi, S. Rossi: Frequency-dependent tuning of the human motor system induced by transcranial oscillatory potentials, J. Neurosci. **31**, 12165–12170 (2011)
28.46 J.M. Kilner, K.J. Friston, C.D. Frith: The mirror-neuron system: A Bayesian perspective, Neuroreport **18**, 619–623 (2007)
28.47 H. Lee, U. Noppeney: Long-term music training tunes how the brain temporally binds signals from multiple senses, Proc. Natl. Acad. Sci. USA **2011**, 1–10 (2011)
28.48 B. Maess, S. Koelsch, T.C. Gunter, A.D. Friederici: Musical syntax is processed in Broca's area: An MEG study, Nat. Neurosci. **4**, 540–545 (2001)
28.49 S. Koelsch, W.A. Siebel: Towards a neural basis of music perception, Trends Cogn. Sci. **9**, 578–584 (2005)
28.50 S. Koelsch, B.-H. Schmidt, J. Kansok: Effects of musical expertise on the early right anterior negativity: An event-related brain potential study, Psychophysiology **39**, 657–663 (2002)
28.51 G. Novembre, P.E. Keller: A grammar of action generates predictions in skilled musicians, Conscious Cogn. **20**, 1232–1243 (2011)
28.52 D. Sammler, G. Novembre, S. Koelsch, P.E. Keller: Syntax in a pianist's hand: ERP signatures of 'embodied' syntax processing in music, Cortex **49**, 1325–1339 (2013)
28.53 R. Bianco, G. Novembre, P.E. Keller, F. Scharf, A.D. Friederici, A. Villringer, D. Sammler: Syntax in action has priority over movement selection in piano playing: An ERP study, J. Cogn. Neurosci. **28**,

41–54 (2016)
28.54 R. Bianco, G. Novembre, P.E. Keller, K. Seung-Goo, F. Scharf, A.D. Friederici, A. Villringer, D. Sammler: Neural networks for musical syntax in perception and in action, NeuroImage **142**, 454–464 (2016)
28.55 V. Krieghoff, M. Brass, W. Prinz, F. Waszak: Dissociating what and when of intentional actions, Front Hum. Neurosci. **3**, 3 (2009)
28.56 L. Fadiga, L. Craighero, A. D'Ausilio: Broca's area in language, action, and music, Ann. N.Y. Acad. Sci. **1169**, 448–458 (2009)
28.57 F. Pulvermüller, L. Fadiga: Active perception: Sensorimotor circuits as a cortical basis for language, Nat. Rev. Neurosci. **11**, 351–360 (2010)
28.58 B.H. Repp: Sensorimotor synchronization: A review of the tapping literature, Psychon. Bull. Rev. **12**, 969–992 (2005)
28.59 B.H. Repp, Y.-H. Su: Sensorimotor synchronization: A review of recent research (2006-2012), Psychon. Bull. Rev. **20**, 403–452 (2013)
28.60 J.L. Chen, R.J. Zatorre, V.B. Penhune: Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms, NeuroImage **32**, 1771–1781 (2006)
28.61 J.L. Chen, V.B. Penhune, R.J. Zatorre: Listening to musical rhythms recruits motor regions of the brain, Cereb. Cortex **18**, 2844–2854 (2008)
28.62 J.L. Chen, V.B. Penhune, R.J. Zatorre: Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training, J. Cogn. Neurosci. **20**, 226–239 (2008)
28.63 J.A. Grahn, M. Brett: Rhythm and beat perception in motor areas of the brain, J. Cogn. Neurosci. **19**, 893–906 (2007)
28.64 J.A. Grahn, J.B. Rowe: Feeling the beat: Premotor and striatal interactions in musicians and nonmusicians during beat perception, J. Neurosci. **29**, 7540–7548 (2009)
28.65 J.A. Grahn, J.B. Rowe: Finding and feeling the musical beat: Striatal dissociations between detection and prediction of regularity, Cereb. Cortex **23**(4), 913–921 (2013)
28.66 S. Kung, J.L. Chen, R.J. Zatorre, V.B. Penhune: Interacting cortical and basal ganglia networks underlying finding and tapping to the musical beat, J. Cogn. Neurosci. **25**(3), 401–420 (2013)
28.67 P. Janata, S.T. Tomic, J.M. Haberman: Sensorimotor coupling in music and the psychology of the groove, J. Exp. Psychol. Gen. **141**(1), 54–75 (2012)
28.68 T. Fujioka, L.J. Trainor, E.W. Large, B. Ross: Internalized timing of isochronous sounds is represented in neuromagnetic Beta oscillations, J. Neurosci. **32**, 1791–1802 (2012)
28.69 R. Bartolo, H. Merchant: Oscillations are linked to the initiation of sensory-cued movement sequences and the internal guidance of regular tapping in the monkey, J. Neurosci. **35**, 4635–4640 (2015)
28.70 S. Nozaradan, I. Peretz, M. Missal, A. Mouraux: Tagging the neuronal entrainment to beat and meter, J. Neurosci. **31**, 10234–10240 (2011)
28.71 S. Nozaradan, Y. Zerouali, I. Peretz, A. Mouraux: Capturing with EEG the neural entrainment and coupling underlying sensorimotor synchronization to the beat, Cereb. Cortex **25**(3), 736–747 (2015), https://doi.org/10.1093/cercor/bht261
28.72 S. Nozaradan, I. Peretz, P.E. Keller: Individual differences in rhythmic cortical entrainment correlate with predictive behavior in sensorimotor synchronization, Sci. Rep. **6**, 20612 (2016)
28.73 R.I. Schubotz: Prediction of external events with our motor system: Towards a new framework, Trends Cogn. Sci. **11**, 211–218 (2007)
28.74 P.E. Keller, G. Novembre, M.J. Hove: Rhythm in joint action: Psychological and neurophysiological mechanisms for real-time interpersonal coordination, Philos. Trans. R. Soc. Lond. B. Biol. Sci. (2014), https://doi.org/10.1098/rstb.2013.0394
28.75 M.J. Richardson, K.L. Marsh, R.W. Isenhower, J.R.L. Goodman, R.C. Schmidt: Rocking together: Dynamics of intentional and unintentional interpersonal coordination, Hum. Mov. Sci. **26**, 867–891 (2007)
28.76 Z. Néda, E. Ravasz, Y. Brechet, T. Vicsek, A.L. Barabási: The sound of many hands clapping, Nature **403**, 849–850 (2000)
28.77 M.C.M. van der Steen, P.E. Keller: The ADaptation and Anticipation Model (ADAM) of sensorimotor synchronization, Front. Hum. Neurosci. **7**, 253 (2013)
28.78 M.T. Fairhurst, P. Janata, P.E. Keller: Being and feeling in sync with an adaptive virtual partner: Brain mechanisms underlying dynamic cooperativity, Cereb. Cortex **23**, 2592–2600 (2013)
28.79 B.H. Repp, P.E. Keller: Sensorimotor synchronization with adaptively timed sequences, Hum. Mov. Sci. **27**, 423–456 (2008)
28.80 S. Kirschner, M. Tomasello: Joint music making promotes prosocial behavior in 4-year-old children, Evol. Hum. Behav. **31**, 354–364 (2010)
28.81 M.J. Hove, J.L. Risen: It's all in the timing: Interpersonal synchrony increases affiliation, Soc. Cogn. **27**, 949–960 (2009)
28.82 S.S. Wiltermuth, C. Heath: Synchrony and cooperation, Psychol. Sci. **20**, 1–5 (2009)
28.83 J. Launay, R.T. Dean, F. Bailes: Synchronization can influence trust following virtual interaction, Exp. Psychol. **60**, 53–63 (2012)
28.84 I. Kokal, A. Engel, S. Kirschner, C. Keysers: Synchronized drumming enhances activity in the caudate and facilitates prosocial commitment – if the rhythm comes easily, PLoS One **6**, e27272 (2011)
28.85 P.F. Mills, M.C.M. van der Steen, B.G. Schultz, P.E. Keller: Individual differences in temporal anticipation and adaptation during sensorimotor synchronization, Timing Time Percept. **3**, 13–31 (2015)
28.86 N. Pecenka, P.E. Keller: The role of temporal prediction abilities in interpersonal sensorimotor synchronization, Exp. Brain Res. **211**, 505–515 (2011)
28.87 N. Pecenka, P.E. Keller: The relationship between auditory imagery and musical synchronization

abilities in musicians. In: *Proc. 7th Triennial Conf. Eur. Soc. Cog. Sci. Music (ESCOM)* (2009) pp. 409–415

28.88 N. Pecenka, A. Engel, P. Keller: Neural correlates of auditory temporal predictions during sensorimotor synchronization, Front. Hum. Neurosci. **7**, 1–16 (2013)

28.89 K. Kornysheva, R.I. Schubotz: Impairment of auditory–motor timing and compensatory reorganization after ventral premotor cortex stimulation, PLoS One **6**, e21421 (2011)

28.90 F. Giovannelli, I. Innocenti, S. Rossi, A. Borgheresi, A. Ragazzoni, G. Zaccara, M.P. Viggiano, M. Cincotta: Role of the dorsal premotor cortex in rhythmic auditory-motor entrainment: A perturbational approach by rTMS, Cereb. Cortex **24**, 1009–1016 (2014)

28.91 M.T. Fairhurst, P. Janata, P.E. Keller: Leading the follower: An fMRI investigation of dynamic cooperativity and leader-follower strategies in synchronization with an adaptive virtual partner, NeuroImage **84**, 688–697 (2014)

28.92 A.E. Cavanna, M. Trimble: The precuneus: A review of its functional anatomy and behavioural correlates, Brain **129**, 564–583 (2006)

28.93 N. Sebanz, G. Knoblich, W. Prinz: How two share a task: Corepresenting stimulus–response mappings, J. Exp. Psychol. Hum. Percept. Perform. **31**, 1234–1246 (2005)

28.94 S. Schütz-Bosbach, W. Prinz: Perceptual resonance: Action-induced modulation of perception, Trends Cogn. Sci. **11**, 349–355 (2007)

28.95 G. Knoblich, S. Butterfill, N. Sebanz: Psychological research on joint action: Theory and data. In: *The Psychology of Learning and Motivation*, ed. by B. Ross (Academic Press, Burlington 2011) pp. 59–101

28.96 P.E. Keller: Joint action in music performance. In: *Enacting Intersubjectivity: A Cognitive and Social Perspective on the Study of Interactions*, ed. by F. Morganti, A. Carassa, G. Riva (IOS Press, Amsterdam 2008) pp. 205–221

28.97 P.E. Keller: Ensemble performance: Interpersonal alignment of musical expression. In: *Expressiveness in Music Performance: Empirical Approaches Across Styles and Cultures*, ed. by D. Fabian, R. Timmers, E. Schubert (Oxford Univ. Press, Oxford 2014)

28.98 G. Novembre, P.E. Keller: A conceptual review on action-perception coupling in the musicians' brain: What is it good for?, Front. Hum. Neurosci. **8**, 1–11 (2014)

28.99 P.E. Keller, G. Knoblich, B.H. Repp: Pianists duet better when they play with themselves: on the possible role of action simulation in synchronization, Conscious Cogn. **16**, 102–111 (2007)

28.100 J.D. Loehr, C. Palmer: Temporal coordination between performing musicians, Q. J. Exp. Psychol. **64**(11), 2153–2167 (2011)

28.101 M. Ragert, T. Schroeder, P.E. Keller: Knowing too little or too much: The effects of familiarity with a co-performer's part on interpersonal coordination in musical ensembles, Front. Psychol. **4**, 368 (2013)

28.102 G. Novembre, L.F. Ticini, S. Schütz-Bosbach, P.E. Keller: Distinguishing self and other in joint action. Evidence from a musical paradigm, Cereb. Cortex **22**, 2894–2903 (2012)

28.103 M.H. Davis: Measuring individual differences in empathy: Evidence for a multidimensional approach, J. Pers. Soc. Psychol. **44**, 113–126 (1983)

28.104 J.D. Loehr, D. Kourtis, C. Vesper, N. Sebanz, G. Knoblich: Monitoring individual and joint action outcomes in duet music performance, J. Cogn. Neurosci. **25**, 1049–1061 (2013)

28.105 G. Novembre, L.F. Ticini, S. Schutz-Bosbach, P.E. Keller: Motor simulation and the coordination of self and other in real-time joint action, Soc. Cogn. Affect Neurosci. **9**, 1062–1068 (2014)

28.106 N.J. Rice, E. Tunik, S.T. Grafton: The anterior intraparietal sulcus mediates grasp execution, independent of requirement to update: new insights from transcranial magnetic stimulation, J. Neurosci. **26**, 8176–8182 (2006)

28.107 N.R. Cohen, E.S. Cross, E. Tunik, S.T. Grafton, J.C. Culham: Ventral and dorsal stream contributions to the online control of immediate and delayed grasping: A TMS approach, Neuropsychologia **47**, 1553–1562 (2009)

28.108 C. Babiloni, P. Buffo, F. Vecchio, N. Marzano, C. Del Percio, D. Spada, S. Rossi, I. Bruni, P.M. Rossini, D. Perani: Brains 'in concert': Frontal oscillatory alpha rhythms and empathy in professional musicians, Neuroimage **60**, 105–116 (2012)

28.109 G. Novembre, D. Sammler, P.E. Keller: Neural alpha oscillations index the balance between self-other integration and segregation in real-time joint action, Neuropsychologia **89**, 414–425 (2016)

28.110 L.V. Hadley, G. Novembre, P.E. Keller, M.J. Pickering: Causal role of motor simulation in turn-taking behavior, J. Neurosci. **35**, 16516–16520 (2015)

28.111 A. D'Ausilio, G. Novembre, L. Fadiga, P.E. Keller: What can music tell us about social interaction?, Trends Cogn. Sci. **19**, 1–4 (2015)

28.112 G. Novembre, M. Varlet, S. Muawiyath, C.J. Stevens, P.E. Keller: The E-music box: An empirical method for exploring the universal capacity for musical production and for social interaction through music, R. Soc. Open Sci. **2**, 150286 (2015)

29. Music and Emotions

Tuomas Eerola

The rapid rise in emotion research in psychology has brought forth a rich palette of concepts and tools for studying emotions expressed and induced by music. This chapter summarizes the current state of music and emotion research, starting with the fundamental definitions and the assumed structures of emotions (Sect. 29.2). A synthesis of the core affects, basic emotions and complex emotions is offered to clarify this complex landscape. A vital development for the field has been the introduction of a set of mechanisms and modifiers for the induction of emotion via music that are here connected to the structures of emotions (Sects. 29.3 and 29.4). Particular attention is given to challenges that are still waiting to be resolved, such as the cultural context and the situal context of music listening (Sect. 29.5).

29.1 The Rise of Music and Emotion Research

The question of how music induces emotions in listeners with such effortless grace is a puzzle worth solving not only for its myriad real-world applications, but also for its ability to address the fundamental reasons for the existence of music [29.1]. Emotions induced through music are intrinsic to most ceremonies [29.2] and may be used therapeutically [29.3], but research has not yet fully dissected the factors contributing to such effects and uses of music; in fact we are neither sure of the ways in which music engenders feelings, nor do we agree on the typology of emotions relevant for most musical episodes. Extensive summaries of music and emotion have been provided in recent dedicated books and journal articles [29.4–7]. This chapter not only reviews the topics in which consensus has emerged, but also attempts to combine some of the key themes that have perhaps become lost amongst the increasing diversity of theories and foci of research.

There has been a long-standing interest in the emotional pull of music. From ancient Greece (Aristotle, Plato) to the Enlightenment (Rousseau, Baumgarten), philosophers have attempted to account for this pull, as have the fathers of evolutionary thought (Charles Darwin) and contemporary psychology (Wilhelm Wundt). While the empirical study of emotions has been pro-

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_29

ceeding now for more than 100 years, the rise of its popularity in music research coincided with the cognitive paradigm loosening its grip on the range of permitted topics in the early 1990s. From this point on, softer topics such as emotions were allowed to flourish after several decades of focus on human information processing. Music research always had an eye on the emotions despite their unpopularity in psychology, and exceptional individuals had presented their views on the subject. For instance, Kate Hevner discovered the low-dimensional structure of affects in the 1930s, fifty years before *James Russell* [29.8] proposed the well-known affective circumplex model. Later, *Leonard B. Meyer* connected emotions to expectations [29.9], *Paul Farnsworth* developed a new emotion vocabulary [29.10], and *Daniel Berlyne* redefined the topic as a mission to find the objective properties of the stimulus that would relate to arousal [29.11].

The 1960s and early 1970s saw research that rediscovered the conceptual cartography of the emotional landscape in music, with *Rigg* [29.12], *Wedin* [29.13], and *Gabrielsson* [29.14] asking listeners to provide free verbal reports and adjective choices for the emotions represented by music examples. This line of research aimed to formulate a definitive emotion taxonomy for music, which even now has not been fully resolved. In the 1980s and 1990s, the field focused more on methodological innovations such as continuous ratings of emotions [29.15], and on specific emotional responses such as thrills [29.16] and strong experiences [29.17].

In the field of affective sciences, the 1990s witnessed a frenzy of activity when influential neuroscientists such as *Antonio Damasio* [29.18] and *Joseph LeDoux* [29.19] outlined their seminal theories of emotions. The neuroscientist *Jaak Panksepp* [29.20] soon applied these ideas to music, quickly followed by others [29.21]. The culmination of research up until the turn of the millennium was collated in the first handbook of the field [29.22], which solidified research terms and concepts. However, in the period that has followed, neither the field of affective sciences, nor the smaller subfield of music and emotion studies, have provided full answers to all fundamental issues.

This chapter comprises questions of *what* and *how* concerning emotions and music. The first section inquires into *what* the perceived and experienced emotions are; here the notions of core affects, basic emotions, and complex emotions are discussed. This is followed by a section exploring *how*, which offers a summary of mechanisms considered relevant for the emotions. Mechanisms are covered in parallel to emotion structures, and the two complementary aspects of emotions – perception and experience – are discussed within a uniform interpretative framework. In the final section, the current challenges are reviewed and discussed.

29.2 Structure of Emotions

To summarize the current state of research in music and emotions, a few definitions are first in order. The terminology used in the field is diverse, but the concentrated effort that created a dedicated handbook [29.4] has remedied the situation considerably. *Affect* is the broader domain encompassing *emotions*, *moods*, and *feelings*. Emotions, which are typically the focus of attention in this field, are distinguished from moods by their relatively short duration and moderate to high intensity, whereas moods tend to be longer and less intense than emotions. However, it has been suggested that this distinction is rather blurry and exceptions have been pointed out [29.23]. *Feelings* refer to the subjective component of the emotion whereas arousal is the physical activation of the autonomic nervous system. Under these categories, there are specialized emotional and physical reactions, such as *chills* or *goosebumps*, and, in addition, *strong emotions*, which can encompass a range of musical experiences [29.5]. Emotions have the capacity to span a broad range of topics such as motivation, preference, intensity, and affect reactivity [29.24], but these are rarely addressed in the context of music and will not be covered in this chapter.

Although consensus on the topic of emotions has been rather elusive [29.25], the main subcomponents of emotions are largely agreed on. These include (1) appraisal (one assesses a situation to be dangerous), (2) expression (one screams), (3) autonomic reaction (one starts to perspire), (4) action tendency (one moves away from the situation), and (5) feeling (one feels threatened), all of which occur more or less simultaneously ([29.26, pp. 6–8], [29.27]). Purely cognitive processing is not usually considered an emotion since all theoretical positions assume some form of arousal to distinguish emotions from processing of information [29.28]. The components occur at conceptually different levels and indeed, the outcomes of these processes are often described at different levels: physiology (changes in brain, hormonal or autonomic nervous system states), psychology (functions, appraisal, or recognition processes), or phenomenology (emotions as experienced). In this chapter, the levels are distinguished in terms

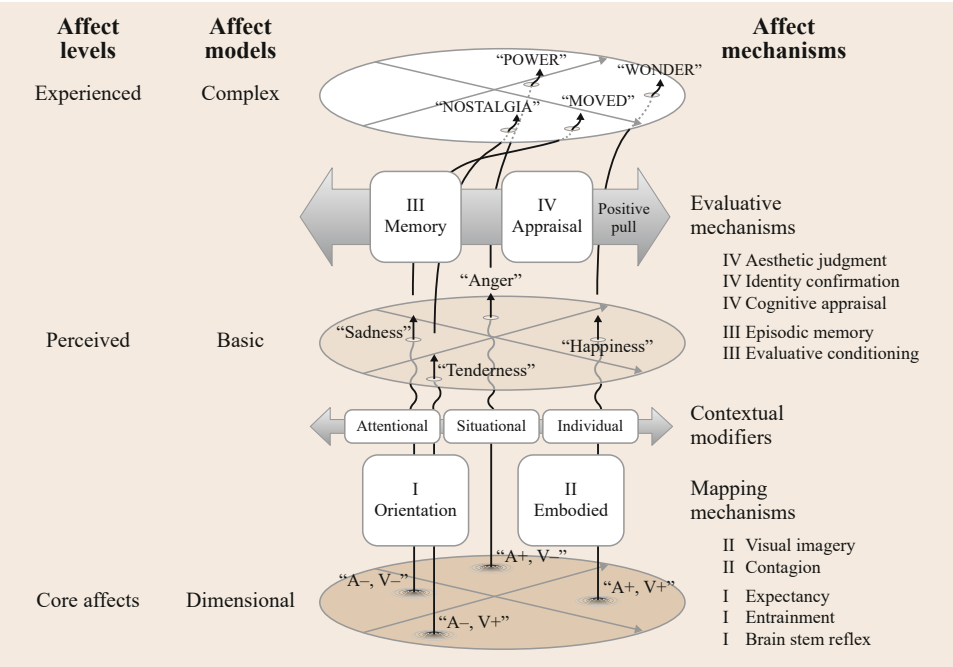


Fig. 29.1 Schematic illustration of affect concepts (levels, models, and mechanisms)

of the processes related to sensation, recognition and experience of emotions and this division shapes the models of emotions and mechanisms proposed at each level.

The aim here is to bridge the gap between what are now largely different avenues of research: the emotions perceived in music (also known as emotion recognition or expressed emotions) and emotions experienced or induced in the listener. The former was an early focus of the field [29.13, 29, 30]. In such studies, listeners were often instructed to describe the music in emotional terms (e.g., this music is sad) or describe what the music might have been expressing (e.g., this music expresses sadness). During the last decade, the emphasis has shifted towards explorations of how music makes the listeners feel. The distinction between recognized and induced emotions has been tempered by solid evidence of the way in which the two often overlap. Empirical evidence suggests that the distinction should be characterized as a question of intensity [29.31] rather than involving completely different processes, despite work considering the latter [29.32]. Although the perception and experience of emotions are intrinsically linked, increasingly refined theorizing and the proliferation of practical examples have taken the two categories in different directions. Here, instead, the intention is clarify the similarities between emotion recognition and induction processes in music by bringing the two into a framework that contains different levels of affects as well as series of mechanisms capable of producing them. This framework also attempts to locate the mechanisms within these different levels of affects.

The theoretical review of emotion models is organized with three explanatory levels of affects, starting from low-level core affects, proceeding to basic emotions, and ending with high-level, complex emotions. The first two levels refer mainly to processes involved in perception of emotions whereas the last one refers to induction of emotions, although this division risks oversimplifying matters. To help the reader interpret and keep track of the affect levels and how they are connected to mechanisms discussed in the next chapters, a schematic illustration of the key concepts is given in Fig. 29.1.

The left part of the illustration concerns the question of what the pertinent emotions are. *Affect levels* refer to distinctions between low-level sensed emotions (core affects), perceived emotions, and experienced emotions. This division corresponds approximately to the models of emotions conceived to date. The organization of affect levels as having low-level measurable properties capable of producing highly different conceptual interpretation is influenced by the hybrid model of emotions proposed by *Lisa Feldman Barrett* [29.33, 34]. In this model, the underlying physical machinery is best described by the dimensions (core affects) but the conscious interpretation of these is categorical, and influenced by the conceptual categories people have for emotions. The synthesis proposed here (Fig. 29.1) assigns different mechanisms of emotions to differ-

ent affect levels. This dynamic model claims that the way people use conceptual knowledge determines how they feel, and, due to variance in context, differences between individuals, and effects of high-level mechanisms, there is increasing variation in emotions at the highest complex and experiential level.

In the affective sciences in general, emotions have been theorized as organized along a few core dimensions (i. e., core affects), belonging to discrete categories (basic emotions), or having a complex, perhaps even a domain-specific structure. Each of these major descriptive schemes is applicable to music and will be explained in the three sections that follow.

29.2.1 Emotion Dimensions and Core Affects

A popular way to conceptualize emotions is to divide emotions into a continuum between positive and negative emotions. This dimensional approach, rooted in the work of Wilhelm Wundt [29.35], is best known as the circumplex model of emotions [29.8]. This bidimensional model assigns emotions as a mixture of two core dimensions, valence and arousal (hyphenated with A and V in Fig. 29.1), that represent two orthogonally situated continuums of pleasure-displeasure and activation-deactivation in the affective space. The circumplex model has received support in large studies of self-reported emotions [29.36], cross-cultural comparison [29.37], and psychometric studies (reviewed in [29.38]) and has thus been used in a large number of emotion studies in psychology as well as in systematic musicology. *Russell* and *Feldman Barrett* [29.34] have characterized the dimensions as *core affects*, to differentiate theirs from other dimensional models. The term *core affect* refers to the idea that such affects arise from the core of the body and in neural representations of body states. Core affects are assumed to present already in infants [29.39] and have psychological universality [29.40, 41].

Another way of expressing the dimensions of emotions has been to divide them into approach- versus avoidance-related emotions [29.42]. There have also been influential variants of the circumplex model such as the rotated circumplex model by *Watson* et al. [29.43] (titled Positive and Negative Affect Scale, or PANAS). *Thayer* [29.44] reorganized the arousal dimension into energetic arousal and tense arousal on the basis that separate psychobiological systems are responsible for energy and tension. Both formulations have been successfully applied to research on emotions in music [29.45, 46], but they remain, nevertheless, fairly unpopular in music research.

The main problem for any dimensional model is that emotions can sometimes be ambivalent [29.47, 48]. It is possible to feel both happy and sad at the same time, and indeed, empirical studies using tragicomic films [29.49] or music [29.31] seem to suggest that mixtures within the dimensions are relevant for emotional experiences. This problem is mitigated by postulating that such mixed emotions are co-occurring combinations of different activities within a dimension, or even rapid switching between the two [29.50].

29.2.2 Basic Emotions and Emotion Perception

Perhaps the most widely known way to organize emotions is to assume they are discrete categories, also referred to as *basic*, *primary*, or even *fundamental*. Such theories posit that all emotions can be derived from a limited set of innate and thus universal basic emotions, such as *fear, anger, disgust, sadness*, and *surprise* [29.51]. The actual number of categories and the label for the categories is still a source of debate, but the basic emotion model has gathered support from neural, cross-cultural, developmental, and physiological research spanning four decades [29.27]. Despite the popular appeal of such categories, doubts about their explanatory power has been raised, as, for instance, brain imaging studies have not yet delivered results consistent with the innate and distinct emotion categories [29.52]. In music, almost half of the studies focusing on emotions experienced have resorted to basic emotions [29.53]. This popularity is related to the fact that the categories are easy to use in emotion recognition studies, which are popular in developmental [29.54, 55] and production research [29.56, 57]. Similarly, when physiological or neural correlates of music-induced emotions are explored, basic emotions tend to be chosen [29.58–62], although this has started to change due to availability of emotion taxonomies made specifically for music-aroused experiences (described in the next section).

The dimensional and basic emotion models offer distinct ways to tackle musical emotions. However, when the two models are mapped onto each other in emotion expressed by music, the results have suggested that the models overlap considerably [29.63].

29.2.3 Complex Emotions and Emotion Experience

When people engage with artworks, or objects in nature, the emotional experiences are not easily explained as dimensions or discrete patterns of survival emotions [29.64]. Any fiction, including music, does not have direct material effects on the physical or psychological wellbeing of the individual in the same way that

events in everyday context do. This freedom of fiction to explore the expanses of the human mind rather than provide handy heuristics for reaction to the environment in everyday activities may expand the scope of possible emotions in fiction. It is perfectly plausible to think that emotions induced by music – or any art in general – are more contemplative, reflective and nuanced, or if we need to use one word – complex. A similar argument has been put forward to other complex emotions such as moral, social, or epistemic emotions that are not necessarily involved in everyday survival [29.27]. Such complex emotions are also more subject to cultural and social interpretations than basic emotions [29.41, 65].

In music, the desire to explain complex emotions has often been to topic of musicological writings [29.66, 67], but the empirical mapping of the topic began in the late 1960s, when *Kaarle Nordenstreng* [29.68] analyzed the structures underlying listener experiences across musical excerpts. A decade later *Edward Asmus* [29.69] carried out a large-scale study ($n = 2057$) of the affect terms relevant for music and proposed nine dimensions of affects accounting for the variety of emotion states. The decisive attempt to account for the emotional experiences induced by music was taken by *Marcel Zentner* and his colleagues [29.70]. They started with a comprehensive list of emotion terms relevant to music and validated these with iterations of surveys, and finally administered the refined lists to a festival audience. Next, the similarity between the terms was able to reduce the list into ten solid factors, which was further validated with separate ratings and confirmatory factor analysis. This final stage resulted in nine emotion factors, a model which is now known as GEMS (GEMS). The model emphasizes positive emotions (at least seven out of nine factors in the model) and provides factors particularly appropriate for contemplative emotions such as *wonder*, *nostalgia* and *transcendence*. For this reason, it fits well with the tradition of complex and aesthetic emotions, and it has been widely adopted in studies of emotional experiences related to music [29.71, 72]. It is worth noting that other proposals for the appropriate emotions induced by music have been offered as well, based on equally large samples of participants [29.73]. This proposal shares many of the factors of emotions such as *nostalgia*, *being moved*, and final differences may just be semantic and the final number of broad categories included.

To summarize, in this section I have discussed the previous models of music-related emotions and presented a new synthesis, outlined in Fig. 29.1. In this scheme, the low-level dimensional representations of core affects can be collapsed into basic emotion categories in perception of emotions using specific, well-known labels (e.g., happy, sad, etc.). However, at the level of experienced emotions, the conceptual act of labeling emotions is fundamentally modified by high-level mechanisms and modifiers as well as by language and culture. Therefore, the experiences as well as the labels applicable are different in emotion perception and experience. This fundamental difference between emotion processes and actual experience is of crucial importance here [29.33]. This perspective may explain some of the dissatisfaction scholars have had with the models designed to explain emotions as processes and objects of recognition (core affects and basic emotions) whereas others have declined to acknowledge the existence of music-specific emotions. The champion for the complex emotions is the GEMS model, but the differences between the complex and basic emotions becomes less of an issue if the differences in the processes are fully acknowledged.

29.3 Mechanisms and Modifiers of Emotions

The reasons and mechanisms of how a particular piece of music might express emotion or evoke particular emotions in listeners is not yet completely understood, although a coherent theoretical framework for experienced emotion has been proposed by *Patrik Juslin* and *Daniel Västfjäll* [29.74]. The individual mechanisms in this framework (dubbed as BRECVEMA according to eight mechanisms in a recent update, [29.75]) will, in this synthesis, be divided into four fundamental processes (physiology, embodied, memory, and appraisal), which are assigned to the different levels of affects described earlier. The purpose is to better acknowledge the interlaced nature of sensed, perceived and experienced emotions, and the thus emotion mechanisms follow the logic of the previous chapter and the diagram outlined in Fig. 29.1.

This reorganization also implies that the boundary between perceived and experienced emotions with respect to emotion mechanisms are redefined, since the BRECVEMA framework only applies to experienced emotions. However, several of the low-level mechanisms (e.g., contagion, entrainment) are appropriate also for emotion perception and by bringing these two closely related processes to the same scheme, the mechanisms previously applied only to induction may help to understand emotion perception principles. This de-

cision is also motivated by the current lack of suitable mechanisms to explain the way emotions are mapped across domains (e.g., visual, auditory). In such processes, the concept of contagion is actually dependent on recognition of the actions giving rise to emotion experiences.

Again, we start with the low-level mechanisms for emotion perception before moving on to higher-level mechanisms of emotional experiences (the right side of Fig. 29.1). The issues that are known to modify the emotion processes in music are also incorporated into this summary of mechanisms.

29.3.1 Mapping Mechanisms of Emotions

In this synthesis of the mechanisms, the BRECVEMA framework can be interpreted to have two low-level *mapping mechanisms*. These are mapping mechanisms between sound and emotion, as direct transfer – mapping – between sound and affect seems to take place. Two categories of mapping mechanisms can be distinguished. The first of these, labeled as *orientation mechanisms*, consist of the brain stem reflex, expectancy and entrainment. The brain stem reflex is hardwired attention response that is activated by any exceptional – loud, sudden, sharp, accelerating – sound, which not only directly influences core affects (e.g., increase in physiological arousal) but is also recognized as surprise and may be experienced as unpleasant or exciting [29.76]. Another orientation mechanism is expectancy, where any violation (pitch, tonal, rhythmic) of the expected musical structure creates an orientation response at the core affect level, which may be interpreted in different ways; violation of the expectations has been observed to lead to anxiety [29.76], surprise [29.77], or even to thrills [29.78]. Entrainment refers to adjustment – though not necessarily synchronization – of the internal oscillators to rhythmic periods in the music. Entrainment itself may be conceived as a multifaceted phenomenon since it can refer to perceptual, motor, physiological, or even social entrainment [29.79], which may each have different implications for emotions. For instance, motor entrainment has been shown to modulate pleasantness [29.80], but social entrainment has been judged to induce feelings of being connected [29.81]. The entrainment mechanism also facilitates orienting and attending to information in music. For this reason, entrainment acts as an attentional focus that may be capable of enhancing the delivery of other mechanisms and emotions in music. In summary, these three orientation mechanisms (brain stem reflex, expectancy, and entrainment) act as strong guides for perceptual processes and attention, are considered to be low-level processes, and often lead to changes in core affects (arousal in particular).

A second set of relatively low-level mechanisms, including contagion and visual imagery from the BRECVEMA framework, are in the synthesis presented here (Fig. 29.1) defined collectively as *embodied* mechanisms. These are embodied in the sense that they refer to reactivation of past sensory and motor mechanisms [29.82–84]). This perspective considers that the body plays a major role in all interactions with environment, and that any response to stimuli will be based on simulations, or reenactments, of others' nonverbal expressions and affective states [29.85]. Such mappings arise within expressive channels, such as matching another person's facial expressions or vocal intonations [29.86], but also across channels (between audio and visual [29.87]).

What makes contagion as a mechanism particularly relevant for embodied explanations is that it is assumed to consist of a process in which the *listener perceives the emotional expression of the music, and then* mimics *this expression internally* [29.74, p. 565]. Taken further, this must mean that most basic emotions are linked to specific bodily reactions and motor patterns that have been influenced by the production or causal results of experiencing that emotion. Because of built-in codes based on states reflecting the emotional experiences or its output (movements, expression, sounds, gestures of such states), an internal mapping between an external stimulus and our representation of the possible cue combinations is possible in contagion. If we take the example of fear, the peripheral nervous system in fear is associated with increased levels of glucocorticoids and norepinephrine [29.88] that lead to states of alarm that quickly raise autonomic arousal, which in turn leads to louder vocal output, higher pitch, brighter timbre and faster movements than in neutral state. These sets of cues, which are very similar in music and speech [29.89], are grounded in physical changes caused by underlying emotion states, in this example by fear and an ensuing state of stress, which affects said individuals' vocal expression, posture, facial expression as well as movement. Knowledge of this multimodal code allows us to decipher the intended emotional expressions, and this knowledge is assumed to be implicit and accessible by embodied simulation (*what would I feel like if I sounded like that* ...). Even when merely activating some of the motor programs used in simulations, we may catch the same emotions, or at least, our emotional reactions may be amplified [29.82].

Another mapping mechanism, visual imagery, may also be associated with embodied processes, since cross-modal correspondences are thought to occur in music [29.90] and in other domains [29.91, 92]. Vi-

sual imagery is believed to generate emotions through such correspondences. Whilst the exact mapping in this mechanism has not yet been pinned down, it is assumed to be related to metaphoric representations [29.93] that are tightly coupled with physical dimensions [29.94].

29.3.2 Evaluative Mechanisms of Emotions

In addition to the low-level automatic mechanisms of emotions, a set of high-level, largely conscious and intentional mechanisms are included in the BRECVEMA framework. In the synthesis, these are cataloged into *memory* and *appraisal* mechanisms. The BRECVEMA framework includes two mechanisms related to memory, *episodic memory* and *evaluative conditioning*. In the latter, a repeated pairing of a particular music or sound example with either a positive or negative stimuli or outcome leads to a conditioned emotional response. This can be either automatic and unconscious or conscious. In music, leitmotifs associated with particular heroes and villains may be the best example of this mechanism, and empirical work on conditioned responses to sounds have demonstrated the malleability of such conditioned responses [29.95]. *Episodic memory* as a mechanism for emotions refers to a recollection of a specific event prompted by the music. This might lead to an emotional experience that could be strikingly different from the emotion expressed by the music. For instance, in recent study, a sad musical expression was turned into happy emotional experience by inserting a short quotation from the Star Wars theme into the sad solo cello piece [29.76]. This mechanism, known also as the *Darling, they are playing our tune* phenomenon [29.96], is a potent mechanism for emotions, as autobiographical memories have been identified as the most frequently listed motive for listening to music [29.73], and, of course, the most essential mechanism for nostalgia [29.97].

Another set of mechanisms, here referred together as *appraisal mechanisms*, consists of three distinct categories: cognitive appraisal, identity confirmation, and aesthetic judgment. Cognitive appraisal is a process in which emotion is the cause of an event regarded to have significant implications for the goals of the individual [29.98]. Though this mechanism is not explicitly included in the BRECVEMA framework, it does function as a bridge between the appraisal theories of emotion [29.99, 100] and their successful use in explaining responses to arts [29.101]. Moreover, the appraisal processes, such as novelty check, goal relevance, goal congruence, and coping potential, can be considered powerful moderators of emotions [29.102]. Mechanisms labeled as *identity confirmation*, although not put forward as a category in the BRECVEMA framework, refer to the potential powerful social effects of music as carrier of self-identity [29.103, pp. 73–74] (see also [29.104]). *Juslin* et al. [29.73, p. 190] found this mechanism to be the second most important explanation of emotional episodes associated with music, yet it remains currently unexplored. The third appraisal mechanism is *aesthetic judgment*, which is simply an evaluation of the aesthetic value of music. Such evaluation may comprise initially of an aesthetic attitude [29.105], but may also comprise multiple criteria, including beauty, skill, novelty, and artistic intention [29.75]. This recent extension to the BRECVEMA framework is considerably broader than the other mechanisms.

To recapitulate the sections on affect models and mechanisms, let us go back to the theoretical synthesis (Fig. 29.1) once more. Here I will take one of the most ubiquitous hits of music and emotion studies as an example, the *Adagio in G minor for Strings and Orchestra* by Tomaso Albinoni, which has been used at least in 16 published studies, despite the probable and dubious associations such frequently heard musical piece might carry. In terms of core affects, it is fairly clear that the slow tempo, soft timbres, low dynamics, legato articulation, and gradually descending melodies lines mimic low arousal state, and perhaps even negative valence due to multiple cues consistent with sadness [29.106]. In this case simple mapping mechanisms such as contagion, entrainment, and expectations support interpretation of the music as calm and low-arousing, due to musical qualities such as predictability and ease of synchronization. Core affects could be measured through psychophysiology, self-reports consisting of dimensions, or by asking the listeners to rate what the music expresses in terms of basic emotion categories. This would most likely lead to the assessment that the piece expresses *sadness* [29.107], and perhaps tenderness. No surprises here, although such ratings of perceived emotions would be subject to minor modifications based on current mood, situation, personality traits and music preferences.

However, it is informative to consider what happens when attention is shifted to the emotions experienced when listening to the Adagio. In this case, appraisal and memory mechanisms kick into play, which may lead experienced emotions to differ markedly from perceived emotions. For example, somebody hearing the Adagio for the first time in a particularly receptive situation (say, in a cathedral when attending a memorial service) might experience the emotion as being *moved*, whereas another listener highly familiar with the piece might experience *nostalgia* due to fond recollections of past performances of the piece. For others, the ap-

praisal mechanism related to aesthetic judgment might steer the experience towards feelings of *wonder* due to the sheer beauty of the piece and aesthetic appreciation of the classical music canon. These trajectories of how the emotional experiences and their labels may systematically – but not arbitrarily – change across the levels of affects are illustrated in the schematic outline (Fig. 29.1). It is worth pointing out that many of the shifts between perceived and experienced emotions caused by the memory and appraisal mechanisms lead to more positive experiences. This *positive pull* is likely to be related to the fact that music does not have an obvious material effect on the listener's well-being; it is voluntary activity that offers a medium in which listeners can safely project their emotions. Taken further, it also implies that the experience of intense emotion of any kind may be inherently pleasurable, as long as the emotions do not come with any real-life consequences attached. The best example of the positive pull is sad music, easily recognized as such by the listeners [29.108], but usually experienced as positive, perhaps involving feelings of peace, tenderness, nostalgia, or being moved [29.53, 72].

The purpose of the concurrent presentation (Fig. 29.1) of affect levels and mechanisms in the present synthesis is to emphasize that the mechanisms together with the modifiers may lead to entirely different experiences, and that core affects, perception of basic emotions, and experienced emotions represent different yet connected explanatory levels of these processes. While the mechanisms themselves undoubtedly have a major influence on shaping the emotions, there are other factors that are known to influence emotions in particular ways, here termed as *contextual modifiers*, which will be described in the next section.

29.3.3 Contextual Modifiers of Emotions

Emotions that are experienced are more likely to be influenced by matters relating to context, whether it is the music, listener, or situation. This idea has been formally expressed by *Scherer* and *Zentner* [29.109] as a multiplicative function between structure of music, performance, listener, and context. The multiplicative nature of this scheme has not been directly tested, which suggests that the entire topic of modifiers has been considered of secondary importance in music and emotion research. Nevertheless, some contextual modifiers have been explored.

Starting with the music itself, music and emotion studies have mainly been conducted in Western art music contexts, utilizing highly educated Western listeners in particularly restricted situations (mainly laboratory settings, see [29.53]). The context created by music itself, its genre, lyrics, and cultural connotations, is perhaps the most obvious modifier of emotions. Musical devices expressing emotions vary across musical genres, periods and cultures. Despite the apparent differences in musical materials, the role played by culture in musical emotion has been considered only a modifier on emotions since it has been shown repeatedly that the basic emotions in music can be recognized across cultures [29.108, 110]. However, complex emotions and experiences appear to be more dependent on cultural knowledge since they rely on aesthetic judgments, memories, and identify formation, which all require learning and exposure to the music. This raises a more fundamental issue, which is that different genres of music have widely different functional uses. In the present synthesis, an attempt has been made to capture this issue, through the notion that situations may alter the perception and experience of emotions, and that this may be more fundamental than a mere contextual modifier. Not all emotion concepts are relevant for all types of music, and for this reason even some of the core affect dimensions (e.g., valence) have been shown to be problematic when applied across different music genres [29.111]. Moreover, when large stimulus sets provided by social tagging of music are harnessed for computational analysis of emotional expression, it has been found that contextual information such as genre brings significant improvements to prediction of emotions [29.112]. Contextual information is also known to affect perception and experience of emotions. For instance, extra-musical information intensifies [29.113], lyrics both enhance and subdue [29.114], and images amplify [29.60] the emotional experiences aroused by music.

Individual traits, moods, expertise and preferences of the listeners also modify emotions. At the broadest level, culture does contribute to the emotions since traditions, customs, and musical genres are byproducts of subcultures. As most music and emotion research has been conducted on Western listeners, there is a need to expand the sphere of investigation outside this realm (e.g., [29.115]). Of the other modifiers to emotions, musical expertise has been suggested to either amplify [29.116], or have little effect on both perceived [29.117] and experienced emotions [29.70]. Personality traits have been shown to have a small but consistent moderator role for both experience of emotions [29.118] and perception of emotions [29.119]. Music preferences and motives for listening to music also have a small but systematic impact on emotions [29.73, 116].

A number of situational factors seem to affect perception of emotions in music but many of these are particularly important for experiencing emotions. Everyday music listening studies [29.120, 121] have shown

that variations across listening contexts – whether at home, at a laboratory, on public transport, or with friends – have an effect on which emotions are likely to be experienced. These differences may not relate only to the situations themselves, but also to differences in emotion regulation goals afforded by the situation. For instance, a gym affords increase in arousal whereas a church has an entirely different set of functions, not only because these places differ acoustically. Finally, the most complex situational modifiers arise from the social dynamics of the situation. For instance, whether a listener is alone or in a group [29.122], and the opinions held by others in the group, both of which are known to affect emotions aroused by music [29.123].

29.4 Measures and Musical Materials

29.4.1 Self-Report Measures of Emotions

Most of the research on music and emotions relies on self-report measures such as Likert ratings and forced-choice designs meant to capture the emotions representing the affect models (dimensions, basic, or complex). There are numerous standardized instruments for collecting mood and emotion evaluations (such as Positive and Negative Affect Scale – PANAS, profile of mood states – POMS, and self-assessment manikin – SAM, or differential emotions scale – DES, etc.), which have been reviewed in detail elsewhere [29.124]. All self-reports measures are subject to caveats and limitations. These often relate to the confusion between perceived and experienced emotions, which can be difficult for participants to dissociate, or more generally to the demand characteristics of such methods, in which participants are biased towards complying with the inferred outcome. Despite these caveats, the standardized measures are efficient and convenient, but may either assume too much about the underlying experience or rely on specific semantic labels. Therefore it is also useful to probe experiences with open responses [29.125], interviews [29.103], and nonverbal measures such as similarity ratings [29.117]. Another way to qualify the emotions experienced is to collect peripheral and indirect measures of emotions.

29.4.2 Peripheral and Indirect Measures of Emotions

Peripheral measures of emotions, such as skin conductance response (SCR), heart rate variability (HRV), facial electromyography (EMG), respiration, and temperature have become increasingly common in studies involving experienced emotions [29.76, 126, 127]. These indicators are well established in terms of the underlying physiology and emotional correlates (reviewed by [29.128]). In research focusing on strong responses to music, it has become customary to record chill reactions [29.129, 130]. Such physiological measures are not always sensitive enough for the emotional experiences examined [29.59, 131], although they do track arousal adequately [29.127]. In addition, there are other behavioral ways of discovering whether the emotional experience is actually taking place, including indirect measures and reaction times, both of which rely on the fact that emotional experience biases cognitive judgment in a systematic fashion. For instance, a sad individual interprets ambiguous faces to be more negative and processes incongruent (e.g., happy) information more slowly than a nonsad individual. Therefore, indirect measures can be used to assess whether listeners are experiencing a negative or positive emotion [29.113], despite the considerable limitations such sensitive measures have. The three levels of affect outlined earlier (Fig. 29.1) are not all accessible with peripheral measures of emotions. Core affects may be measured with physiology, and with some caveats, the basic emotions might serve as good targets of physiological measures if the question is really about emotional experiences rather than emotion perception. Measuring complex emotions with physiology may not be feasible unless one is interested in a very specific type of emotional reaction such as chills or extremely strong, aesthetic reactions.

29.4.3 Neural and Endocrine Measures of Emotions

Measuring brain activation has become increasingly common in emotion studies, although the purpose is often not to verify the emotions experienced or perceived, but rather to pinpoint when and where processing take place. In this, electroencephalography (EEG), magnetoencephalography (MEG), and functional magnetic resonance imaging (fMRI) are the key techniques, with EEG and MEG indicating electric activity, and fMRI indicating blood flow and blood-oxygen level, both of which index the underlying brain activity. fMRI is accurate in terms of the location of the neural activity but imprecise in timing. Therefore, fMRI studies have revealed which areas are involved in emotion experiences [29.130, 132] and perceived emotions, in terms

of stimulus valence [29.133–135]. However, discovering where emotional information is processed is not the main aim of such studies as the real insights come from functional explanations of what each area of the brain area is responsible for. Such explanations allow us to associate the areas in question to plausible mechanisms. For example, a part of the limbic system called the amygdala, which itself has further specialized areas, is known to be relevant in detecting dangers and fast processing of visual information, as well as being able to code the core affect of arousal [29.136]. In neuroscience studies of music, the amygdala has been shown to be heavily involved in processing of strong emotions, such as chills, but also in strong negative stimuli, demonstrating that it captures the intensity of the stimuli [29.137]. In addition, the anterior cingulate cortex (ACC), an area involved in autonomic activity, is often implicated in music and neuroscience research. This could be taken as another index of core affects, but such activations may take place due the involvement of this area with movement and motivation.

Other brain areas such as ventral striatum and anterior insula have been associated with pleasure induced by music [29.138], which in more general terms has been connected to reward-related brain areas that regulate the release of dopamine in the brain (demonstrated by [29.139]). In research connecting brain areas to emotion mechanisms, the amygdala and ventral striatum have been associated with low-level mapping mechanisms as well as the outcomes of the emotions via higher-level mechanisms. In addition, there are more specific areas related to memory mechanisms such as the hippocampus. The anterior hippocampal formation has been observed to display activity changes in a numerous music studies [29.132, 140], and this area has a crucial role in learning and memory, as well as in expectation and the detection of novelty [29.141]. However, this is not to say that only musically induced emotions generated via memory mechanisms (episodic memory, evaluative conditioning) necessarily show hippocampus activation, as the hippocampus also serves as a switchboard between cortical and subcortical areas, and is thought to be involved in positive social emotions, and even in the experience of being moved [29.142].

In addition, studies of brain-damaged patients that demonstrate selective emotional impairments can reveal which functional brain areas are necessary for emotion perception and experience [29.143]. Electrophysiological measures (mostly EEG) have typically investigated valence as a core affect in association with music [29.133, 144, 145]. The results have shown that positive and negative emotions produce different hemispheric lateralization of neural activity; the right hemisphere is involved more during negative and the left during positive emotions. However, the results from these studies fail to paint a consistent picture, since the results tend to be dependent on the type of techniques employed.

29.4.4 Musical Materials

The availability of good-quality audio is nowadays virtually limitless, but the number of excerpts used in studies is usually limited by the length and feasibility of the experimental task. In behavioral studies that focus on perceived emotions, a large number of stimuli could potentially be used. Despite this, there are only a few exceptional cases where the number of stimuli utilized in a music and emotion study is truly large: for example the several thousands of music examples used by *Schuller* et al. [29.146]. Online annotation schemes based on self-reports have typically produced larger datasets than laboratory experiments, with up to 500 music excerpts in some cases [29.147]), but this may be at the expense of audio quality, listener attention, and overall control.

Crowdsourcing is one way of harnessing the power of the masses, often through implementation of online annotation games [29.148, 149]. It has been suggested that crowdsourcing seems to yield data of similar accuracy to that generated by expert annotations [29.150], since the sheer volume of data compensates for variation between participants and evaluation settings. Another way of obtaining large amounts of ecologically valid information pertinent to perceived emotions is to tap into social tagging services such as *Last.fm* or curated databases such as *I Like Music*. The data in these services contain, among other things, user-defined *tags* for each music track. The tags may represent genres, preferences, and situations, but a significant proportion of them relate to emotions. For this reason, the collections of tags in these services have been used to build emotion models [29.112] and to connect the emotions to musical and acoustic properties of the music [29.151]. For music and emotion research, these massive datasets have tremendous potential for formulating predictive models, and also for connecting the everyday uses of music to situations, emotions, and individual profiles of listeners (such as age, gender, music preferences, musical expertise and personality). The validity of emotion structures inferred through analysis of such noisy folksonomies remains, however, to be determined in a rigorous comparison of these and carefully constructed psychometric laboratory studies.

29.5 Current Challenges

The topic of music and emotions in systematic musicology faces several challenges, many of which have been acknowledged for some time [29.53, 75, 152]. Perhaps the most important challenges to meet concern the widening of research contexts, both in terms of culture and music listening situation, as this could lead to significant revisions of the main concepts in music and emotion research in the near future.

29.5.1 Widening the Research Context

The majority of music and emotion studies have been carried out in Western countries and amongst particularly elite groups of people (affluent, young college students) using music from a narrow repertoire (classical, jazz, and film soundtracks). It is clear that the range of people under investigation should be broader. This could mean cross-cultural studies; however, such research is increasingly difficult due to the effects of continuing globalization. Nevertheless, useful tests of generalizability of findings can be performed within different cultural practices, regions, and subcultures found in the West, and by turning attention to representative rather than convenience samples.

Another challenge for the field is the applicability of the results to situations in which music is listened to. Most studies in this field have been conducted in laboratory settings [29.53], which is unlikely to be an environment conducive to complex emotional experiences. This is particularly the case if memory and appraisal mechanisms are tightly controlled, such as when listeners are exposed to unfamiliar music that does not fit their musical identities. Although such an artificial setting is not entirely problematic for studying perceived emotions, it is likely to suppress experienced emotions to some extent. One solution is to tap into the listening activities and emotion of listeners with the experience sampling method (ESM, [29.120]), which can now be directly incorporated into smartphones [29.121]. This method, coupled with relevant information about music, environment, individual features and current activities, can lead to significantly more realistic research settings.

29.5.2 Narrowing Down the Causal Influences

Only a small minority of music and emotion studies have attempted to establish causal links between musical features and emotions perceived or mechanisms and the ensuing emotions. In order to understand how low-level mapping mechanisms contribute to contagion, entrainment, or expectancy mechanisms, factorial manipulations of key musical features are indispensable. This line of research was started in the 1970s [29.30] but has not taken hold in the field despite promising studies on production [29.56], analysis-by-synthesis [29.153], and synthetic stimulus creation [29.154]. With the formulation of emotion induction mechanisms [29.74], cause and effect has been established between some emotions and mechanisms [29.76, 155], but a great deal more has to be done in order to explain how music is able to generate emotions through separate mechanisms at different levels of explanation. The synthesis provided in this chapter attempted to bridge the gap between the existing mechanisms of music-induced emotions, emotion perception, and core affects. The low-level processes involving recognition, physical characteristics and embodiment of emotions are necessary building blocks for high-level emotional experiences.

Most of the research in the field relies on retrospective ratings of emotions. Since both music and emotional experiences unfold in time, moment-by-moment fluctuations should be better incorporated into experiment design and theories. Although continuous self-reports have been used numerous times [29.129, 156–158], the approach has not yet provided full insights into the processes, perhaps because the modeling techniques are still unresolved. The most promising way to solve this problem seems to come from modeling the neural responses to music [29.159].

The topic of music and emotion has been undergoing a profitable expansion of themes, approaches and definitions during the last decade. As a topic for systematic musicology, it offers a rich *interdisciplinary* object of study that will benefit from close cultural and historical readings of the emotions in different eras, places, and subcultures, using methods ranging from strict laboratory experiments designed to tease apart the theoretical constructs, to experiments involving biological markers of emotions. Moreover, the technological advances that have altered the way music is being consumed and analyzed (such as music information retrieval), offer yet additional motivation and research tools to make progress on this topic in a transparent, empirical, and systematic fashion.

References

29.1 A.J. Lonsdale, A.C. North: Why do we listen to music? A uses and gratifications analysis, Br. J. Psychol. **102**(1), 108–134 (2011)

29.2 E. Dissanayake: Root, leaf, blossom, or bole: Concerning the origin and adaptive function of music. In: *Communicative Musicality: Exploring the Basis of Human Companionship*, ed. by S. Malloch, C. Trevarthen (Oxford Univ. Press, Oxford 2009) pp. 17–30

29.3 R. MacDonald, G. Kreutz, L. Mitchell: What is music, health, and wellbeing and why is it important. In: *Music, Health and Wellbeing*, ed. by R. MacDonald, G. Kreutz, L. Mitchell (Oxford Univ. Press, Oxford 2012) pp. 3–11

29.4 P. Juslin, J.A. Sloboda (Eds.): *Handbook of Music and Emotion: Theory, Research, Applications* (Oxford Univ. Press, Oxford 2010)

29.5 A. Gabrielsson: *Strong Experiences with Music: Music is Much More Than Just Music* (Oxford Univ. Press, Oxford 2011)

29.6 P.N. Juslin: From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions, Phys. Life Rev. **10**(3), 235–266 (2013)

29.7 S. Koelsch: *Brain and Music* (Wiley, Oxford 2012)

29.8 J.A. Russell: A circumplex model of affect, J. Personal. Soc. Psychol. **39**(6), 1161–1178 (1980)

29.9 L.B. Meyer: *Emotion and Meaning in Music* (Univ. of Chicago Press, Chicago 1956)

29.10 P.R. Farnsworth: *The Social Psychology of Music*, 2nd edn. (Iowa State Univ. Press, Ames 1969)

29.11 D.E. Berlyne: *Aesthetics and Psychobiology* (Appleton-Century-Crofts, East Norwalk 1971)

29.12 M.G. Rigg: The mood effects of music: A comparison of data from four investigators, J. Psychol. **58**(2), 427–438 (1964)

29.13 L. Wedin: A multidimensional study of perceptual-emotional qualities in music, Scand. J. Psychol. **13**(4), 241–257 (1972)

29.14 A. Gabrielsson: Adjective ratings and dimension analyses of auditory rhythm patterns, Scandinavian J. Psychol. **14**(1), 244–260 (1973)

29.15 F.V. Nielsen: *Oplevelse af musikalsk spænding [The experience of Musical Tension]* (Akademisk Forlag, Copenhagen 1983)

29.16 J. Panksepp: The emotional sources of chills induced by music, Music Percept. **13**(2), 171–207 (1995)

29.17 A. Gabrielsson, S. Lindström Wik: Strong experiences related to music: A descriptive system, Musicae Scientiae **7**, 157–217 (2003)

29.18 A. Damasio: *Descartes' Error: Emotion, Reason and the Human Brain* (Random House, New York 1994)

29.19 J.E. LeDoux: *The Emotional Brain: The Mysterious Underpinnings of Emotional Life* (Simon Schuster, New York 1996)

29.20 J. Panksepp: *Affective Neuroscience: The Foundations of Human and Animal Emotions* (Oxford Univ. Press, Oxford 1998)

29.21 A.J. Blood, R.J. Zatorre, P. Bermudez, A.C. Evans: Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions, Nature Neurosci. **2**(4), 382–287 (1999)

29.22 P.N. Juslin, J.A. Sloboda (Eds.): *Music and emotion. Theory and Research* (Oxford Univ. Press, Oxford 2001) pp. 583–646

29.23 C. Beedie, P. Terry, A. Lane: Distinctions between emotion and mood, Cogn. Emot. **19**(6), 847–878 (2005)

29.24 M. Gendron, L.F. Barrett: Reconstructing the past: A century of ideas about emotion in psychology, Emot. Rev. **1**(4), 316–339 (2009)

29.25 C.E. Izard: Emotion theory and research: Highlights, unanswered questions, and emerging issues, Annual Rev. Psychol. **60**, 1–25 (2009)

29.26 P.M. Niedenthal, S. Krauth-Gruber, F. Ric: *Psychology of Emotion: Interpersonal, Experiential, and Cognitive Approaches* (Psychology, New York 2006)

29.27 D. Sander: The cambridge handbook of human affective neuroscience. In: *The Cambridge Handbook of Human Affective Neuroscience*, ed. by J.L. Armony, P. Vuilleumier (Cambridge Univ. Press, Cambridge 2013) pp. 5–53

29.28 N.H. Frijda, K.R. Scherer: Emotion definition (psychological perspectives). In: *Oxford Companion to Emotion and the Affective Sciences*, ed. by D. Sander, K.R. Scherer (Oxford Univ. Press, Oxford 2009) pp. 142–143

29.29 K. Hevner: Expression in music: A discussion of experimental studies and theories, Psychol. Rev. **42**(2), 186–204 (1935)

29.30 K.R. Scherer, J.S. Oshinsky: Cue utilization in emotion attribution from auditory stimuli, Motiv. Emot. **1**(4), 331–346 (1977)

29.31 P.G. Hunter, E.G. Schellenberg, U. Schimmack: Mixed affective responses to music with conflicting cues, Cogn. Emot. **22**(2), 327–352 (2008)

29.32 P. Evans, E. Schubert: Relationships between expressed and felt emotions in music, Musicae Scientiae **12**, 75–99 (2008)

29.33 L. Barrett: Emotions as natural kinds, Perspect. Psychol. Sci. **1**(1), 28–58 (2006)

29.34 J.A. Russell, L.B. Feldman: The circumplex model of affect. In: *The Oxford Companion to Emotion and the Affective Sciences*, ed. by D. Sander, K.R. Scherer (Oxford Univ. Press, New York 2009) p. 104

29.35 R. Reisenzein, S. Doring: Ten perspectives on emotional experience: Introduction to the special issue, Emot. Rev. **1**(3), 195–205 (2009)

29.36 J.A. Russell, L. Feldman Barrett: Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant, J. Personal. Soc. Psychol. **76**(5), 805–819 (1999)

29.37 J.A. Russell: Pancultural aspects of the human conceptual organization of emotions, J. Personal. Soc. Psychol. **45**(6), 1281–1288 (1983)

29.38 J. Posner, J.A. Russell, B.S. Peterson: The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology, Dev. Psychopathol. **17**(03), 715–734 (2005)

29.39 M. Lewis: The emergence of human emotions, Handb. Emotions **2**, 265–280 (2000)

29.40 J.A. Russell: Culture and the categorization of emotions, Psychol. Bull. **110**(3), 426–450 (1991)

29.41 B. Mesquita: Emotions as dynamic cultural phenomena. In: *Handbook of Affective Sciences*, Series in Affective Science, ed. by R.J. Davidson, K.R. Scherer, H.H. Goldsmith (Oxford Univ. Press, Oxford 2003) pp. 871–890

29.42 R.J. Davidson, W. Irwin: The functional neuroanatomy of emotion and affective style, Trends Cogn. Sci. **3**(1), 11–21 (1999)

29.43 D. Watson, L.A. Clark, A. Tellegen: Development and validation of brief measures of positive and negative affect: The PANAS scales, J. Personal. Soc. Psychol. **54**(6), 1063–1070 (1988)

29.44 R.E. Thayer: *The Biopsychology of Mood and Arousal* (Oxford Univ. Press, New York 1989)

29.45 G. Ilie, W. Thompson: A comparison of acoustic cues in music and speech for three dimensions of affect, Music Percept. **23**(4), 319–329 (2006)

29.46 T. Eerola, R. Ferrer, V. Alluri: Timbre and affect dimensions: Evidence from affect and similarity ratings and acoustic correlates of isolated instrument sounds, Music Percept. **30**(1), 49–70 (2012)

29.47 J.T. Cacioppo, G. Berntson: Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates, Psychol. Bull. **115**, 401–423 (1994)

29.48 U. Schimmack, S. Colcombe: Eliciting mixed feelings with the paired-picture paradigm: A tribute to kellogg (1915), Cogn. Emot. **21**(7), 1546–1553 (2007)

29.49 J.T. Larsen, A.P. McGraw, J.T. Cacioppo: Can people feel happy and sad at the same time?, J. Personal. Soc. Psychol. **81**(4), 684–696 (2001)

29.50 J.T. Larsen, A.P. McGraw: Further evidence for mixed emotions, J. Personal. Soc. Psychol. **100**(6), 1095–1110 (2011)

29.51 D. Matsumoto, P. Ekman: Basic emotions. In: *Oxford Companion to Affective Sciences*, ed. by D. Sander, K.R. Scherer (Oxford Univ. Press, Oxford 2009) pp. 69–72

29.52 K. Vytal, S. Hamann: Neuroimaging support for discrete neural correlates of basic emotions: A voxel-based meta-analysis, J. Cogn. Neurosci. **22**(12), 2864–2885 (2010)

29.53 T. Eerola, J.K. Vuoskoski: A review of music and emotion studies: Approaches, emotion models and stimuli, Music Percept. **30**(3), 307–340 (2012)

29.54 S.B. Dalla, I. Peretz, L. Rousseau: N. Gosselin A developmental study of the affective value of tempo and mode in music, Cognition **80**(3), B1–10 (2001)

29.55 E. Nawrot: The perception of emotional expression in music: Evidence from infants, children and adults, Psychol. Music **31**(1), 75–92 (2003)

29.56 P.N. Juslin: Emotional communication in music performance: A functionalist perspective and some data, Music Percept. **14**(4), 383–418 (1997)

29.57 P. Laukka, A. Gabrielsson: Emotional expression in drumming performance, Psychol. Music **28**(2), 181–189 (2000)

29.58 P. Gomez, B. Danuser: Affective and physiological responses to environmental noises and music, Int. J. Psychophysiol. **53**(2), 91–103 (2004)

29.59 J.A. Etzel, E.L. Johnsen, J. Dickerson, D. Tranel, R. Adolphs: Cardiovascular and respiratory responses during musical mood induction, Int. J. Psychophysiol. **61**(1), 57–69 (2006)

29.60 T. Baumgartner, M. Esslen, L. Jancke: From emotion perception to emotion experience: Emotions evoked by pictures and classical music, Int. J. Psychophysiol. **60**(1), 34–43 (2006)

29.61 I. Peretz, L. Gagnon, B. Bouchard: Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage, Cognition **68**(2), 111–141 (1998)

29.62 N. Gosselin, I. Peretz, M. Noulhiane, D. Hasboun, C. Beckett, M. Baulac, S. Samson: Impaired recognition of scary music following unilateral temporal lobe excision, Brain **128**(3), 628–640 (2005)

29.63 T. Eerola, J.K. Vuoskoski: A comparison of the discrete and dimensional models of emotion in music, Psychol. Music **39**(1), 18–49 (2011)

29.64 J. Robinson: Aesthetic emotions (philosophical perspectives). In: *The Oxford Companion to Emotion and the Affective Sciences*, ed. by D. Sander, K.R. Scherer (Oxford Univ. Press, New York 2009) pp. 6–9

29.65 J.L. Tsai, B. Knutson, H.H. Fung: Cultural variation in affect valuation, J. Personal. Soc. Psychol. **90**, 288–307 (2006)

29.66 V. Bennett: Music and emotion, Music. Q. **XXVIII**(4), 406–414 (1942)

29.67 D. Cooke: *The Language of Music* (Oxford Univ. Press, Oxford 1959)

29.68 K. Nordenstreng: A comparison between the semantic differential and similarity analysis in the measurement of musical experience, Scand. J. Psychol. **9**(1), 89–96 (1968)

29.69 E.P. Asmus: The development of a multidimensional instrument for the measurement of affective responses to music, Psychol. Music **13**(1), 19–30 (1985)

29.70 M. Zentner, D. Grandjean, K. Scherer: Emotions evoked by the sound of music: Characterization, classification, and measurement, Emotion **8**(4), 494–521 (2008)

29.71 W. Trost, T. Ethofer, M. Zentner, P. Vuilleumier: Mapping aesthetic musical emotions in the brain, Cerebral Cortex **22**(12), 2769–2783 (2012)

29.72 L. Taruffi, S. Koelsch: The paradox of music-evoked sadness: An online survey, PLoS ONE **9**(10), e110490 (2014)

29.73 P.N. Juslin, S. Liljeström, P. Laukka, D. Västfjäll, L.-O. Lundqvist: Emotional reactions to music in a nationally representative sample of swedish adults prevalence and causal influences, Musicae

Scientiae **15**(2), 174–207 (2011)
29.74 P. Juslin, D. Västfjäll: Emotional responses to music: The need to consider underlying mechanisms, Behav. Brain Sci. **31**(05), 559–575 (2008)
29.75 P.N. Juslin, J.A. Sloboda: Music and emotion. In: *Psychology of Music*, 3rd edn., ed. by D. Deutsch (Academic Press, New York 2013)
29.76 P.N. Juslin, L. Harmat, T. Eerola: What makes music emotionally significant? Exploring the underlying mechanisms, Psychol. Music **42**(4), 599–623 (2013)
29.77 D. Huron: *Sweet Anticipation: Music and the Psychology of Expectation* (MIT Press, Cambridge 2006)
29.78 J.A. Sloboda: Music structure and emotional response: Some empirical findings, Psychol. Music **19**(2), 110–120 (1991)
29.79 W. Trost, P. Vuilleumier: Rhythmic entrainment as a mechanism for emotion induction by music: A neurophysiological perspective. In: *The Emotional Power of Music: Multidisciplinary Perspectives on Musical Arousal, Expression, and Social Control*, ed. by T. Cochrane, B. Fantini, K.R. Scherer (Oxford Univ. Press, Oxford 2013) pp. 213–225
29.80 W. Trost, S. Frühholz, D. Schön, C. Labbé, S. Pichon, D. Grandjean, P. Vuilleumier: Getting the beat: Entrainment of brain activity by musical rhythm and pleasantness, NeuroImage **103**, 55–64 (2014)
29.81 A.P. Demos, R. Chaffin, K.T. Begosh, J.R. Daniels, K.L. Marsh: Rocking to the beat: Effects of music and partner's movements on spontaneous interpersonal coordination, J. Exp. Psychol. Gen. **141**(1), 49–53 (2012)
29.82 P.M. Niedenthal, P. Winkielman, L. Mondillon, N. Vermeulen: Embodiment of emotion concepts, J. Personal. Soc. Psychol. **96**(6), 1120–1136 (2009)
29.83 L.W. Barsalou: Simulation, situated conceptualization, and prediction, Philos. Trans. R. Soc. B **364**(1521), 1281–1289 (2009)
29.84 P.M. Niedenthal: Embodying emotion, Science **316**(5827), 1002–1005 (2007)
29.85 L.W. Barsalou, P.M. Niedenthal, A.K. Barbey, J.A. Ruppert: Social embodiment, Psychol. Learn. Motiv. **43**, 43–92 (2003)
29.86 U. Dimberg, M. Thunberg, K. Elmehed: Unconscious facial reactions to emotional facial expressions, Psychol. Sci. **11**(1), 86–89 (2000)
29.87 S.R. Arnott, A. Singhal, M.A. Goodale: An investigation of auditory contagious yawning, Cogn. Affect. Behav. Neurosci. **9**(3), 335–342 (2009)
29.88 S.M. Rodrigues, J.E. LeDoux, R.M. Sapolsky: The influence of stress hormones on fear circuitry, Annu. Rev. Neurosci. **32**, 289–313 (2009)
29.89 P.N. Juslin, P. Laukka: Communication of emotions in vocal expression and music performance: Different channels, same code?, Psychol. Bull. **129**, 770–814 (2003)
29.90 Z. Eitan, R.Y. Granot: How music moves: Musical parameters and listeners images of motion, Music Percept. **23**(3), 221–248 (2006)
29.91 C. Spence: Crossmodal correspondences: A tutorial review, Atten. Percept. Psychophys. **73**(4), 971–995 (2011)
29.92 A.B. Hostetter, M.W. Alibali: Visible embodiment: Gestures as simulated action, Psychonomic Bull. Rev. **15**(3), 495–514 (2008)
29.93 G. Lakoff, M. Johnson: The metaphorical structure of the human conceptual system, Cogn. Sci. **4**(2), 195–208 (1980)
29.94 L.E. Crawford: Conceptual metaphors of affect, Emot. Rev. **1**(2), 129–139 (2009)
29.95 A.C. Bolders, G.P. Band, P.J. Stallen: Evaluative conditioning induces changes in sound valence, Front. Psychol. **3**, 106 (2012)
29.96 J.B. Davies: *The Psychology of Music* (Hutchinson, London 1978)
29.97 F.S. Barrett, K.J. Grimm, R.W. Robins, T. Wildschut, C. Sedikides, P. Janata: Music-evoked nostalgia: Affect, memory, and personality, Emotion **10**(3), 390–403 (2010)
29.98 K.R. Scherer: Emotion. In: *Introduction to Social Psychology: A European Perspective*, 3rd edn., ed. by M. Hewstone, W. Stroebe (Blackwell, Oxford 2000) pp. 151–191
29.99 R.S. Lazarus: Progress on a cognitive-motivational-relational theory of emotion, Am. Psychol. **46**(8), 819–834 (1991)
29.100 K.R. Scherer: Appraisal considered as a process of multilevel sequential checking. In: *Appraisal Processes in Emotion: Theory, Methods, Research, Series in Affective Science*, ed. by K.R. Scherer, A. Schorr, T. Johnstone (Oxford Univ. Press, Oxford 2001) pp. 92–120
29.101 P. Silvia: Emotional responses to art: From collation and arousal to cognition and emotion, Rev. General Psychol. **9**(4), 342 (2005)
29.102 K. Millis: Making meaning brings pleasure: The influence of titles on aesthetic experiences, Emotion **1**(3), 320–329 (2001)
29.103 T. DeNora: *Music in Everyday Life* (Cambridge Univ. Press, Cambridge 2000)
29.104 D. Hesmondhalgh: Towards a critical understanding of music, emotion and self-identity, Consum. Mark. Cult. **11**(4), 329–343 (2008)
29.105 H. Leder, B. Belke, A. Oeberst, D. Augustin: A model of aesthetic appreciation and aesthetic judgments, Br. J. Psychol. **95**(4), 489–508 (2004)
29.106 D. Huron: Why is sad music pleasurable? A possible role for prolactin, Musicae Scientiae **15**(2), 146–158 (2011)
29.107 C.L. Krumhansl: An exploratory study of musical emotions and psychophysiology, Can. J. Exp. Psychol. **51**(4), 336–352 (1997)
29.108 P. Laukka, T. Eerola, N.S. Thingujam, T. Yamasaki, G. Beller: Universal and culture-specific factors in the recognition and performance of musical emotions, Emotion **13**(3), 434–449 (2013)
29.109 K.R. Scherer, M.R. Zentner: Emotional effects of music: Production rules. In: *Music and Emotion: Theory and Research*, ed. by P.N. Juslin, J.A. Sloboda (Oxford Univ. Press, New York 2001) pp. 361–392

29.110 T. Fritz, S. Jentschke, N. Gosselin, D. Sammler, I. Peretz, R. Turner, A.D. Friederici, S. Koelsch: Universal recognition of three basic emotions in music, Current Biol. **19**(7), 573–576 (2009)

29.111 T. Eerola: Are the emotions expressed in music genre-specific? An audio-based evaluation of datasets spanning classical, film, pop and mixed genres, J. New Music Res. **40**(4), 349–366 (2011)

29.112 P. Saari, T. Eerola, M. Barthet, G. Fazekas, O. Lartillot: Genre-adaptive semantic computing and audio-based modelling for music mood annotation. In: *IEEE Trans. Audio Speech Lang. Process. (TASLP)* (2015), https://doi.org/10.1109/TAFFC.2015.2462841

29.113 J.K. Vuoskoski, T. Eerola: Extramusical information contributes to emotions induced by music, Psychol. Music **43**(2), 262–274 (2015)

29.114 S.O. Ali, Z.F. Peynircioğlu: Songs and emotions: Are lyrics and melodies equal partners?, Psychol. Music **34**(4), 511–534 (2006)

29.115 L.-L. Balkwill, W.F. Thompson, R. Matsunaga: Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners, Jpn. Psychol. Res. **46**(4), 337–349 (2004)

29.116 G. Kreutz, U. Ott, D. Teichmann, P. Osawa, D. Vaitl: Using music to induce emotions: Influences of musical preference and absorption, Psychol. Music **36**(1), 101–126 (2008)

29.117 E. Bigand, S. Vieillard, F. Madurell, J. Marozeau, A. Dacquet: Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts, Cogn. Emot. **19**(8), 1113–1139 (2005)

29.118 J.K. Vuoskoski, T. Eerola: Measuring music-induced emotion: A comparison of emotion models, personality biases, and intensity of experiences, Musicae Scientiae **15**(2), 159–173 (2011)

29.119 J.K. Vuoskoski, T. Eerola: The role of mood and personality in the perception of emotions represented by music, Cortex **47**(9), 1099–1106 (2011)

29.120 P. Juslin, S. Liljeström, D. Västfjäll, G. Barradas, A. Silva: An experience sampling study of emotional reactions to music: Listener, music, and situation, Emotion **8**(5), 668–683 (2008)

29.121 W.M. Randall, N.S. Rickard: Development and trial of a mobile experience sampling method (m-eSM) for personal music listening, Music Percept. **31**(2), 157–170 (2013)

29.122 H. Egermann, M.E. Sutherland, O. Grewe, F. Nagel, R. Kopiez, E. Altenmüller: Does music listening in a social context alter experience? A physiological and psychological perspective on emotion, Musicae Scientiae **15**(3), 307–323 (2011)

29.123 H. Egermann, R. Kopiez, E. Altenmüller: The influence of social normative and informational feedback on musically induced emotions in an online music listening setting, Psychomusicol. Music Mind Brain **23**(1), 21–32 (2013)

29.124 M.R. Zentner, T. Eerola: Self-report measures and models. In: *Handbook of Music and Emotion: Theory, Research, Applications*, ed. by P.N. Juslin, J.A. Sloboda (Oxford Univ. Press, Oxford 2010) pp. 187–221

29.125 H.-R. Peltola, T. Eerola: Fifty shades of blue: Classification of music-evoked sadness, Musicae Scientiae **20**(1), 84–102 (2015)

29.126 L.-O. Lundqvist, F. Carlsson, P. Hilmersson, P.N. Juslin: Emotional responses to music: Experience, expression, and physiology, Psychol. Music **37**(1), 61–90 (2009)

29.127 C.V.O. Witvliet, S.R. Vrana: Play it again Sam: Repeated exposure to emotionally evocative music polarises liking and smiling responses, and influences other affective reports, facial EMG, and heart rate, Cogn. Emot. **21**(1), 1–23 (2007)

29.128 D. Hodges: Psychophysiological measures. In: *Handbook of Music and Emotion: Theory, Research, Applications*, ed. by P.N. Juslin, J.A. Sloboda (Oxford Univ. Press, Oxford 2010) pp. 279–312

29.129 O. Grewe, F. Nagel, R. Kopiez, E. Altenmuller: Listening to music as a re-creative process: Physiological, psychological, and psychoacoustical correlates of chills and strong emotions, Music Percept. **24**(3), 297–314 (2007)

29.130 A.J. Blood, R.J. Zatorre: Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion, Proc. Natl. Acad. Sci. **98**(20), 11818–11823 (2001)

29.131 U.M. Nater, E. Abbruzzese, M. Krebs, U. Ehlert: Sex differences in emotional and psychophysiological responses to musical stimuli, Int. J. Psychophysiol. **62**(2), 300–308 (2006)

29.132 S. Koelsch, T. Fritz, D.Y. v. Cramon, K. Müller, A.D. Friederici: Investigating emotion with music: An fMRI study, Hum. Brain Mapp. **27**(3), 239–250 (2006)

29.133 E. Altenmüller, K. Schuermann, V.K. Lim, D. Parlitz: Hits to the left, flops to the right: Different emotions during listening to music are reflected in cortical lateralisation patterns, Neuropsychologia **40**(13), 2242–2256 (2002)

29.134 L.A. Schmidt, L.J. Trainor: Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions, Cogn. Emot. **15**(4), 487–500 (2001)

29.135 E. Flores-Gutiérrez, J. Díaz, F. Barrios, R. Favila-Humara, M. Guevara, Y. del Río-Portilla, M. Corsi-Cabrera: Metabolic and electric brain patterns during pleasant and unpleasant emotions induced by music masterpieces, Int. J. Psychophysiol. **65**(1), 69–84 (2007)

29.136 T. Colibazzi, J. Posner, Z. Wang, D. Gorman, A. Gerber, S. Yu, B.S. Peterson: Neural systems subserving valence and arousal during the experience of induced emotions, Emotion **10**, 377–389 (2010)

29.137 T. Ball, B. Rahm, S.B. Eickhoff, A. Schulze-Bonhage, O. Speck, I. Mutschler: Response properties of human amygdala subregions: evidence based on functional MRI combined with probabilistic anatomical maps, PLoS One **2**(3), e307 (2007)

29.138 V. Menon, D. Levitin: The rewards of music listening: Response and physiological connectivity of the mesolimbic system, Neuroimage **28**(1), 175–184 (2005)

29.139 V.N. Salimpoor, M. Benovoy, K. Larcher, A. Dagher, R.J. Zatorre: Anatomically distinct dopamine release during anticipation and experience of peak emotion to music, Nature Neurosci. **14**(2), 257–262 (2011)

29.140 M.T. Mitterschiffthaler, H.Y. Cynthia, C.H.Y. Fu, J.A. Dalton, C.M. Andrew, S.C.R. Williams: A functional MRI study of happy and sad affective states induced by classical music, Human Brain Mapp. **28**, 1150–1162 (2007)

29.141 L. Nadel: Hippocampus and context revisited. In: *Hippocampal Place Fields: Relevance to Learning and Memory*, ed. by S. Mizumori (Oxford Univ. Press, New York 2008) pp. 3–15

29.142 S. Koelsch, A. Remppis, D. Sammler, S. Jentschke, D. Mietchen, T. Fritz, H. Bonnemeier, W.A. Siebel: A cardiac signature of emotionality, Eur. J. Neurosci. **26**(11), 3328–3338 (2007)

29.143 N. Gosselin, I. Peretz, E. Johnsen, R. Adolphs: Amygdala damage impairs emotion recognition from music, Neuropsychologia **45**, 236–244 (2007)

29.144 D. Sammler, M. Grigutsch, T. Fritz, S. Koelsch: Music and emotion: Electrophysiological correlates of the processing of pleasant and unpleasant music, Psychophysiology **44**(2), 293–304 (2007)

29.145 I. Daly, A. Malik, F. Hwang, E. Roesch, J. Weaver, A. Kirke, D. Williams, E. Miranda, S.J. Nasuto: Neural correlates of emotional responses to music: an EEG study, Neurosci. Lett. **573**, 52–57 (2014)

29.146 B. Schuller, J. Dorfner, G. Rigoll: Determination of nonprototypical valence and arousal in popular music: Features and performances, EURASIP J. Audio Speech Music Process. **2010**(5), 735854 (2010)

29.147 D.R. Turnbull, L. Barrington, G. Lanckriet, M. Yazdani: Combining audio content and social context for semantic music discovery. In: *Proc. 32nd Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.* (2009) pp. 387–394

29.148 Y.E. Kim, E. Schmidt, L. Emelle: Moodswings: A collaborative game for music mood label collection. In: *Proc. Int. Symp. Music Inf. Retr* (2008) pp. 231–236

29.149 E.L. Law, L. Von Ahn, R.B. Dannenberg, M. Crawford: TagATune: A game for music and sound annotation. In: *Int. Conf Music Inf. Retr. (iSMIR'07)*, Vol. 3 (2007) pp. 361–364

29.150 L. Barrington, D. Turnbull, G. Lanckriet: Game-powered machine learning, Proc. Natl. Acad. Sci. **109**(17), 6411–6416 (2012)

29.151 P. Saari, T. Eerola: Semantic computing of moods based on tags in social media of music. In: *IEEE Trans. Knowl. Data Eng* (2013), https://doi.org/10.1109/TKDE.2013.128

29.152 S. Koelsch: Music and emotion. In: *The Cambridge Handbook of Human Affective Neuroscience*, ed. by J. Armony, P. Vuilleumier (Cambridge Univ. Press, New York 2013) pp. 286–303

29.153 R. Bresin, A. Friberg: Emotion rendering in music: Range and characteristic values of seven musical variables, Cortex **47**(9), 1068–1081 (2011)

29.154 T. Eerola, A. Friberg, R. Bresin: Emotional expression in music: contribution, linearity, and additivity of primary musical cues, Front. Psychol. **4**(487), 1–12 (2013), https://doi.org/10.3389/fpsyg.2013.00487

29.155 P.N. Juslin, G. Barradas, T. Eerola: From sound to significance: Exploring the mechanisms underlying emotional reactions to music, Am. J. Psychol. **128**(3), 281–304 (2015)

29.156 E. Schubert: Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space, Aust. J. Psychol. **51**(3), 154–165 (1999)

29.157 E. Schubert: Modeling perceived emotion with continuous musical features, Music Percept. **21**(4), 561–585 (2004)

29.158 J.A. Sloboda, A.C. Lehmann: Tracking performance correlates of changes in perceived intensity of emotion during different interpretations of a Chopin piano prelude, Music Percept. **19**(1), 87–120 (2001)

29.159 V. Alluri, P. Toiviainen, I.P. Jääskeläinen, E. Glerean, M. Sams, E. Brattico: Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm, NeuroImage **59**(4), 3677–3689 (2012)

Part D

Psychoph

Part D Psychophysics/Psychoacoustics

Ed. by Albrecht Schneider

This part deals with sensation and perception of pitch, timbre, and loudness in humans, largely from a psychoacoustic perspective. The part is organized into four chapters. Chapter 30 covers fundamentals of psychophysics in regard to sensation and perception including some theoretical and historical background information (in particular on the concepts of Fechner and of Stevens) as well as a detailed discussion of the so-called *sone* scale and the *mel* scale, taken as examples for scaling of subjective loudness and pitch respectively. Also included in this chapter is a section on types of sound phenomena relevant for auditory perception, and in particular for sensation and perception of musical sound.

Chapter 31 addresses sensation and perception of pitch mainly from a functional perspective. Anatomical and physiological facts concerning the auditory pathway are provided to the extent necessary to understand excitation processes resulting from sound energy in the middle ear as well as within the cochlea. Place coding and temporal coding of sound features are viewed as two parameters relevant for pitch perception, in regard to frequency and period The Wiener–Khintchine theorem is taken as a basis to explain the correspondence between temporal periodicity and spectral harmonicity as two principles fundamental to perception of pitch and timbre. The basics of some models of the auditory periphery suited to extracting pitch from complex sounds either in the time or frequency domain is outlined, along with examples demonstrating how such models work for certain sounds. Sections of this chapter also address tone height and tonal quality as components of pitch as well as the rather dubious nature of the so-called tone chroma. Issues such as isolating tone quality from height (as in *Shepard tones*) and an alleged preference of subjects for stretched octaves are covered in a critical assessment. Section 31.6 on psychophysics includes just-noticeable difference (JND) and difference limen (DL) for pitch, the concept of auditory filters known as critical bands, the sensation of roughness and dissonance as well as special pitch phenomena (the *residue* and the missing fundamental, the concept of virtual pitch, and combination tones). Another section covers spectral fusion, Stumpf's concept of *Verschmelzung*, and the sensation of consonance. Further, there are sections on categorical pitch perception as well as on absolute and relative pitch followed by a brief survey of scale types, tone systems and intonation. The chapter closes with a section on geometric pitch models and some basic features of tonality in music.

Chapter 32 deals with perception of *timbre* or *sound color*. Both concepts can be distinguished by terminology as well as their historical and factual background, even though both relate to some common features for which an objective (acoustic) basis exists. Sections of this chapter review, in brief, developments in (traditional and electronic) musical instruments as well as in research on timbre and sound color. Section 32.2 on sensation and perception of timbre offers a retrospect on classical concepts of tone color or sound color and reviews some modern approaches from Schaeffer's *objet sonore* to semantic differentials and multidimensional scaling. Taking a functional approach, acoustical features (such as transients and modulation) and perceptual attributes of timbre as well as interrelations between *pitch* and *timbre* are discussed. In a final section, fundamentals of sound segregation and auditory streaming are outlined. For most of the phenomena covered in this chapter examples are provided, including sound analyses obtained with signal processing methods.

Chapter 33 is on sensation of sound intensity and perception of loudness. Since some of the relevant matter (on scaling concepts of loudness) has been presented in Chap. 30 of this part, and because a considerable portion of research on loudness is done outside musical contexts (namely, in industrial and environmental noise control as well as in audiology), this chapter condenses facts and models more than the previous ones on pitch and timbre. Section 33.1 offers physical and physiological bases for sound intensity sensation while Section 33.2 discusses features of some models of loudness sensation that have been established in psychoacoustics over the past decades. Since these models originally were designed for stationary sound signals and levels, and have been tested mostly in lab situations, they could not adequately cover a range of real-world sound types found in natural or technical environments. Section 33.3, therefore, discusses features of sound material contained in techno music productions and presented, for example, to audiences in live music clubs at very high sound levels, which calls for measurement of relevant sound parameters (such as the rate of energy transmitted by pulse sequences embedded in music) as well as an appropriate assessment of sensory effects. Different from perception of pitch (where samples of subjects respond more or less in similar ways to certain types of sound signals), perception of loudness shows a high degree of variability even within groups of musically trained subjects reflecting their musical background and preferences. Finally,

there is a concluding section in which recent empirical evidence demonstrating that subjects judge loudness for various musical genres on an ordinal category scale (from very soft to very loud) is presented. However, the range for such scales and the absolute sound pressure levels that correspond to certain categories differ widely among subjects, calling into question the concept of a unitary scale for loudness (such as the *sone* scale).

30. Fundamentals

Albrecht Schneider

This part of the handbook deals with sensation and perception of pitch, timbre, and loudness in humans, largely from a psychoacoustic perspective. Since there is broad range of publications available on subjects such as hearing (including anatomy and physiology), psychoacoustics as well as signal processing in relation to perceptual modeling (for comprehensive summaries see, e.g., [30.1–12]), it would be quite difficult, if possible at all, to condense relevant matter that has found detailed discussion elsewhere so as to fit into one part of this handbook. Rather, the approach taken here is selective such that phenomena that have found extensive treatment in works on psychoacoustics and audio processing (e.g., masking) will only be briefly addressed while a number of aspects usually given less attention shall be included. The perspective chosen aims at presenting facts and models but also turns to theoretical and methodological issues deemed necessary to understand lines of development in research. To this end, Chap. 30 of this part addresses fundamental concepts such as sensation, perception, and apperception. Since such concepts have been developed in a long process of research, and from certain philosophical backgrounds, it seems adequate to refer to at least some of the discussion found in disciplines such as philosophy, psychology, and neuroscience with respect to epistemology and research strategies. Further, ideas that are of special interest in regard to the history of psychophysics and in particular concepts developed by Theodor Fechner and Stanley Stevens are given a critical examination. To illustrate certain facts or problems, examples are provided (such as sound analyses or other empirical data). Though music perception in humans seems to be unique with respect to the involvement of cognitive factors, in research involving sensation and perception (e.g., of pitch and loudness) reference must also be made to other mammals we share basic anatomical, physiological and neuronal structures and functions with [30.13, 14]. Sensation and perception of pitch and timbre basically is viewed as a functional relation between certain types of sound and processing of sounds within the sensory organ as well as on several levels along the auditory pathway (AuP). Functional in this context means that sensations and perceptions in general can be related to features inherent in sounds and that, notwithstanding variability in capabilities and performance among individuals, a relation of cause and effect holds that permits us to assume intra-individual as well as interindividual similarity and consistency of sensations and perceptions for the same set of stimulus conditions. At least it seems reasonable, as a working hypothesis, to assume the same objective causes will provoke similar effects, in human subjects, within a certain range.

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_30

30.1 Theoretical and Methodological Background

Dealing with perception and cognition presupposes decisions as to epistemology and methodology. Over the past decades, there has been an intensive discussion in philosophy, psychology, and neuroscience, as well as in areas such as artificial intelligence, concerning fundamental issues like: is there any justification for separating the physical brain and its specific functions from intellectual capacities subsumed under the heading of *mind*? What is the relation of neural brain states to our experience of *consciousness* and *self-consciousness*? What exactly constitutes a *mental representation* of something? How is cognition accomplished in regard to complex tasks such as production and understanding of language or music?

30.1.1 Realism, Naturalism, Reductionism, Empiricism

A review of ongoing debates on, for example, monistic or dualistic approaches to brain and mind is beyond the scope of the present article. For in-depth treatment of central issues in philosophy, psychology and cognitive sciences, the reader is referred to publications such as [30.15–24]. We may, however, briefly address four basic concepts that have played a major role in almost all discussions of brain/mind relations as well as in regard to fundamentals of perception and cognition. These four concepts can be stated as:

1. Realism
2. Naturalism
3. Reductionism
4. Empiricism.

Realism

Ontological realism accepts there is a material world that exists irrespective of human perception and consciousness. This world contains a manifold of things including plants and animals that make up types of environments. Humans are one species (Homo sapiens) of primates living in various environments. Things have properties (e.g., solids have a mass and a density). Epistemological realism holds that humans are capable of gaining valid and reliable knowledge of things existing in this world as well as of physical, chemical, and biological processes governing nature in all its aspects. The basic ontological assumption that there is a physical world independent of the mind is idealistic since, according to *Rescher* [30.25, p. 115],

> *ultimately, it does not represent a discovered fact, but a methodological presupposition of our praxis of inquiry; it is not constitutive (fact-descriptive) but regulative (praxis-facilitating).*

While *naïve* realism guides much of our everyday experience of the world, *critical* realism seeks for principles that make coherent experience possible for individual subjects, on the one hand, and that limit our knowledge derived from such experience, on the other. For example, knowledge based on experience beyond the single case is achieved by prototyping and categorization [30.26]. Experience involves identification of things or objects presented under varying conditions as being *the same* or belonging to *the same category*. Critical realism concedes, however, that perception and even sensation, which is at the base of much of our experience, can incorporate subjective factors (giving rise to the variance found in the data from samples in experiments on sensation and perception).

Naturalism

Naturalism holds that this world (and also the universe) is governed by natural laws that are discovered by science, and suited to explain things that make up the world as well as their relations and functions. Hence, valid and reliable knowledge is as such based on natural laws and acquired by the scientific method. In a long and ongoing process of inquiry, scientific research revealed that this world consists of many systems and subsystems, many of which are dynamic in nature (as is evident from, for example, the study of molecular genetics, or of relations between climate and vegetation). Conducting *systematic* inquiry according to scientific methodology, and thereby discovering the principal *systematicity* of nature, has led to results that also can claim systematicity in regard to coherence and explanatory power. As a consequence, naturalism (that either excludes or scrutinizes whatever is deemed *supernatural*) rests on scientific methodology comprising empirical observation, controlled experiments, measurement, and mathematical modeling. It has set the standard for research including disciplines such as psychology as well as major areas of sociology.

The *naturalist* approach common now in cognitive science and psychology, as well as in philosophy, proposes that our mind, and in particular phenomenal consciousness, ultimately rests on brain structures and their functions [30.19, 20, 22]. However, rather than taking consciousness as an epiphenomenon, one may view consciousness as emerging from increasingly more complex brain structures and processes in the course of evolution that led to Homo sapiens [30.27, 28]. To explain mental phenomena in terms of neuronal activity is

an attempt at overcoming the mind/body or mind/brain dualism. It should be noted, though, that even strong dualistic concepts such as advanced by *Karl Popper* and *John Eccles* [30.27] acknowledge that mental operations rest on brain states. The dualistic concept rather implies that there are certain states of the mind and processes of *mental activity* that as yet cannot be sufficiently explained in terms of brain states, and which, for the specifics of experiences addressed as *mental*, deserve to be studied from a psychic and intellectual point of view. In this context, the sphere of *intentionality* is a matter of debate. In regard to intentionality, subjects are concerned with analyzing and evaluating their experiences and motivations, they attribute meaning to things or persons, form opinions, have beliefs or have expectations, they can understand the double meaning of certain sentences or jokes. Though all such acts can, ultimately, be performed only because subjects have a brain, it is not quite clear how brain operations may lead to specific mental experiences usually subsumed under *intentionality* [30.29]. One will remember *Brentano*'s [30.30] notion of psychic objects for which *intentional inexistence*, that is, mental immanence, obtains. According to Brentano, in an act of conceiving something is conceived as an object, in judgment something is judged (as true or false), etc. While *outer perception* effected by means of sensory organs leads to seeing certain colors or hearing certain tones, *inner perception* involves the act of hearing (of which we become conscious) as well as conceptualizing phenomena so that they are *evident*. In this respect, the German term for perceiving includes judgment: the term *wahrnehmen* literally means to take something for granted and *evident* that has been perceived properly. To *Husserl* [30.31], the manifold of inner perceptions we can bring about and the experience of *evidence* is an implication of intentionality. Husserl maintained that psychic phenomena thereby can be distinguished from physical phenomena.

Within a naturalist and monistic approach, one would argue that the brain realizes neural states that somehow represent mental states, and that neural states can in principle account for intentionality, and for mental acts such as thinking, desiring, imagining. Also, it is common to adopt a third-person approach to mental phenomena and to avoid, for reasons of scientific objectivity, a first-person perspective as was often adopted in works on the philosophy of mind as well as in psychology when dealing with *phenomenal consciousness*.

There are various models of how *phenomenal content* can be represented. An issue that has caused major controversy is that of subjective experience of certain things or states as evoking a specific *quality*. From a monistic and naturalistic view, the existence of *qualia* beyond *the sum total of all the innate and learned associations and reactive dispositions* has been denied [30.15, p. 388], or has been turned into a problem of representational states [30.22]. However, the tendency of many subjects to include properties of *qualia* into their experiences can be explained within the naturalistic framework by a dispositionalist higher-order theory of phenomenal consciousness [30.20].

Reductionism

Reductionism is a basic strategy in science aimed at understanding complex systems by reducing, for example, the number of variables operative in a model or the number of dimensions representing a system. In this way, in acoustics, vibration of strings often is treated as if the string is a one-dimensional continuum (where an *ideal string* is assumed to be of extreme or even infinite length and so thin that string diameter is negligible when compared to string length). Further, it is assumed that elongation of each mass point of the string is very small (so that the restoring force also is very small according to Hooke's law, and vibration amplitude is a linear function of excitation force). Similarly, a number of cochlea models have been constructed as one-dimensional, and very small deflections of basilar membrane (BM) structures are assumed in order to maintain linearity for certain parts and processes where in fact nonlinear behavior can be observed already for medium sound pressure level (SPL), and more so for high SPL.

Reduction of complexity is a necessary step to develop heuristics for solving certain problems. In some philosophical and neuropsychological approaches to perception and cognition, however, reductionism has gained a different status. Obviously not the least because of disappointment with perennial epistemological disputes in philosophy, and confessing *eroded confidence in 'the grand old paradigm', a framework derived mainly from Logical Empiricism* [30.32, p. 546] since the days of Hume, prospects that current scientific endeavors such as cognitive neurobiology and neural network models might hold for the reductionist epistemology were considered more promising than the continuation of traditional discourses. One has to remember, in this context, certain developments and debates in areas that, taken together, would make up *cognitive science* (for details, see [30.33]). To name but a few of the central issues, there had been the brain/computer analogy and the metaphor of the mind carrying out procedures similar to operations of a computer program that were of consequence for conceptualization and heuristics since about 1950 [30.34, Chap. I,2]. Further, findings in neuroscience according to which cortical areas and other brain structures are

said to work like task-specialized neural *modules* seems to have inspired concepts that suggested the *modularity of the mind* as a functional equivalent to the brain. Also, models of neural networks (comparable, at least in principle, to operations found in neurons, cell assemblies and similar substrates) soon proved the usefulness of the *connectionist* approach. These developments all had their impact on attempts at establishing *neurophilosophy* as well as philosophically oriented neuroscience that seeks to overcome the dichotomy of brain and mind [30.16, 18, 35, 36]. Common to such approaches is to explain perception and higher-order phenomena of cognition on the basis of lower-order phenomena, and to take into account facts and models available from biology and neuroscience as well as from computing [30.17]. Processes pertaining to perception and cognition are thus addressed *bottom-up* from their natural foundations rather than along abstract principles. Such approaches have been labeled *reductionist* insofar as cognition is investigated and explained largely from a biological and also from a computational perspective (whereby *mental phenomena* might be *reduced* to representations of neural brain states). In line with both naturalism and reductionism is the thesis stating *that the function of cognition is to enable the agent to deal with environmental complexity* [30.37, p. 3]. In order to adapt to the environment in an optimal way, humans (and other mammals) had to develop their cognitive skills.

There are several approaches to reductionism that differ, among other aspects, in the way the monistic perspective is implemented. While, in a more radical perspective, some claim identity of mind and brain so that mind states are nothing but brain states (taken to be observable as such), others admit there are mind states that can be experienced by subjects, but which are *epiphenomena* derived from brain states. Still others suggest the sheer complexity of brain structures and processes can result in *emergent experiences* that are difficult to trace in neural substrates (an excess of unpredictable, creative brain activity has also been accessed in theories of *autopoiesis* as well as in ideas pertaining to *radical constructivism*). However, concepts of *emergence* besides self-organization also include the aspect of complex interactions between systemic components [30.21, pp. 114–116] giving rise to salient patterns that seem relevant for, among other areas, auditory perception.

Empiricism

Empiricism in regard to perception and cognition evolved from several different angles in the 19th century. Among relevant sources one will have to reckon sensualism as proposed by John Locke and the belief that knowledge for the individual results largely from experiences that make up an aggregate. Further, association psychology (as sketched by James Mill) and inductive logic as developed by John Stuart Mill became of importance for scientists like Helmholtz and Mach who both advocated empiricism in the study of sensation and perception. The paradigm developed by Helmholtz can be described under the heading of *experimental interactionism*. To Helmholtz, acquiring experience starts from elementary sensations resulting from the input received from peripheral sensory organs. In regard to vision (which he uses for example), however, we can control what we actually see by moving our head and directing our eyes towards certain objects. Such movements are effected by means of innervations, which in turn can be activated through volitional impulses. Hence, the observer can select by innervations and motor control what he or she perceives, and what becomes *present* from a given object or from a range of different objects of which we have a sequence of sensations. According to Helmholtz, directed observation is an active process suited to realize that, by willful innervations, sensations can be turned into perceptions where certain aspects or features become clear. To Helmholtz, a lawful relation exists between (willful) innervation and perceptual focus. Taking the example of seeing objects in a space, *Helmholtz* [30.38, p. 237] argued that each willful movement of our head and eyes constitutes an elementary *experiment* since it leads to changing the phenomenal appearance of objects and permits us to check whether we have correctly realized its distinctive features.

We all perform many such perceptual *experiments* in ontogenesis in order to build up experience. Also, repeated perceptions that contain features *typical* of certain objects leave traces in our memory. Further, *Helmholtz* [30.38, p. 233] pointed to a cognitive mechanism he at times has called *unconscious inferences*. According to this concept, current sensation leads to a perception that makes use of a sum of previous experiences (resulting in, for example, *prototypes* or other typifications/generalizations) stored in memory. The sum of such experiences constitutes what, in modern terminology, would be *implicit knowledge*; *Helmholtz* [30.39, p. 85] distinguished between *Kenntnis*, i. e., knowledge accumulated through experience, and *Wissen*, that is, verbalized knowledge. For example, sensing the sound of a trumpet may evoke the instant and perhaps unconscious inference in a subject that it is a *brass instrument* one has heard. In regard to empiricist methodology, the concept of *unconscious inference* can be taken as applying inductive generalization to perception [30.40, Chap. 3]. Helmholtz' concept implies that, due to an ongoing process of acquiring

experience in each subject, sensory input is likely to *trigger* inferences (from categories, prototypes, learned schemata, etc. stored in memory) that influence actual perception. Hence, as has been discussed with respect to loudness perception and other areas, the perceptual response to physical stimuli, though based on peripheral pickup of sensory data and processing along specific neural pathways, typically involves also cognitive structures [30.41]. Of course, empiricism also had consequences in regard to methodology. In psychology (including psychophysics, [30.42]) and in the area that would later become psychoacoustics, experimental investigations aiming at *measurable* attributes of sensation and perception (Sect. 30.1.4) were, from the very beginning, deemed necessary.

30.1.2 Sensation, Perception, Apperception, Imagination

Historically [30.43], Thomas Reid perhaps was the first to make a clear distinction between sensation and perception. Summarizing his ideas, sensation is viewed as causing a change in experience, while perception is said to include a conception of an object or a relationship that is perceived, plus the immediate and irresistible conviction of the existence of such object or of a spatial organization [30.44, p. 88]. From Helmholtz's writings as well as from works by Wilhelm Wundt and other psychologists of the time, a useful distinction between concepts becomes evident:

1. Sensation
2. Perception
3. Apperception
4. Imagination.

A short summary of central features of these four concepts is given here that serves also as an introduction to the next section where a functional model of sensation and perception is outlined.

Sensation

By sensation (German: *Empfindung*) typically a process involving sensory organs and addressing *the senses* (vision, hearing, etc.) is understood, which starts at the periphery (skin, eyes, ears, tongue, nose) with some physical and/or chemical stimulus and leads to a mechanical (as in our ears) or other excitation that is followed by a physiological reaction that in turn results in a neural signal [30.45]. Such a signal can immediately trigger a fast response (subcortical, e.g., brainstem reflexes) that usually includes some form of motor behavior or other action suited to prevent damage from sensory organs or from other parts of the body, and to adjust to a new situation. Sensations, which may start also from within the body (e.g., visceral sensations, headache) can, yet must not, rise to some level of awareness depending on the conditions of the stimulus as well as on sensitivity of the sensory organ and the ascending neural pathway (for which a threshold is assumed; Chap. 31). Sensations, which somehow correlate to the neurally coded pattern of stimulation/excitation, often include emotive coloring (e.g., sensation of a very loud sound may go along with affright or even horror). Sensation can be regarded as based on a direction outward → inward, that is, from peripheral sensory organs used for *stimulus information pickup* to processes within the central nervous system. In general, for the emergence of perceptions, circuits comprising receptors (sensory organs) and neural transmission lines to the central nervous system (CNS) are required [30.46, Chap. 14]; such circuits also can include effectors whereby the system responds to sensory input [30.47, Chap. 5]. In psychology and psychophysiology, it has been customary for a long time ([30.48, p. 273 ff.], [30.49, p. 30 ff.]) to categorize sensations as to their quality, intensity, duration, extension, etc. If a subject notices qualities and/or intensities, his or her sensations must entail some elementary categorization process. It has been argued at times that subjects are likely to become aware of peripheral stimuli if the level of excitation exceeds a certain threshold (so as to bring the sensation to the subject's attention). In this respect, a certain level in intensity of a sensation would give rise to a perception while low-intensity sensations would remain unnoticed. In psychophysics, researchers since the days of Weber and Fechner have drawn relations between the physical intensity of stimuli presented to sensory organs and the intensity of sensations (Sect. 30.1.4, *Historical and Theoretical Background*).

Perception

Perception in this respect is directed towards the *content* of the sensation whereby the specifics of this content are interpreted in regard to certain properties or features (*what* is it that has been sensed?). In a straightforward *realist* approach, the content of what has been sensed should be explainable in an unambiguous way since, in such a perspective, sensation of *something* is causally related to the nature of the stimulus and the transfer function of the sensory channel. This aspect will be discussed in regard to pitch perception in more detail in Chap. 31.

The faculty to perceive, however, can also be understood as a general activity of interacting with the environment to which our senses are directed. For instance, walking in a natural or in a man-made environment (comprising spatial and temporal order) with eyes

and ears open implies that we see very many objects as they enter the field of vision, and we hear sounds as they appear in a sound field. At this stage, perceiving means that we distinguish objects on the basis of their phenomenal appearance where we may identify certain objects along some salient features. For the proponent of a *direct realism*, perceiving (seeing, hearing) generates

> *immediate knowledge* [...] *which is not inferred from, or suggested by, any further knowledge, or any ground or basis for knowledge.* [30.50, p. 191]

Such knowledge can, yet must not, be verbalized. Perceiving as a quasicontinuous process means things in a field of vision or in a sound field are captured in what could be called first-order approximation, where things become perceptual objects (often for a short time only, in kind of a *frame-by-frame* mode) and are categorized in a rough or global access with little if any attention to details. One can address such roughly categorized objects as first-order percepts. However, even rough categorization involves judgment (and probably also some *unconscious inferences*). The *theater* where judgment pertaining to properties of things that are perceived as features of objects takes place traditionally is *the mind*; in line with current debates [30.20, 22], some stage of *phenomenal consciousness* can be assumed to account for these tasks. It was argued long ago that perceiving, in particular of visual objects, involves operations that in certain respects are similar to geometrical measurement (or at least to making estimates of geometrical properties) as well as abstraction. For example, the observer, after noticing some objects, accomplishes perception of them by correctly identifying relevant properties under varying, and often poor, conditions [30.51]. Within this process, cognitive evaluation is needed to facilitate and to complete perception.

Apperception

According to a model outlined by *Wundt* [30.52], perceiving as an ongoing process involves a steady influx of things to become objects of perception as they enter the field of vision or as they emerge (as events in a temporal sequence as well as in some spatial coordinates) in a sound field. From the perspective of an average viewer or listener, following a brief and global categorization, most of such objects disappear from phenomenal consciousness with the influx of new things or events. It is only if things or events become the focus of attention as objects of perception that phenomenal consciousness gains a detailed understanding of their structural organization. In such a situation, objects are singled out of a quasicontinuous sequence of first-order percepts in an active process that *Wundt* [30.52, p. 244 ff.] has described as *apperception*. It requires the viewer or listener to focus on a perceptual object with due attention whereby its *content* becomes clear in regard to structural composition and functional aspects. For example, the relational structure of tones and intervals making up complex chords such as used by Wagner in his *Tristan* will hardly be grasped in a situation where a subject is listening to the music as it is performed in an opera house. One can analyze such chords though by attentive listening to portions of recorded music in several runs with the aim of determining as many intervals (representing the relations between tones in each chord) as is possible. The results of the analysis, i. e., the number and topological structure of tones in a given complex chord can be visualized mentally, for example, by means of a two-dimensional (2-D) or three-dimensional (3-D) tone net (as to tone nets, see [30.53]). Also, one can label certain chords according to their tone and interval structure (e.g., *underseventh-ninth chord*). Similarly, listening closely to a melody or to several melodic lines interwoven in a polyphonic setting results in perceiving a *melodic contour* or several such contours each of which contains intervallic (diastematic) and temporal information. Such contours that can be graphically presented in a 2-D format (with *pitch* and intervallic relations on the ordinate, and time on the abscissa) are apparently perceptual objects whereby subjects conceive of melodic structures. *Wundt* [30.52] would have addressed perceptual objects such as melodic contours derived from analytic listening as *formations* (*Gebilde*), indicating that, on the one hand, a melody in fact does have an intrinsic structural property of order in regard to pitch and time, and, on the other, a subject attentively listening to such a melody will build a perceptual object of which he or she becomes aware of as a *formation*. To facilitate apperception, a subject needs a *mental representation* of features distinctive of an object that is thus apprehended (one may regard this process as a second-order perceptual approach). A *mental representation* understood both as a process and an entity resulting from the process is essential for apperception as well as for other cognitive operations. It can be characterized in certain ways (see also [30.54]). First, a representation typically is achieved in an active process of perceptual analysis involving judgment and also knowledge stored in memory. Second, though the basis from which analysis starts in general will be sensations (in this case, auditory), the final representation is the result of abstraction of which subjects become *aware* as such. The *content* of the object that is represented is mental, and may have the format of an *image*. Though the semantic field of the term *image* is quite wide (covering or alluding to meanings

such as *idea*, *conception*, *metaphor*, *picture*; [30.55]), it is central to an *image* to present or to re-present something. In this respect, the image of a melodic contour that we abstract from listening to a tune (say, *We shall overcome*) is a useful *format* for (mental) representation and memory storage of musical structure (also [30.56]). Visual imagery has been addressed as an efficient cognitive tool employed in learning and memory [30.57]. One important aspect of *representation* is that it can bring back certain objects and their content that resulted from previous perceptions and have been stored in some format in memory. Such objects are retrieved from memory (typically, long-term memory LTM) and are re-presented in a *mental state* that of course rests somehow on neural states. Third, since the process of abstraction *strips down* the sensory input so that structural and relational features are left over and condensed into a perceptual object, the *content* of this object serves as a model that represents essentials of the sensory input, and which also can represent (depending on conditions, see Chap. 31) properties of the physical stimulus. Perception includes not just categorization yet to some extent an interpretation of sensory data and even a tendency to *extrapolate* from sensory data in that perception, though based on sensory input, may lead to constructing, for example, *illusionary* objects as percepts. In regard to auditory perception, the Risset/Shepherd sound constructs of seemingly constantly falling or rising pitch (Sect. 31.1.2) is a good case in point.

Imagination

Auditory perception can go along with visual imagery whereby skilled musicians may form, for example, images of the score of the music they actually listen to. Imagery also is essential in conceiving of objects that must not necessarily exist in the environment but which could be imagined to exist (like the enchanted mermaid singing in the sea or some giant trumpets suited to produce a sound level that causes town walls to collapse).

30.1.3 Functional Model of Sensation and Perception

The basic relations of sensation to perception outlined in Sect. 30.1.2 can be condensed into a functional model (sketched also in [30.58, Chap. 4] and [30.46, Chap. 14]) that combines bottom-up and top-down processes. The source always consists of the environment that offers a wide range of (physical, chemical) stimuli, and of the mammalian body with its sensory organs capable of picking up such stimuli. Sensory organs work when a stimulus transmits a minimum quantity and intensity of energy needed for excitation. This means a certain energy transfer into the sensory system is required beyond a threshold to yield, as in the case of the ear, a mechanical and then a hydromechanical excitation that in turn will result in a physiological and neurophysiological response in the inner ear (for details, see [30.8, Chaps. 2–5]). At the core is a generation of sensor potentials and of action potentials as the result of electrochemical transduction processes that lead to neural spike patterns in the auditory nerve (AN) suited to code auditory-relevant information. Such information is routed *bottom-up* along the AuP and brings about auditory sensations, and also may evoke early responses to stimuli such as brainstem reflexes even before reaching the level of cortical sensory areas. In regard to auditory stimuli, it is likely that integration of incoming information and analysis as to stimulus features starts on a subcortical level (for which thalamic nuclei, in particular the medial geniculate body (CGM), have been addressed). Analysis of neural information on the CNS level, in regard to binding parts into more or less coherent objects and segregating objects from each other, involves previous experience and memory, that is, elementary or more complex top-down processes. In this respect, even a rough (and preliminary) categorization of incoming information in regard to stimulus features that might be compared to certain templates and prototypes activates processing beyond mere sensation. This level often is addressed as perception, for which at least some *lower* mental processes such as elementary comparison and categorization are needed; according to the distinction of *lower* and *higher* mental processes, the latter apply to cognition and for many tasks involve explicit knowledge that can, yet must not always, be verbalized. Acts of classification, though directed to actual stimulus features (e.g., timbres of musical instruments one hears at a certain moment), are supported by experience and knowledge. Access to previously stored information can remain implicit, especially in situations such as listening to music in a concert, which means quasi-*real-time* processing of sensory input and fast formation of perceptual objects on the basis of prototypes, schemata, etc. However, subjects can also draw on explicit knowledge (of which a subject becomes aware when making decisions and judgments) especially if they are skilled musicians. Further, even elementary categorization as well as subsequent detailed formation of perceptual objects can be influenced by motivation as well as by expectancies. In addition, perception of sounds almost always includes some emotive component, which is also an element in categorization and classification. Single sounds and textures of sounds are experienced as loud, shrill, dull, or as harmonious, smooth, etc. but also as joyous or frightening, as pleasant, soothing, etc.

As one may study in many situations and contexts, humans do not always experience clear demarcation lines between sensation and perception, on the one hand, and between perception and cognition, on the other. Among the reasons for such a crossover one has to reckon both the sheer amount of information from a broad range of (visual, auditory, olfactory etc.) stimuli that can affect our senses while the information processing capacity of each sense modality or *channel* is limited. For the auditory modality, the channel information capacity has been estimated as $K = 40\,000$ bit/s. Conscious processing of (auditory, visual, etc.) information by the CNS, however, is much more limited and may be as small as 100 bit/s or less [30.59, p. 156 ff.]. In many situations related to music and sound, density of flow of (auditory and often also visual) information is very high, hence processing requires significant reduction of stimulus information relative to only a few features or dimensions to maintain perception. Dimensional reduction is a typical response in perception to keep up with a heavy processing load or even total overflow of information per time unit, or with a heavy load of information that needs to be processed in parallel (the limits being largely set by the capacity of the working memory [30.60]). Of course, the use of templates and schemata helps in forming perceptual objects and in perceiving the relations between them. While templates even support elementary auditory perception (as will be discussed in regard to pitch perception in Chap. 31), learned schemata apparently play a role in understanding musical syntax in the process of listening [30.61, 62].

The sequence of processing that relates sensation to perception and cognition can be tentatively sketched in a block diagram (Fig. 30.1). Since hearing is the sense modality that is of prime interest here, the stimuli will be sounds (that may be combined with visual, tactile, or olfactory stimuli), and the relevant peripheral sensory organ of course is the pair of our ears.

Hearing is a complex process in which feedback loops are essential, and in which bottom-up and top-down mechanisms closely interact. Feedback circuits such as implemented in the brainstem reflexes and in centrifugal pathways [30.8, Chaps. 6–8] lead to adjustment of transduction parameters such as motility of the outer hair cells (OHCs).

30.1.4 The Measurement of Sensation

Historical and Theoretical Background

Psychophysics is a field of scientific study investigating the relations between physical stimuli and the sensations as well as perceptions evoked in subjects. Such relations are subjected to measurement (if possible), and are thus given quantitative formulations. The idea that even the *strength* of conceptions is quantifiable goes back to Johann Fr. Herbart and to the mathematician and philosopher Moritz W. Drobisch, who laid the foundations for a discipline labeled *mathematical psychology* [30.63]. In regard to empirical studies, Ernst H. Weber, a physiologist, had carried out experiments aiming at just noticeable differences (JNDs) in sensations of weight and touch. He found ratios of 39 : 40 (weights of two bodies lifted by hand) and 29 : 30 (sensitivity for pressure two different weights exert on a hand; [30.64, p. 546 ff.]) as differential thresholds. Weber said he himself had registered *many different degrees of sensation* to conclude that we must test our *innate instruments of sensation* just as a physicist is testing the sensitivity of his measuring instrument. Taking concepts from both Drobisch and Weber, Gustav Th. Fechner, a physician and physicist, offered a comprehensive treatise on fundamental issues of psychophysics [30.65]. He took off from the basic ques-

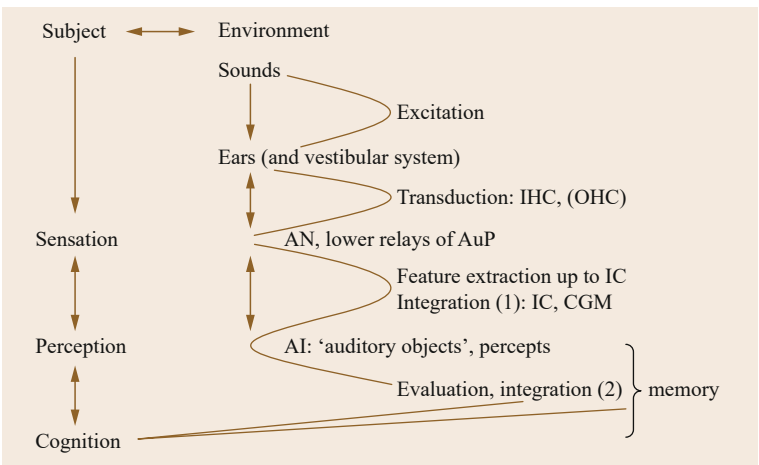


Fig. 30.1 Block diagram: main processing stages in hearing

tion: how does the human body relate to the *soul* (he still used this term), or, more specifically, how does the mind relate to our body? Fechner of course knew that the brain, like other parts of our body, is regulated by physiological mechanisms. However, as can be concluded from personal experience, human beings possess a *mind* capable of forming ideas and concepts, and of abstract thinking in general. Also, in addition to using our senses for whatever *information pickup* from the environment, we can perceive *certain* things as well as imagine things not observed so far (like the legendary unicorn). Such experiences led to the view of the mind being a sphere of its own; even if resting ultimately on brain functions, the *mind* is often regarded a superstructure based on, yet not reducible to, cortical functions [30.29]. This is the dualistic perspective that has been tackled by reductive approaches to the brain/mind gap (Sect. 30.1.1) leading to monistic theories [30.15]. Avoiding overt decision in favor of either *dualism* or *monism*, Fechner simply claimed a lawful relationship between physical stimuli, certain psychophysical functions (as embedded in the human body including the brain) and the sensations resulting from the stimuli. Hence, there is a tripartite relation of which stimulus and sensation are the endpoints that allow empirical investigation while the psychophysical operations (i. e., physiological and neurophysiological processes) activated by stimuli such as musical sound were not accessible to *direct inspection* (given the methodology and tools available to Fechner).

Fechner distinguished between outer and inner psychophysics whereby the former connects physical stimuli and sensations, and the latter would be concerned with specifics of the neural structures and activity underlying sensations and perceptions (as well as with a range of issues in regard to the fundamental relation of body and soul or body and mind Fechner was seeking to investigate). As part of inner psychophysics, *Fechner* [30.65, Vol. II] treats certain psychic conditions (e.g., attention, vigilance) and also memory phenomena such as after-images. Fechner's ideas on inner and outer psychophysics have been discussed in great detail, and have been used also for modern reformulations of psychophysics [30.66, 67].

The symbols ∩ and ◊ both denote measurement, however, of two different kinds. While physical stimuli in general can be measured in an objective way according to certain dimensions, parameters, and units that are well defined and accepted as such by the scientific community (as is, for example, the MKSA system), sensations are based on neural activity that, due to the complex distributed structure of the CNS, is difficult to measure. Accessible, though, are reactions of individual subjects that reflect their experiences relative to certain stimuli. What is measured, therefore, are stimulus parameters (∩) representing the physical *input* into a system that performs physiological and neurophysiological processing. The system output under certain conditions is accessible to measurement, or at least certain parameters can be measured and taken as indices of system performance; for example, the effect of very high SPL acting on dancers in techno discotheques can increase secretion of adrenaline as well as of acetylcholine that can be measured from blood samples. The extent to which sensations are measurable depends on modality, on the type of stimulus as well as on experimental conditions (as in measurement of pain, loudness, etc.). Quite often, the reaction of subjects is registered, which is taken to represent the sensation. The reaction itself can be measurable in a physical unit (e.g., pulse frequency (Hz) or skin resistance (Ohm) as in polygraphic recordings while a subject is listening to different types of music; also, reaction time (s) as an indicator of perceptual and cognitive processing complexity, etc.). Also, the sensation a subject might have, can lead to behavioral responses (e.g., facial expressions) recorded in an appropriate way. Further, reactions can be measured by asking subjects to mark the *intensity* or *strength* of a sensation using a scale offered by the experimenter. Such scales typically are ordinary, offering a number of alternatives that can be regarded as *degrees* of a sensation as experienced by a subject. Large samples of behavioral response data are gathered from such types of measurement (◊ in Fig. 30.2) that rest on subjective estimates and evaluations. Since, however, behavioral response data correspond in some way to the physical input it is possible to express the magnitude of sensation in terms of the input variables. For example, empirical findings demonstrate that, in techno discotheques, visitors scale subjective loudness into categories such as *moderate*, *loud*, *very loud* in correspondence with the increase of SPL measured in dB(A) or, more appropriately, dB(C). Hence, the scaling of relative loudness (Chap. 33) in this case can be tied to measurement of the *input* in a physical unit (indicated by an arrow from sensation to stimulus in Fig. 30.2). Expressing the *output* (magnitude of reactions of subjects) in terms of the

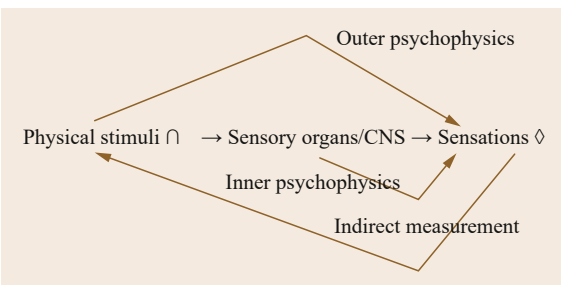


Fig. 30.2 *Outer* and *inner* psychophysics (Fechner)

input is one aspect of so-called *indirect measurement* of sensation.

Fechner foresaw difficulties inherent in the measurement of sensations [30.65, Vol. I, Chap. VII]. The main reason for concentrating on outer psychophysics was that stimuli and sensations are both accessible to empirical study. Fechner would have wanted to investigate neural brain states as well; however, there was no objective methodology available then (about 65 yr before human electroencephalogram (EEG) recordings, and more than a hundred years before positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) techniques were introduced). In a general form, outer psychophysics postulates that there is a lawful relation between a parameter S of a physical stimulus and a sensation E effected by S (via the receptor organs and internal processing) that can be expressed as a function $E = f(S)$. One parameter that had been studied even before Fechner was intensity (for the sense of touch, by *E.H. Weber* [30.64]). Applying $E = f(S)$, one could hypothesize that the sensation of loudness increases or decreases in proportion to the intensity of sound, which is defined as $I = p_{\text{eff}}\, v_{\text{eff}} = \rho\, c\, v^2$ ($p =$ sound pressure; $\rho =$ density of a certain medium, e.g., air; $c =$ sound velocity in this medium; $v =$ particle velocity). Hence, sensation of loudness can be expected to increase with rising sound pressure and increasing particle velocity. At least within the realms of classical physics, a common view is that exact measurement is possible for physical stimuli with parameters that can be varied along certain dimensions conceived as continua (so that, in theory, any value can be represented on a dimension between its minimum and maximum respectively). The notion of a *dimension* implies, in a strict sense, that such constructs must be quasicontinuous (consisting of differentiable elements), interval or ratio scaled, equipped with a well-defined or natural zero, and, as entities representing some acting parameter (e.g., a force of some kind), independent of each other so as to form orthogonal vectors in an n-dimensional metric space.

It has been argued that such conditions do not apply to sensations, since these would have to be treated as intensive and not as extensive magnitudes. Such a distinction points to *Kant*'s epistemology [30.68, pp. 202–205] where one of the axioms states that conceiving of phenomena in time and space implies extensive magnitudes (for instance, geometrical figures are conceived of in spatial dimensions). The notion of an extensive magnitude entails quantity (which in turn leads to operations of addition, multiplication, etc.). Sensations and perceptions resulting from real world phenomena, according to *Kant* [30.68, pp. 205–214], bear to intensive magnitudes, that is, they may represent different grades of the same basic quality. The distinction between quantifiable extensive magnitudes and intensive magnitudes that allow for a gradation of qualities has played a role also in the early debates concerning *Fechner*'s psychophysics (for essential arguments, see [30.66, 69–71]). The view according to which sensations can be graded (as subjects experience different degrees of phenomenal qualities) yet might not be accessible to true measurement as this would presuppose a sensory continuum plus a well-defined *unit* of measurement has had some consequences. One was to regard sensations basically subjected to *categorical* judgments and scalable on ordinal yet not on interval or ratio scale level (Sect. 30.1.4). Another was to call into question applicability of the concept of quantification to psychic phenomena, as well as comprehensive debates on what *measurement* entails in (a) the sciences and (b) in psychology and related fields. Such debates have been complex because they involve both fundamental principles as well as historical processes in scientific research.

Without going into details of historic background and theoretical argument, it is generally agreed that mathematics, including geometry, number theory, and even analysis, has played a significant role already in antiquity, and in particular in the so-called Pythagorean tradition of science [30.72]. In fact, Greek mathematical thought already included concepts such as the continuum and the infinitesimally small [30.73] that were elaborated on ever since, up to modern approaches (such as advanced by Dedekind, Poincaré, Hölder, etc.). Also, concepts of quantity, quantification, and measurement were discussed in Greek philosophy. *Aristotle* [30.74] defines quantity as a set of discrete elements that can be counted, while he regards magnitude as a continuous entity that can be segmented into parts. The elements that make up a countable quantum are numbers while magnitudes are realized in the length of lines, the width of a plane, and the depth of a body. Hence magnitudes rest fundamentally in geometric constructs, which are extensive, and comprise of continuous elements such as lines and circles. Therefore, length has been the paradigm of measurement from antiquity to the present (even in publications on theory and methodology of psychological measurement [30.75], and has served as an illustrative model to study fundamental principles (e.g., continuity, additivity) of measurement. In this context, one should remember developments in analytic geometry as well as efforts to provide a complete axiomatic-deductive foundation for geometry that have led to the treatment of geometry in terms of arithmetic, a development greeted as necessary to achieve utmost logical coherence [30.76, Chap. 2,3]. As one of the consequences, the concepts of *magnitude* and *measure* were subjected

to axiomatic justification. Though concepts such as *quantity*, *magnitude* and *proportion* of magnitudes are still related to geometric entities such as point and line, there is a clear shift to arithmetic reasoning whereby, for instance, the concept of *proportion* (known from Pythagorean tradition and systematically treated in Euclid's *Elementa*) is given a new treatment based on real numbers [30.77]). In effect, each magnitude can be represented by a (positive) real number (implying that a continuum of points forming a line can be represented by a continuum of real numbers). Also, distances between any pairs of points $(a, b, c, d, \ldots)$ can be expressed as *line numbers* (Hölder uses the German term *Streckenzahlen*) and can be given numerical treatment as real numbers.

Though Helmholtz played a significant role in modern approaches to analytical geometry (including non-Euclidean spaces and theory of manifold), he maintained that even complex geometrical constructs can be related to perception, and that geometry as well as measurement rests ultimately on our experience of forms and bodies in the environment (where we compare their sizes, assess their distance and location, observe motions of such bodies, etc. [30.78–80]). Hence, Helmholtz advocated an empiricist foundation of measurement, taking into account our ability of focused observation as well as to compare objects as they occur in the environment. Such comparisons may be directed to sizes, distances, weights (but may also include degrees of brightness and differences in pitch), and may lead to tentative estimates while precise measurement for Helmholtz requires *units* modeled close to natural objects or processes. Once such *units* are at hand, measurement of *magnitudes* is secured by operations such as counting, addition, and multiplication. Though Helmholtz viewed measurement in regard to physics in the first place, measuring was also defined in psychology as *comparing a magnitude with a fixed unit, and then counting how often the unit is contained in the magnitude* [30.81, p. 72].

In the 20th century, measurement theory has become a large field of its own, deeply rooted in mathematical and probabilistic foundations [30.82–84]. In one approach labeled axiomatic and representational, the basic idea is to *map* qualitative relations found between n empirical objects into a set of numerical relations (expressed, in most cases, as positive real numbers). Qualitative relations between objects may result from different *degrees* to which certain observable attributes exist in these objects (which implies an ontological statement), or which seem to exist according to our perceptions (including evaluation Sect. 30.1.2). The mapping of qualitative differences observed between objects into quantitative indices is homomorphic insofar as the resulting numerical structure represents the underlying empirical structure. The structure is defined in regard to identical empirical and numerical relations; it has been proposed [30.85, pp. 4–5, fn. 1] to use the term *isomorphic* for such cases. Under certain conditions (for which logical and mathematical definitions are given [30.86, Chap. 21], [30.75]) the mapping is regarded as isomorphic if each real number represents one (and only one) magnitude on a continuum.

In regard to empirical research such as is carried out in psychophysics, measurement has been given a rather broad definition as *the assignment of numerals to objects according to rules* [30.87, p. 677]. This definition, which Stevens himself saw as a generalized version meeting specific demands of the behavioral sciences, and of psychology in particular [30.88, Chap. 2] has been reiterated in quite many psychology textbooks and other publications, and thus has become kind of a standard. However, Stevens' definition as well as the concepts of measurement he has advocated in a range of publications [30.89–94], have been severely criticized on grounds both theoretical and empirical. Among the objections ([30.71, 95–97] and Sect. 30.1.4, *Loudness Scaling and Scaling of Pitch: Two Illustrative Examples*) that have been raised are these:

1. Stevens' approach where he, in numerous experiments directed to finding scales for attributes of sensation and perception, had subjects *assign numerals to objects according to rules*, is not compatible with established procedures of scientific measurement, which typically aim to find numerical relations between properties of objects that exist in this world (or, in regard to measurement in astronomy, somewhere *out there*). Measurement of such objects (or, rather, of magnitudes representing certain properties and attributes) thus often has led to the discovery of laws of nature; the relations found by way of measurement (most of all, in *classical* physics) are independent of the individual observer, and are largely invariant within a range of conditions. Relations can be expressed as ratios and proportions, as was done prominently in Greek mathematics and music theory, both of which seem to stem from Babylonian geometric traditions of rational division of straight lines as well as string lengths. The Pythagorean theory of proportions precedes Euclid by far [30.72, Part 2]; the legendary Pythagoras is said to have discovered that two strings of same length and tension yield harmonic intervals if divided in certain proportions based on small integer ratios such as $2:1, 3:2, 4:3, \ldots, m:n$ (where $m = n+1, n = 1, 2, 3, \ldots$). To be sure, *this was the first known example of a law of nature ruled by the arith-*

metic of integers ... [30.98, p. 138]. The theory of proportions was given systematic treatment in Euclid's *Elementa* (in particular, book VII and VIII), and again in *Hölder*'s [30.77] axiomatic approach to continuous magnitudes.

2. In regard to measurement, certain principles for defining a metric have been established; the principle of additivity (whereby $f(x \circ y) = f(x) + f(y)$; the sign $\circ$, known as *Hempel's operator*, indicates a well-defined operation to combine two physical objects into a new one) is perhaps best known. Also, the unit of measurement shall be defined by some standard object, which in turn should have some meaningful relation to natural objects or configurations as was the case with the meter (the *mètre des archives* of 1793), which represents – in good approximation – $1/10\,000\,000$ of the distance at sea level from the North Pole to the Equator. Also, the length of a pendulum that had a half cycle of exactly 1 s was very close to 1 m. Several of the units that Stevens employed for *subjective magnitude estimation* seem to be quite arbitrarily chosen; this holds true also with respect to anchor and zero points (we will address this issue below in Sect. 30.1.4, *Loudness Scaling and Scaling of Pitch: Two Illustrative Examples* in regard to the so-called *mel* scale of *subjective pitch*).
3. The methods of measurement Stevens was using in experiments with subjects in order to develop ratio scales and exponents for power law formulations (see below) have been challenged. Among the critical issues are some of the instructions given to subjects as well as the order and range of stimuli presented, which are regarded as factors having introduced a bias. Also, averaging applied to datasets from various samples (many of which differed in size as well as in regard to conditions) is viewed with some concern.

Stevens defended the work he and his co-workers had conducted at Harvard and maintained that measurement cannot, and should not, be confined to such variables for which the principle of additivity holds, for, in psychology and the social sciences, *there is a need to scale new kinds of variables and to measure the previously unmeasured* [30.88, p. 46]. One has to bear in mind that, in modern science from the Renaissance to the present, methodology directed to measurement, scaling, and to quantification of variables and magnitudes has been essential. This holds true also for musical acoustics and music theory [30.98, 99]. The quest to measure sensations, and possibly even complex mental properties (such as *intelligence* or *musicality*), therefore, fits into a general trend of modern science, which is looking for causality, lawful functional relations, and for exactitude (implying operations of *counting and measuring*, as was underpinned by [30.79, 80]). It might be added that, in psychology, there have been proposals for extended and alternative concepts of measurement [30.100–103]; some concepts either avoid the principle of additivity, or give it a broader meaning as was done by *Stevens* [30.89] when proposing a new scale of *subjective loudness* (Sect. 30.1.4, *Loudness Scaling and Scaling of Pitch: Two Illustrative Examples*). Also, attempts have been made at demonstrating that measurement for phenomena such as (musical) pitch and loudness in a form that adheres to the principle of additivity *by procedures essentially similar to those for length and weights* [30.95, p. 277] is feasible. These procedures were discussed in detail, however, with a view on proving the principle rather than providing a model that has practical relevance.

In his works on measurement, Stevens distinguished between two types of scales (prothetic and metathetic) and four scale levels (nominal, ordinal, interval, ratio). *Prothetic* scales are regarded to represent *intensity continua*. According to *Stevens* [30.94, pp. 371–372], intensity of sensations such as brightness or loudness can grow since there is *a physiological process in which excitation is added to excitation already there*. On the other hand, a sensation of a constantly rising pitch (as brought about by a sine sweep where the frequency is modulated so as to pass through a band (e.g., octave) per time unit) does not result from *added excitation but because new excitation has been exchanged for excitation that has been removed* [30.94, p. 372]. Hence, the assumption is that prothetic continua depend on physiological processes with additive patterns of excitation while metathetic continua would require a process of substitution. As paradigms for the two types of continua, Stevens himself has pointed to loudness and pitch respectively (though not exclusively; there are more examples for each type, [30.104]).

The term *continuum* as occurring in psychophysics and theories of perception implies that a random variable X can express very many values x_i; in principle and taking the mathematical concept of *continua* into account, any value x_i within the interval $x_{min} \leftrightarrow x_{max}$ is possible. However, while parameters of the physical stimulus can often be varied on a quasicontinuum (e.g., gain (dB) in an amplifier used for experiments in hearing), sensations are subjected to a discriminal process (Sect. 30.1.4, *Loudness Scaling and Scaling of Pitch: Two Illustrative Examples*, Sects. 31.1 and 31.7, Sect. 31.6.1), which leads to segmentation of continua. Though a continuum is assumed to exist for certain stimulus variables and for sensations corresponding to such stimuli (e.g., frequency of pure tones $\rightarrow$ pitch;

sound intensity → loudness), this assumption needs to be investigated on an empirical level. Consequently, one can segment such continua according to JNDs or other units that have empirical relevance in order to establish what is called a *sensory scale* [30.103, Chap. 3]. For such a scale to be valid and reliable, it would be requested that *it explains a large body of sensory data, collected with a variety of methods* [30.105, p. 131].

Basically, measurement and scaling as developed in the sciences call for certain conditions to be fulfilled: one may often start from (preliminary) observations that demonstrate that a number of objects or events differ with respect to some properties, or at least with respect to their phenomenal appearance. Differences can be viewed to exist in objects of the natural world (ontological and realist interpretation), or to be attributed to objects and events due to their phenomenal appearance to subjects (psychological interpretation made from a realist point of view). If differences among a set of objects are to be measured on a metric so as to capture differences precisely [30.106, pp. 16–20], one would need a gauge with a scale, which in turn calls for a proper unit as well as for a zero. Such a metric would be an interval or a ratio scale, for which additivity applies.

The well-known division of scale types into nominal, ordinal, interval and ratio [30.87, 90] includes the two metric scales as well as two that fall short of a strict concept of measurement. Scale types can be distinguished in regard to permissible transformations (such that do not affect the relational structure inherent in data, which is kept invariant) as well as other operations [30.101, Chap. 20], [30.86, 103]. Nominal scales often result from classification where a large number of objects is sorted into classes according to certain distinctive features (as is the case, for example, with musical instruments that can be classified with respect to the sound-producing element into idiophones, membranophones, aerophones, chordophones, electrophones). Within each class, all objects are regarded as equivalent with respect to the criterion defining the class. One can assign numbers to such classes; however, these numbers do not indicate a quantity or a magnitude, and can be chosen arbitrarily. With ordinal scales, a monotonous relation ($<$ or $>$) between elements is introduced as a number of objects or subjects are sorted so as to yield a rank order (like $A > B > C > D > E$). In case several objects have the same rank, there are ties that can be of relevance in statistical procedures (for statistical methods, see textbooks such as [30.107]). Ordinal scales, which are often used with variables representing subjective attitudes and performance, have been also employed in experiments directed to sensation and perception where ordinal scales can result from estimates of phenomenal differences that are experienced. For example, observers can be asked to judge the loudness of music examples they hear according to categories (which might have a range, with respect to loudness, from *very soft* to *extremely loud*). Though such estimates can be fairly consistent within samples of subjects, leading to a rank order that can represent *degrees* of sensational and perceptual qualities, an ordinal scale derived from rank order expresses relative differences not yet quantifiable in a well-defined unit. Such units as well as interval constancy are expected for an interval scale where the zero can be chosen at the discretion of the experimenter (who will, however, seek to anchor a scale at a zero that does have empirical meaning, e.g., a threshold of sensation). For a ratio scale, an *absolute zero* is requested (this is a theoretic postulate that can be difficult to attain in practice).

In regard to scale level, it has been argued that we cannot truly measure, for example, two different sensations of loudness resulting from two audio stimuli presented in succession since we have no *natural* unit of loudness at hand. Rather, we employ some comparison and may express that two sensations (A, B) are either *equal* ($A = B$, including $A \approx B$) or *unequal* (the latter implying $A > B$, or $A < B$). In case two sensations are found unequal, we may try to estimate their relative difference ($A \to B$ or $B \to A$), or may even guess what the ratio between A and B might be $(A : B)$. Subjective estimates as to *how much louder* A appears relative to B can be achieved on a scale comprising a number of categories ordered in accordance with a monotonous function and satisfying the criterion of transitivity. Judgments then involve assigning A and B to a certain category on an ordinal scale. However, true measurement of sensations, as already expressed by *Fechner* [30.65, T. I, Chap. VII], would require a unit derived from empirical investigation relevant to fundamentals of psychophysics. Measurement in the psychic realm to Fechner meant *summing so many times the same*, whereby *the same* in this respect means the same basic unit. The relevant unit for sensation was the just noticeable difference (JND) found in discrimination tasks.

To be registered as unequal by subjects, two sensations must differ by a certain magnitude, which depends on (a) the stimulus parameters and (b) the modality. *Weber* [30.64] found ratios for two stimulus magnitudes that led to just discriminable sensations; in these tasks, in which he acted as experimenter and as subject, sensation of course was registered in a situation of focused attention. Weber observed that the ratio of two stimuli just discriminable as to the intensities they produce in sensation is constant (or almost so) over a wide range

of stimulus settings. Taking one stimulus, of magnitude S, as a reference and the other that leads to a just noticeable difference in sensation as ΔS, the magnitude of ΔS in proportion to S denotes a differential threshold. The quotient $\Delta S/S = k$ ($k =$ constant) became known as the *Weber fraction* (and also as *Weber's law*). In case the quotient $\Delta S/S$ for a certain modality and type of sensation is constant within its range, there is but one value for k per modality (e.g., hearing) and type of sensation (e.g., pitch, loudness) relative to physical stimuli such as sine tones or other types of sound. For example, from empirical findings according to which most subjects can detect a shift in frequency for a sine tone from 1000−1003 Hz (at medium SPL), it has been concluded that for pitch of pure tones $k = 0.003$ (or $1/333$). Values for k of course differ from one modality to the next. One may start the process of determining JNDs from an absolute lower threshold, for example the minimum stimulus intensity applied to evoke a sensation of sound, which can be taken as a reference, S_0.

The *Weber fraction* can be viewed in terms of differentials since a small stimulus increment ΔS is needed to bring about a small change in sensation, ΔE, the just noticeable difference (JND) corresponding to a change of stimulus magnitude from S to ΔS. Hence, $\Delta S/S = k\,\Delta E$ and $\Delta E = 1/k\,\Delta S/S$. Fechner saw Weber's law as a basis for a differential equation [30.100, p. 37 ff.]. As a differential equation, Weber's law can be expressed thus

$$E = k\frac{\mathrm{d}S}{S}\,.$$

Fechner took

$$\mathrm{d}E = \frac{k\mathrm{d}S}{S}$$

as his *fundamental formula*.

From the basic $\Delta S/S =$ constant, one can expand its interpretation to conclude that *equal relative increases of the stimulus correspond to equal increases in sensation* [30.65, T. I, p. 64]. Hence, two stimuli that form a certain ratio with respect to a physical parameter (such as SPL) will bring about a *difference* in sensation corresponding to this ratio. As a particularly striking example, Fechner pointed to musical intervals, saying it is a simple and *notorious* experience to find, by way of hearing, that equal ratios of frequencies (of two pure tones) result in an identical *Tondifferenz* in various octaves [30.65, T. I, p. 181]. The term *Tondifferenz* as used by Fechner is a bit awkward as it includes arbitrary ratios while, at least for listeners familiar with music, it is the small integer ratio constituting a certain musical interval that leads to such *notorious* experiences; perceiving two pure tones as forming an interval of a perfect fifth always calls for a frequency ratio of $3/2$, no matter if this is realized in a rather low frequency region (say, 120 : 80 Hz) or at a high one (say, 4500 : 3000 Hz). Of course, thereby the *absolute* increase in stimulus frequency grows towards higher frequency regions. *Fechner* [30.65, T. I, Chap. IX] concluded that this peculiar sensation of musical intervals based on stimulus frequencies forming small integer ratios is unparalleled in other sense modalities. One can further argue that musical intervals are perceived as entities (or even as *Gestalten*) distinct in quality that do not fall under discrimination tasks and just noticeable differences (with the possible exception where subjects discriminate intonation deviations leading to perceivable differences in the quality of intervals (Sects. 31.6 and 31.7)).

Taking the Weber fraction as a differential, integration yields $E = k\ln S + C$ (where C is the integration constant and ln indicates the natural logarithm). Fechner's concept included the absolute minimum of stimulus intensity needed to evoke a faint sensation. For a very small stimulus S_0, sensation can be expected to vanish ($E = 0$) when $S_0 < b$, and to become just noticeable when $S_0 > b$. For $E = 0$ the equation is $0 = k\ln b + C$ and, hence, $C = -k\ln S_0$, which gives the threshold of sensation, S_0 (Fechner's *Schwelle*). *Fechner*'s [30.65, T. II, Chap. XVII] *Maßformel* then is

$$E = k\log\frac{S}{S_0}\,,$$

and the integrated formula is $E = k\ln S + C$. Fechner argued that the magnitude of sensation is not simply a function of the stimulus magnitude as such, yet is dependent on this magnitude in relation to the threshold. This is why the *Maßformel* contains the quotient S/S_0. It is only under certain conditions that the functional relation between stimulus and sensation simply obeys $E = \log b$ where b denotes the threshold, indicating that stimulus magnitude for every sensation must be measured from the lower threshold point. The concept is that the difference between an actual stimulus magnitude b and the threshold b can be measured like $E = k(\log \mathrm{b} - \log b)$, which equals

$$E = k\log\frac{S}{S_0}\,.$$

Fechner cautiously warned that strict proportionality of ΔE to ΔS holds only for very small increments. In order to turn the *Weber fraction* into a differential equation, and to use Fechner's *fundamental formula*, the increments would have to be infinitesimal and, hence, much smaller than the JNDs actually found in em-

pirical study of sensation. Fechner's approach led to the consideration of the difference between b and *b* as corresponding to a certain number of JNDs [30.65, T. I, Chap. VII]. In this way, measurement of sensation comes down to summing up or *counting* a number of JNDs to bridge the gap between b and *b*. As a result, the magnitude of sensation *E* corresponds to *n* times the JND counted from *b* to b ([30.103, pp. 106–108], who offers an algorithm for calculation of the number of sensation increments S_N/S_0 that fall between *b* and b). Judging the difference between b and *b* as a sum of JNDs (even if conceived as a series of concatenated JNDs) has been viewed as *indirect measurement* [30.108] as opposed to *direct measurement* where the difference b $\leftrightarrow$ *b* would be judged as one magnitude. (It has been mentioned above that, in Fechner's approach, *indirect measurement* refers also to another aspect, namely, measurement of sensation magnitude through measurement of the stimulus magnitude.)

Fechner originally (in 1851) had postulated that *psychic intensity* is the logarithm of the corresponding physical intensity; he subsequently modified his concept in regard to sensation. The general idea is that increases in the stimulus parameter must form a geometric series to bring about an arithmetic series of corresponding increases in sensation. Scaling of an attribute of sensation relative to a physical parameter typically starts from the lower threshold, which is *not* a discontinuity in the magnitude of sensation but the level of excitation where an observer notices a (sometimes very faint) sensation. Since the threshold is reached in small increments of stimulus intensity, starting from zero or a subliminal level, there have been considerations that one may regard the threshold itself as a JND (that marks the difference between *no sensation* and *minimal sensation*). For the threshold of sensation, the physical stimulus parameter must have a certain magnitude. For example, a sine tone of 200 Hz becomes audible for many persons at a SPL of $\approx$ 15 dB [30.109, p. 32]. If the threshold would qualify as a first JND, one would expect the next JND above 15 dB to be of similar magnitude. Fechner apparently did not make this assumption but argued instead that the JNDs above the threshold can be taken as equal provided both the JND of sensation and the stimulus increment are very small.

Fechnerian scaling of a sensory attribute yields a logarithmic function for which the graph would have a shape similar to that shown in Fig. 30.3.

The Weber fraction has been generally accepted as empirically proven within certain limits; observations near a threshold of intensity sensation may indicate a *near miss* to the *Weber law* that must be slightly modified accordingly [30.110, pp. 121 ff.]. Fechner's assumption that a sensation magnitude can be measured as a sum of JNDs has been rejected by many. The main argument (also in regard to the *Weber fraction*) is that the JNDs are not constant in size over the range of measurement. Some of his critics [30.111, Chap. 4], [30.112] denied measurability of sensations as magnitudes at all and instead held that sensations must be viewed in a qualitative rather than in a quantitative way.

As an alternative to Fechner's approach, a concept in psychophysics advanced in particular by Stanley S. Stevens et al. at Harvard in the decades from 1940–1970 has gained prominence. Relating stimulus magnitudes to subjective magnitudes, Stevens claimed *that equal stimulus ratios produce equal subjective ratios* [30.92, pp. 153, 172]. In short, the concept is *that the sensation* Ψ *is proportional to the stimulus* S *raised to a power* n [30.92, p. 162]. Hence, Stevens' power law is $\Psi = kS^n$. Leaving aside constants depending on the choice of units, the formula simply is $\Psi = S^n$. What needs to be determined, then, is the exponent *n* relevant for a given sensory modality as well as certain classes of stimuli. Different from Fechner's approach, which was based on *indirect* measurement (Fig. 30.2), Stevens proposed a methodology of *direct* measurement that includes four types of experiments all designed to achieve ratio scaling of sensory attributes [30.92]:

1. Ratio estimation (including direct judgment of ratios)
2. Ratio production (by either fractionation or multiplication)
3. Magnitude estimation (with or without a modulus prescribed by the experimenter)
4. Magnitude production.

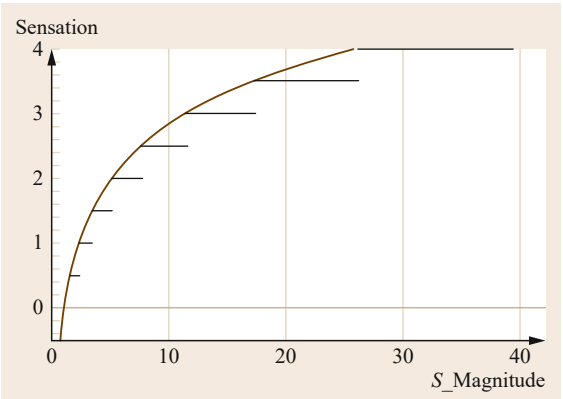


Fig. 30.3 Graph illustrating *Fechner's law* (increases in stimulus intensity corresponding to subsequent JNDs indicated by *horizontal lines*)

This scheme was somewhat expanded later on to include cross-modality matching and cross-modal ratio matching [30.93]. Stevens' approach to ratio scales and to establishing the power law deserves special notice because it led to two scales proposed as being of relevance to the sensation and perception of sound, namely, the so-called *mel scale* of pitch [30.113] and the *sone scale* of loudness of which several versions were published [30.104, 114]. We will examine both scales in detail below after briefly looking into some of the experimental methods.

The idea to have subjects insert a stimulus R_m between pairs of stimuli R_1 and R_2 so that R_m is right in the middle, and the distance $R_1 \leftrightarrow R_m$ is of the same size as $R_m \leftrightarrow R_2$, has played a role already in Wundt's lab at Leipzig; *Merkel* [30.115], working with visual stimuli, reported that R_m was significantly closer to the arithmetic mean than to the geometric mean (with the latter to be expected due to Weber's law and Fechner's law). In 1890, Wundt and Stumpf engaged each other in a famous controversy concerning sensation of *tone distances* [30.116, Chap. III.1], and Stumpf later had his coworkers Abraham and von Hornbostel perform experiments on *tone distances* (where the magnitude of a tone distance is defined as the difference of two frequencies [Hz]: $f_2 - f_1$). The results in particular for a large distance ($74 \leftrightarrow 4115$ Hz) showed that a tone intended to bisect the distance had a frequency close to the geometric mean (which is 551.8 Hz) and far from the arithmetic mean. Moreover, in several experiments directed at fractionation of a *tone distance* into equal appearing smaller distances, estimates were not independent of musical intervals (which apparently served as categories and perceptual *yardsticks*; [30.117]). Observations reported by *Pratt* [30.118, 119], on a small sample of subjects, also indicate that frequency distance estimates interfere with musical intervals (in particular for tone distances smaller than an octave), and that subjects with some musical background tend to bisect distances by a frequency corresponding to the geometric mean rather than the arithmetic mean.

Stevens started his experiments on sensation of loudness and pitch (as well as on other sensory attributes) in the 1930s. In a number of ratio estimation experiments, subjects had to judge the ratio of pairs of stimuli presented to them in succession. In ratio production experiments, subjects typically were asked *to adjust a stimulus to produce a certain ratio to a standard*. With respect to fractionation, subjects might be asked to adjust a tone or noise so that it appeared to them as *half as loud* as a standard. Conversely, multiplication might be done by adjusting a pure tone from a sine generator so as to produce a pitch *four times as high* as that of a standard. The method of magnitude estimation did not employ ratios yet

> *requires the subject to assign numbers to a series of stimuli under the instruction to make the numbers proportional to the apparent magnitudes of the sensations produced.* [30.92, p. 165]

Complementary to magnitude estimation, subjects were engaged in magnitude production where *the experimenter might name various magnitudes and ask the subject to adjust stimuli to produce proportionate subjective values* [30.92, p. 165].

From a broad range of experiments addressing several modalities and sensory attributes [30.104], *Stevens* [30.92] came to the conclusion that the relation of stimulus to sensation almost always can be scaled according to a power function of the form $\Psi = S^n$ (with an appropriate exponent for each attribute). Taking the range of modalities and attributes where a power function seems to apply, he made a generalization towards *the* psychophysical law (labeled *power law*, and also *Steven's law* or *Plateau–Stevens law*).

Scaling according to a power function $\Psi = S^n$ would yield a graph of the form shown in Fig. 30.4 where both the y-axis and the x-axis are logarithmic. Displayed is the sensation of loudness (in sone; Sect. 30.1.4, *Loudness Scaling and Scaling of Pitch: Two Illustrative Examples*) as a function of sound level (dB). In a log–log plot, the resulting graph is a straight line (in this case, the loudness function is very much idealized).

The quest for ratio scales favored experimental operations such as fractionation, multiplication, ratio estimation and ratio production. Stevens' attempt at building a *new psychophysics* based on what he called a *schemapiric* approach, combining experimental manipulation and measurement done mostly by *assign-*

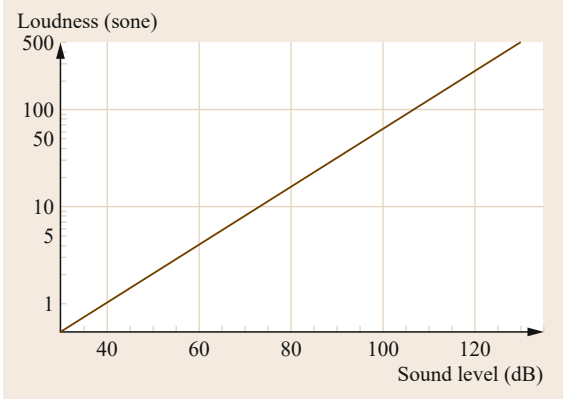


Fig. 30.4 Graph illustrating *Stevens' power law* (for loudness)

ment of numerals to objects [30.88, 91–93], has been met with skeptical reservation. Critics objected that the instructions given to subjects were not neutral and that, typically, *power law data are generated by inducing subjects to assign an approximately geometric distribution of numbers to a geometric series of stimuli* [30.97, pp. 183, 191]. Analysis of the model employing mathematical tools and scrutiny [30.97, Chap. 11] reveals that, given the methodology used by Stevens for experimental design and procedures, the chance to end up with the power law is very high. In this respect, the outcome appears as a consequence of assumptions and modeling rather than of empirical investigation. Even the logical and mathematical concepts of Stevens et al. have been attacked as flawed [30.96]. Other objections raised in the literature were directed against specifics of experimental procedures, data acquisition, and interpretation of data in regard to finding ratio scales as well as to generalize findings to the power law [30.95, 97, 120–123].

Loudness Scaling and Scaling of *Pitch*: Two Illustrative Examples

To illustrate the problems, two scales established by Stevens et al. shall be examined in more detail. One is the sone scale for *subjective loudness*, the other is a scale for *subjective pitch* known as the mel scale. The sone scale was regarded a paradigm for scaling of a prothetic continuum, the mel scale a paradigm for scaling of a metathetic continuum [30.88]. Again, the development in particular of the *sone* scale must be viewed against a background of previous research (see also [30.124]) since there are quite many experiments on loudness sensation antedating Stevens' own contributions to scaling.

To begin with, it was long known that the sensation of loudness relates to the amplitude of vibration of a (periodic or nonperiodic) signal. The source of such a signal may be, for instance a mass of air enclosed in an organ flue pipe in which resonance results in a standing wave. Energy from this pipe emitted into the sound field ([30.125, Chap. 6], [30.126, Chap. 6]) leads to periodic changes of pressure. Though static pressure under normal conditions is quite high (about $1.013\,\mathrm{hPa} = 101.300\,\mathrm{Pa}\ (\mathrm{N/m^2}) = 1.013\times10^6\,\mathrm{dyne/cm^2}$), our ear, which has gained its specific morphology and functions in the evolution of pressure-sensitive organs, is capable of detecting even extremely small pressure changes. In regard to sound, periodic changes of the static pressure are sensed from a threshold for hearing that is as low as 0.00002 Pa.

The intensity of sound, which can be defined as the mean density of the energy flux in a given medium with density ρ (for details, see [30.4, Chap. 4], [30.127]) can be expressed as

$$I = \frac{\rho}{2} c (2\pi f)^2 \hat{A}^2 \,,$$

where $\hat{A}$ is the peak value, c is the wave velocity in air ($\approx 340\,\mathrm{m/s}$ at $20\,^\circ\mathrm{C}$), and ρ denotes density in air, which is $0.00121\,\mathrm{g/cm^3}$. In a more practical view, the intensity of sound equals the product of the root mean square (rms) value of the sound pressure and the rms value of the particle velocity, that is

$$I = p_{\mathrm{eff}}\, v_{\mathrm{eff}} = \rho\, c\, v^2 \,.$$

Since sound pressure is the field quantity that is most decisive for loudness perception, intensity relates to p like $I = p^2/\rho c$, which can be expressed, in the cgs system, in erg ($\mathrm{g\,cm^2/s^2}$).

The characteristic acoustic impedance of air is quite small ($\rho c = 42\,\mathrm{g/(cm^2\,s)}$) so that, approximately, $I \approx p^2$. The reference point for intensity of sound usually is $I_0 = 10^{-12}\,\mathrm{W/m^2}$, which is taken as 0 dB intensity level. The sound pressure correspondingly is $p_0 = 20\times10^{-5}\,\mathrm{Pa} = 0.00002\,\mathrm{Pa}\ (\mathrm{N/m^2})$. Some correspondences of sound pressure (Pa(rms)) and sound intensity to the decibel (dB or db) scale are given in Table 30.1.

Since the range of pressure the ear can deal with spans seven magnitudes, a logarithmic measure is more apt.

It was known from experience that the intensity of a sound emitted from a source decreases with distance if a listener moves away from the source. An early experiment, taking perceived intensity as a function of distance (mm) from the source (a pocket watch placed in a wooden frame with a hole towards the ear of the observer that could be opened and closed for periods

Table 30.1 Decibels, sound pressure, and sound intensity

Decibel	Sound pressure (Pa)	Sound intensity ($\mathrm{W/m^2}$)
0	0.00002	0.000000000001
10	0.000063	0.00000000001
20	0.0002	0.0000000001
30	0.00063	0.000000001
40	0.002	0.00000001
50	0.0063	0.0000001
60	0.02	0.000001
70	0.0632	0.00001
80	0.2	0.0001
90	0.632455	0.001
100	2	0.01
110	6.32455	0.1
120	20	1
130	63.2	10
140	200	100

of time), reported that two intensities were discriminable with certainty if their ratio was 100 : 72, and hardly discriminable (the number of correct and incorrect judgments being almost equal) if the ratio was 100 : 92. Though several hundred observations were made by the two experimenters [30.128], they abstained from fitting a *curve* to their yet preliminary data. Another study (also from Tübingen) underpinned that we tend to perceive stronger stimuli as closer to our sensory organs, that is, intensity is correlated with perceived or imagined (if the sound source is not visible) distance [30.129, p. 8]. This is the core of what much later became the *physical correlate* theory of loudness [30.121, 122]. The relation of intensity to distance depends on the nature of the sound source (point or line [30.130]) and can be given as

$$I \sim \frac{1}{r^2}$$

for point sources and

$$I \sim \frac{1}{r}$$

for line sources.

In the 20th century, among the aspects studied were differences in the sensitivity of the ear with respect to tones of different frequency [30.131]. Progress in experimental setups and procedures was made when electric signal generation, reproduction and measurement equipment became laboratory tools for hearing research. *Fletcher* and *Wegel* [30.132] found a minimum range for the lower threshold at 1–4 kHz where the pressure variation needed to evoke a sensation was 0.001 dyne/cm^2. *Wegel* [30.132, p. 156] published a sketch of the auditory sensation area as defined by curves for the lower and upper limits of hearing respectively. The threshold has its lowest point at about 3 kHz with an rms pressure of ≈ 0.0007 dyne/cm^2 (which is very close to actual threshold values); the peak of the upper limit is located at ≈ 4000 dyne/cm^2 for frequencies of ≈ 0.5–1 kHz. *Fletcher* and *Steinberg* [30.133] investigated how the energy contained in complex sounds might contribute to the overall loudness and suggested this *may be obtained by summing up the contributions throughout the frequency range* [30.133, p. 316]. *Steinberg* [30.134] continued this early approach to loudness summation and gave a formula to calculate the loudness, L, as a function of the energy spectrum of a sound and its level above threshold. To this end, his formula included a weighting of components to make sure that, with increasing intensity of a complex sound, low frequency components contribute increasingly more to the loudness sensation. He used logarithms to define sensation level as

$$S = 20 \log_{10}\left(\frac{P}{P_0}\right),$$

where P is the rms pressure of the sound wave and P_0 the minimum audible pressure for subjects with normal hearing. Steinberg employed the *transmission unit* (TU) customary in telephone technology and defined as $10 \log_{10}(P_1/P_2)$ (where P_1 and P_2 are two powers). *Kingsbury* [30.135], in *behavioral* experiments with 22 observers (11 male, 11 female) established equal loudness contours for pure tones that were compared to a standard at 700 Hz. Since he conducted behavioral experiments, Kingsbury presented statistics of his data and discussed the variability found therein. Still inclined to a JND approach in the realm of Weber and Fechner, Kingsbury employed a *loudness unit* that *expresses the ratio of the smallest perceptible difference in the energy of a tone to its total energy*, and he derived loudness functions for various pure tone frequencies from a function established for a 1 kHz standard. For the JNDs, Kingsbury used data published by *Knudsen* who had suggested that

> *under favorable circumstances, the normal ear can distinguish about 400 gradations from threshold to a painful intensity 10^{12} times as loud.* [30.136]

Knudsen argued that, from his data, a modification of the Weber and Fechner law (as had been proposed by Nutting for light sensation) would be apt for audition. A few years later, *Riesz* [30.137], also from empirical data (12 subjects), concluded that the Weber fraction for differential intensity sensitivity ($\Delta E/E$) of pure tones was, from a certain level above the lower threshold onwards, almost constant ($\Delta E/E = 0.05$–0.15, depending on frequency). In regard to frequency position, the minimum $\Delta E/E$ was found at about 2.5 kHz. Riesz calculated that, at about 1300 Hz, the number of JNDs for loudness would be 370.

Parallel to scholars in the USA (in particular, Harvey Fletcher et al. at Bell labs), Heinrich Barkhausen, a physicist working in electronics and electroacoustics, began to explore loudness sensation and scaling of loudness in the 1920s. One of the issues was how the power of a signal (e.g., a tone from a generator fed into headphones) relates to the loudness sensed by subjects. While the electrical signal could be defined using physical magnitudes (voltage, current, gain), a scale for loudness had yet to be established. Taking a tone of 800 Hz, Barkhausen, with the aid of a *sound meter* (built to detect intensities whose ratios are 1 : 2 : 4 : 8 :

16 : . . .) established a scale from the lower threshold (where the signal fed into earphones was barely audible) to a sensation experienced as painful. The unit he proposed was *Wien* (to honor the physicist Max Wien), where 1 Wien would be the threshold and the upper limit would be at about 16 000 Wien. According to this scale, the difference from 2–4 Wien would be sensed as equally large as that between 2000–4000 Wien. To meet practical demands, *Barkhausen* [30.138, p. 601] offered a logarithmic scale (with the logarithm at base 2) for which the unit would be *phon*. He tabled the two scales, adding some dynamic levels used in music for comparison (Table 30.2).

The unit *phon* is still in use (though with some modifications). In a modern definition (DIN 45630), the loudness level of a sound is n phon if this sound is judged, by observers with normal hearing, as being of equal loudness as that evoked by a pure tone of 1000 Hz, which is sensed as a plane wave incidental normal to the listener, and which has a SPL of n dB (reference sound). If the threshold at 1000 Hz is taken as 0 dB, calculation of loudness can be done according to

$$\Lambda_{1000} = L = 20 \log_{10} \frac{p}{p_0} = 10 \log_{10} \frac{I}{I_0} \text{ [phon]} .$$

From Table 30.1 it is clear that, on the decibel scale, an increase of 10 dB means that the pressure amplitude rises by a factor of about 3.16 while intensity rises by a factor of 10 (as it should to make p^2 equal I). *Kingsbury* [30.135] also addressed the issue of a *direct* comparison of two tones (B, C) in regard to loudness without reference to a standard (A). *Barkhausen* and *Tischner* [30.139] found that comparisons of complex tones (and noise signals) with a standard (pure tone) of varying frequency and intensity were possible within an error margin of, in most trials, ≈ 10%. *Richardson* and *Ross* [30.140] had their observers ($n = 11$) make *numerical mental estimates* of the loudness of tones varied in intensity and frequency relative to a standard (for which the loudness was declared unity). The intensity in many experiments of the time being controlled by the current fed into a telephone receiver, it was fairly easy to determine the ratio of currents used for different intensity levels and to see whether the corresponding loudness levels did yield a similar ratio. *Ham* and *Parkinson* [30.141], stating that, in the previous experiment of Richardson and Ross, on average *the logarithm of the mental estimate ratio of loudness was directly proportional to the logarithm of the current ratio in the phones*, had their (in total, more than 175) observers estimate (1) the percentage (%) to which a stimulus seemed to have decreased in intensity relative to a standard; then, observers were asked (2) to record when a (decreasing) stimulus appeared to evoke 1/2, 1/3, or 1/5 of the loudness of a standard, and, conversely (3), when the loudness of an (increasing) stimulus was two, three and five times as great as that of the standard. From quite many (≈ 4500) observations, *Ham* and *Parkinson* [30.141] concluded that observers can make reliable judgments with respect to the relative loudness of sounds according to a function

$$mL = \mathrm{a} + \mathrm{b}e^{nx} ,$$

where mL denotes the multiple change in loudness, a and b are constants, and x is the difference in the levels of the sound expressed in decibels. Struggling (as many researchers of the time) with *the Weber–Fechner law*, they inferred

> *that the increase (or decrease) of sound energy measured in decibels is a logarithmic function of the increase (or decrease) in loudness.*

Judgments on loudness

> *appear in all cases to be made as a per cent, fraction, or multiple of the original level. In no case did the judgments indicate that the loudness change was an additive or a subtractive function.* [30.141]

The state of affairs reached in loudness measurement and scaling was summed up by *Fletcher* and *Munson* [30.142] with definitions of basic concepts such as intensity level (number of dB above reference intensity) and loudness level (loudness of any sound measured as intensity level of the equally loud reference tone of 1 kHz relative to a threshold). For pure tones, loudness as a sensory magnitude depends on both frequency and intensity. The basic function for loudness,

Table 30.2 Sound and loudness levels in *Wien* and in *phon* (after [30.138])

Music	Wien	Phon
	1	0
	2	1
Pianissimo	4	2
	8	3
Piano	16	4
	32	5
Mezzoforte	64	6
	125	7
Forte	250	8
	500	9
Fortissimo	1000	10
	2000	11
	4000	12
	8000	13
Painful	16 000	14

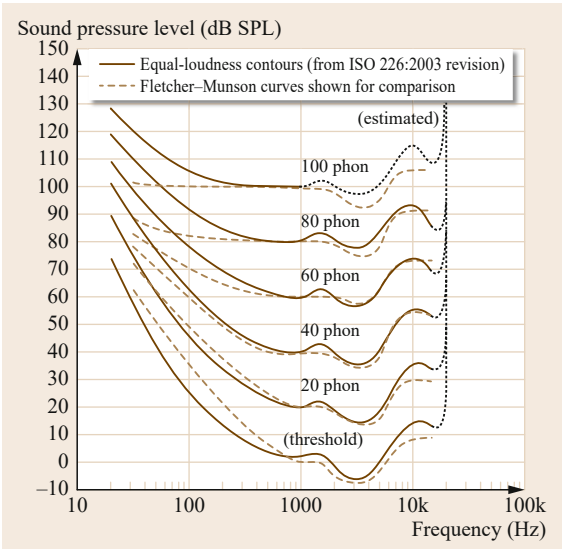


Fig. 30.5 Equal loudness contours (revised 2003 version)

N, thus is $N = G(f, I)$, where f is the frequency in Hz and I is the intensity ($I = p^2/\rho c$).

From comprehensive measurements and analyses of data, *Fletcher* and *Munson* [30.142, pp. 90–91] derived their famous loudness level contours (also known as equal loudness curves or Fletcher–Munson curves), which reflect the fact that the sensitivity of our ears is different for (pure) tones of different frequency (an important factor being the transfer function of the outer ear and middle ear respectively [30.8]). One can combine the equal loudness-curves with the *phon* unit (modern definition) to yield *isophones* (curves of loudness level for different frequencies expressed in phons). The latest version of equal loudness contours/isophones as described in ISO 226 (2003, 2008) is shown in Fig. 30.5.

The loudness of a pure tone presented at arbitrary frequencies and with varying SPL can be judged by comparison to the loudness of a reference tone at 1 kHz; in many experiments subjects had to adjust the stimulus (a sine tone of a certain frequency) in SPL so that the stimulus and the reference seem to evoke equal loudness. The SPL of both the stimulus and the reference can be measured in dB. At 1 kHz, phon scale and dB scale converge for various levels (or nearly so) while they differ significantly at very low and very high frequencies. To appear as *equally loud* as a 1 kHz reference tone of 40 phon (= 40 dB), a pure tone of 50 Hz must be presented with a SPL of ≈ 75 dB. The flattening of contours at higher levels probably is due to a nonlinear compressive loudness function in hearing (Chap. 33). The JND for loudness sensation of pure tones found in a number of experiments has been 1 dB or 1 phon respectively.

The phon scale can be used to measure loudness levels, though measurement in practice is laborious because it involves two parameters per stimulus (energy, frequency) even for simple sounds like sine tones as well as comparison of a stimulus to a reference sound [30.143, Chap. 4]. Therefore, a method of *direct* measurement as offered by Stevens seems attractive for reasons of simplicity. It is of interest to note that, in his article introducing the *sone* as a unit for measuring *subjective loudness*, *Stevens* [30.89] discusses concepts of measurement including the criterion of additivity (Sect. 30.1.4, *Historical and Theoretical Background*). Being a strong proponent of operationism [30.144, pp. 656–658], [30.145], *Stevens* outlines his concept in terms of operations stating that

> *the attribute of sensation to which the number 10 is assigned should appear to be half as great to the experiencing subject as that to which the number 20 is given,*

he concludes that, with a scale derived from such relations,

> *the operation of addition consists of changing the stimulus until the observer gives a particular response which indicates that a given relation of magnitudes has been achieved.* [30.89, pp. 406–407]

The loudness scale thus established is one based on *assigning numerals* to sensations that appear *half as loud* or *twice as loud* as a reference stimulus. A sine tone of 1000 Hz at 40 dB SPL above threshold heard with both ears served as the anchor point of the scale and was defined as 1 sone. A tone of 1 kHz heard *twice as loud* thus had the loudness of 2 sone, a tone heard *half as loud* 0.5 sone, and so on. The plot of the loudness function obtained for the 1 kHz stimulus [30.146, p. 118, Fig. 43], indicated a decrease in intensity of nearly 20 dB for a drop of loudness from 1 sone to 0.1. Stevens later revised his loudness scale on the basis of averaging data he and others had obtained in experiments. He [30.114, p. 820] concluded that, for the 1 kHz sine tone stimulus, the appropriate increase in intensity would be 10 dB for a loudness ratio of 2 : 1. The exponent of the power function then is $\log_{10} 2 = 0.3$ so that $L = kI^{0.3}$ (with $k =$ constant). The intensity I is taken as the density of energy flux and as proportional to p^2, $I \sim p^2$. If I is measured in units relative to the reference level of $10^{-16}\,\mathrm{W/m^2} = 0\,\mathrm{dB}$, then for $40\,\mathrm{dB} = 1$ sone, $k = 0.06$ (the threshold for loudness being $L = 0.06$ sone, which corresponds to 0 dB). The power function for loudness thus is $L = 0.06I^{0.3}$. For

sound levels > 50 dB, Stevens assumed loudness L to grow by a factor of two per 10 dB also for a range of continuous noise signals.

Taking the 10 dB increase or decrease of intensity as appropriate for producing loudness ratios of 2 : 1 and 1 : 0.5, respectively, the sone scale for the 1 kHz tone (where 1 sone = 40 dB = 40 phon) turns out nicely (see also Fig. 30.4) like in Table 30.3. The loudness function corresponding to this scale of 10 dB steps is at odds with a principle of sound radiation, which states that pressure and particle velocity in the sound field decrease with growing distance from a point source like p, $v \approx 1/r$ where r is the radius (however, in the nearfield $v \approx 1/r^2$ applies).

In practice, by doubling the distance from a source (e.g., a loudspeaker radiating constant sound in one direction), a decrease in SPL by −6 dB can be measured with a sound meter. In experiments, Warren et al. found that, to most subjects, estimates of half-loudness of a sound are equivalent to twice the distance from a source, where doubling distance means 1/4 of the intensity or a reduction by 6 dB. Absolute values for half-loudness equivalent to doubling the distance varied from about −6 to −8 dB depending on stimulus and source conditions, however, with a clear tendency towards −6 dB [30.122], [30.123, pp. 111–122]. In his later studies, *Stevens* [30.147] adopted a step of 9 dB as equivalent to a doubling or halving of loudness. As a rule of thumb, for sine tones in a convenient frequency and intensity range (say, 200−1000 Hz, 40−80 dB SPL) most observers will agree that *somewhere between* 5−10 dB *difference in intensity is twice as loud* [30.148, p. 76].

Judgments obtained from subjects on what is *twice as loud* or *half as loud* depend on physical parameters (the frequency, level, and duration of pure tones; spectral energy distribution, level and duration in complex tones; the center, bandwidth, spectral and phase structure, level and duration of noise bands) as well as on experimental conditions (use of loudspeaker(s) or headphones, quality of audio and test equipment, monaural or binaural presentation, order of stimuli in the presentation, etc.). The variability inherent in the data sets used by *Stevens* [30.114] to derive the 10 dB equivalence to doubling or halving loudness is considerable. Analysis of sample sizes and of the distribution of mean and median values shows that the 10 dB median finally chosen for doubling or halving loudness seems problematic [30.149, pp. 46 ff.]. Moreover, there were objections that the experiments Stevens et al. had conducted were biased in regard to instructions, the range of stimuli presented to observers as well as the range available for responses and other effects [30.97, 121–123, 149]. However, *Stevens* [30.150] seems to have foreseen such objections as he tried to assess effects of stimulus range, level, context, etc. Proposing methods such as direct magnitude estimation or cross-modal matching he believed to reduce or even prevent possible biases, Stevens in many publications up to his final book [30.88] maintained the validity of the power law as applying to loudness as well as to other sensory attributes. Other researchers agreed that the power law *appears to have a general validity for prothetic sensory continua* [30.151, p. 82]; see also [30.110, Part IV, Chap. 3].

One important aspect to justify the assertion simply is that increasing the intensity of a stimulus fed into a system such as a sensory organ coupled to a neural pathway suited to processing sound or light should reasonably also increase the *intensity* of the respective sensation. This was the assumption already made by Fechner (Sect. 30.1.4, *Historical and Theoretical Background*). Stevens regarded prothetic continua as such where *excitation is added to already existing excitation*. The term *excitation* is central to modeling of pitch and loudness perception [30.152, 153] where a correspondence between parameters of the sound stimulus (typically, amplitudes and frequencies of spectral components, bandwidth and level of noise bands) and places or areas of excitation on the BM are considered. Assuming a tonotopic organization for the BM taken as a chain of (slightly overlapping) bandpass-type auditory filters (AF), one can expect certain frequencies to produce excitation at certain places on the BM (Sect. 31.4, Fig. 31.4 and Fig. 31.5). Also, a correspondence of sound intensity to the area of excitation is reasonable since, in an elastic mechanical and hydromechanical system such as the cochlea [30.154, 155], the area of excitation on the BM should spread in proportion to the intensity of energy supplied (Chap. 33). It has been suggested that, at least for simple sounds (pure tones and similar stimuli), sensation of loudness *is likely to be closely associated with the physical vibrations of the BM* [30.156, p. 100]. In this respect, a close correspondence between BM vibration amplitude and loudness sensation could be expected. However, it has been argued from observations in animal experiments that, even for pure tones, *no one location in the cochlea can encode the entire auditory dynamic range of sound intensities* [30.157, p. 178]. The *loudness code*, therefore, must cover the activity from an area of neurons. The magnitude of the sensation of course should relate, in a lawful manner, to the intensity of the sound stimulus as well as to the corresponding pattern of excitation. While BM excitation is accessible to measurement, neural processing that leads to the sensation of loudness in humans has become a topic of research just recently [30.158–160]. In the 1940s and 1950s when

Table 30.3 Relation of subjective loudness (sone) to loudness level (phon)

Sone	0.5	1	2	4	8	16	32	64	128	256	512	1024
Phon	30	40	50	60	70	80	90	100	110	120	130	140

Stevens conducted most of his experiments, *inner psychophysics* of loudness was difficult to investigate so that *outer psychophysics* – having observers *assign numerals* to stimuli according to the apparent strength with which sensory attributes were registered – seemed a viable option. In addition, since many observations on loudness sensation could not be satisfactorily explained in terms of Fechner's law, seeking an alternative was certainly justified. The question seems whether the *new psychophysics*, which deems magnitude estimation experiments and attempts at *direct measurement* as well as scaling of sensory attributes in general essential [30.161], has led to significant results in regard to factual evidence and methodology.

One has to remember the *psychophysical law*, which claims that *equal stimulus ratios produce equal subjective ratios* ([30.92, pp. 153, 172]; Sect. 30.1.4, *Historical and Theoretical Background*). Attributes of sensation hence are ratio-scaled so that subjective magnitudes relate to the physical stimulus by a power function [30.70, Chap. 3]. *Luce* [30.162, Theorem 1] has demonstrated that the only function that will preserve physical intensity ratios is of the form $u(x) = \mathrm{a}x^{\beta}$ where x is a typical value of the independent variable, $u(x)$ is the value of the dependent variable, and the exponent $\beta > 0$ is a constant independent of both the unit for u and for x. Consequently, in regard to loudness [30.110, p. 136], the function is $L(I) = AI^{\beta}$, where I denotes intensity of sound and A and β are positive constants.

To facilitate such neat formula, assumptions have to be made. One of the most fundamental assumptions is *that stimuli and responses both form continua*. If this assumption cannot be justified, this *would mean casting out much of psychophysical theory* [30.162, p. 92]. In fact, many textbooks on psychophysics as well as on psychology in general seem to take it for granted *that stimuli and responses both form continua* while empirical data quite often do not support this assumption. Judgments received from subjects in many experiments conform to an ordinal scale (at best), and this is also found in experiments on loudness in particular when stimuli are chosen from real music ([30.163] and Chap. 33). But also in laboratory experiments observations were made that indicate that subjects might use a single (ordinal) scale for either ratio or difference estimates of loudness [30.164]. Moreover, it has long been noted that *scales based on direct judgments of subjective differences are linearly related to those based on subjective ratios* [30.165, p. 203], and that a logarithmic transformation of a (prothetic) magnitude scale often corresponds closely to a category scale [30.100, p. 56 f.]. In regard to the power law, *Stevens* [30.147, p. 586] himself noted that, *at frequencies below* 400 Hz *there is a dramatic change in the exponent of the power function that governs loudness*. *Stevens* [30.88] maintained, however, that notwithstanding necessary corrections, the basic approach (largely based on magnitude estimation and magnitude production tasks leading to ratio scales) was sound, and that the power law applies to sensation of loudness.

To assess matters properly, one has to recall that the search for a loudness function that relates the intensity of a (pure or complex) tone to the sensation of loudness had begun in the 1920s. Attempts at fitting data to functions corresponding to the laws of Weber and Fechner [30.141] met with great difficulties. In addition, *Newman* [30.166] had argued, against Fechner's approach, that the JND is not a unit suited to measure the loudness of tones of different frequency. One might object to his conclusion, though, since the differences in the number of JNDs he calculated (largely from data taken from [30.135]) for six frequencies {80, 200, 400, 1000, 1900, 4000 Hz} to fit in between the lower limen and a loudness level of 70 dB are probably due to the lower limen varying with frequency. In principle, observers might attempt to estimate the difference between two sensory magnitudes (A, B) by imagining the size of the relevant JND, and then *summing up* the number of JNDs corresponding to the difference A $\leftrightarrow$ B; such a process might work if the difference between A and B is rather small (so that the JND appears as a fraction of the difference A $\leftrightarrow$ B, and can also be regarded as of constant size within this range). Though a Fechnerian process of *indirect* sensory measurement cannot be precluded always (depending on the type of stimulus and experimental conditions as well as on the observer's strategy to deal with magnitude estimation), it became clear from many experiments that subjects could estimate the magnitude of two stimuli (in terms of either *difference* or *ratio*) by *direct* comparison.

Several versions of *corrected* loudness functions have been issued previously [30.167, 168]. In regard to the striking variability of data reported in many publications on loudness, it seems questionable that a single loudness function could account well for a very broad range of possible stimuli (from sine tones to full orchestra, and from distant traffic noise to types of sound loaded with impulses such as produced by machinery

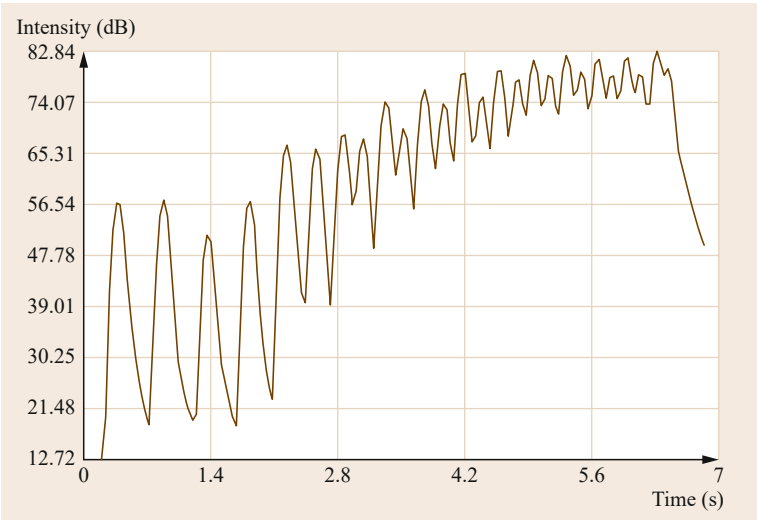


Fig. 30.6 Sequence of strokes on a snare drum; increase of intensity (dB)

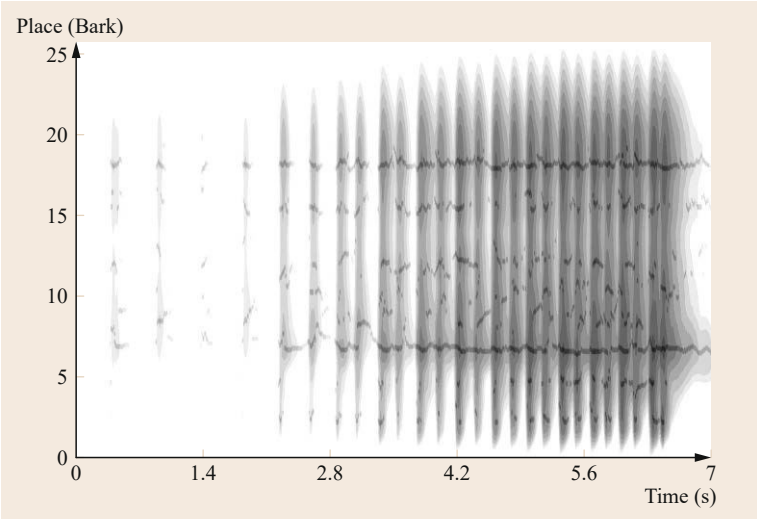


Fig. 30.7 Cochleagram, sequence of drum strokes on a snare drum

or, for that matter, techno music). One aspect that might possibly explain, at least in part, the extreme variability of data on loudness is that sound stimuli comprise more than one physical dimension (and, correspondingly, more than one parameter), which in turn address more than one sensory attribute. To illustrate the case, we take a sequence of strokes applied to a snare drum with different force. Of course, intensity (dB) of the sound (recorded with a single condenser microphone from ≈ 50 cm above the upper drum skin) rises with increasing force as input variable (Fig. 30.6).

Because the force applied increases, more modes of vibration are elicited. From the cochleagram of the same sequence, it becomes evident that spectral energy also increases so as to involve more of the Bark filters at the low end as well as the high end (Fig. 30.7). Thus, there is a *spread of excitation* in that more critical bands (CBs) are activated at higher levels of intensity; consequently, there is significant energy in most of the 25 (Bark) filters representing the BM. According to models of *loudness summation* [30.152, 153, 169], sensation of loudness increases in proportion to the number of CBs activated (Chap. 33).

However, there is more than an increase in loudness. As more modes in the drum are elicited due to an increase in the force applied, spectral energy both at low and at high frequencies increases and the sounds in the sequence are perceived as developing both more *depth* (the relevant sensory attribute being *volume* at low frequencies) and more *edge* (the attribute being *sharpness* due to amounts of energy at high frequencies). Since spectral energy increases both at low and

at high frequencies, the spectral centroid (a weighted mean of spectral energy [30.9]) does not vary much from one stroke to the next (its position stays close to 3–3.2 kHz). However, especially in brass instruments (trumpet, trombone, etc.), increasing physical parameter values (pressure of blowing, which affects also velocity of the particles) at the input also will excite more of the higher modes, which typically leads to a shift of the spectral centroid towards a higher frequency. A higher spectral centroid is sensed as an increase in brightness, and may be experienced also as an increase in sharpness (Sect. 32.2.2, *Semantic Attributes of Timbre*). In sum, by changing a single stimulus parameter more than one sensory attribute may be affected.

Stevens of course knew this problem (even from his early work; [30.170, 171]). *Stevens* [30.88, p. 54] argued that, in experiments, *the subject has the help of instructions designed to direct his attention to a single attribute of a stimulus*. To direct a subject to act this way means he or she must suppress information contained even in seemingly simple stimuli (such as pure tones, which can contain information relevant to sensation of *pitch* and *loudness* as well as to phenomenal attributes such as *brightness*, *volume*, *density*, and *vocality*; [30.116], Chap. 32). Though it is possible to focus attention on a single attribute, there are *integral* dimensions of sensation and perception [30.172, Chap. 5] so that one dimension implies the existence of the other. Experiments have shown that sensation of *pitch* is not totally independent of *loudness* (since a change in SPL can influence the *pitch* of pure tones to some degree; Sect. 31.1, Sect. 31.6.1) and that both *pitch* and *loudness* relate to a dimension of *brightness* [30.173, 174]. Since *brightness* plays a significant role also in vision, there are intermodal correspondences (accountable to neural networks connecting input from different sensory channels). Given the often intricate interrelation between *dimensions* and *attributes*, an approach comprising combinations of variables to study cross-modal information integration and multimodal effects has been proposed, with the aim of establishing a *psychocognitive law* [30.175]. The multimodal and multidimensional paradigm (that is necessary to understand, for example, timbre perception; Chap. 32) may be suited to overcome limitations of *classical* psychophysics where, in quite many experiments, single attributes and dimensions have been investigated. There were certain reasons to proceed this way, among them technical issues as is evident also in many publications on loudness from the years 1920–1950. In general, experimental designs and theoretical frameworks quite often relate to current information and communication technology (as the setup of later and more advanced theories such as signal detection shows [30.176, 177]).

In retrospect, *Stevens* [30.89, 146] took the quest for *summation* and *additivity* of sensory magnitudes from the research on loudness undertaken by *Fletcher* and *Munson* [30.142, Chap. 3], [30.178]. He probably also found inspiring the aspect of a physiological explanation of magnitudes of sensory attributes that had been pursued since the 19th century, and had been adopted by Fletcher early on [30.179]. We will explore this issue, to some degree, by taking a look at the so-called *mel scale* [30.113, 180] devised to study *melodic pitch*. The English-American term *pitch* has many different meanings (only some of them relating to sensation and perception). A well-known definition tells us pitch is *that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high* (ANSI 1973, after [30.181, p. 267]). Traditionally, the physical parameter considered most relevant to determine the sensation of pitch, especially of pure tones, was frequency.

While loudness of pure tones was found to depend on both intensity and frequency, the point of departure for research that led to the *mel scale* apparently was that the pitch of a pure tone of given frequency (Hz) as sensed by observers can be changed to some extent by manipulating intensity. Experiments to this effect were undertaken by Stevens and some other researchers in the 1930s, and have been repeated (with some modifications) thereafter. The effect of SPL on the pitch of pure tones is small at medium frequencies (400–1000 Hz) and medium SPL (60 dB) yet increases at both low and high frequencies with SPL where the pitch shift exceeds 1% [30.152, p. 114]. *Stevens* et al. [30.180, p. 185] argued that, because of the effect intensity can have on sensation of the pitch of pure tones, pitch cannot be simply equated with frequency. Therefore, a new scale for pitch would be apt to account for *the perceiving organism*. This scale, taking a reference loudness of 40 dB, should consist of *numbers whose values are directly proportional to the magnitude of the perceived pitch*. The scale was derived, analogous to loudness scaling, from fractionation involving the *half as ...* task (in this case: *half as high*), and resulted in a pitch function. However, in regard to difference limen (DLs), the experiment yielded that *all DLs for pitch, at a constant loudness level, are of equal subjective magnitude* [30.180, p. 189]. This finding was interpreted in regard to the structure of the BM such that the DLs are associated with *the location of stimulation on the BM*, that is, in terms of a place theory of pitch (comprehensive summaries of pitch theories are given by [30.6, 182–185]). Consequently, the pitch scale as devised by Stevens et al. should closely correspond to BM tonotopicity (from Greek tonos = tone and topos = place). Finally, in regard to musical intervals, *Stevens*

et al. [30.180, p. 190] concluded that, *throughout the useful musical range, intervals increase in subjective magnitude with increasing frequency of the stimulus*. They believed to have thereby proven a hypothesis expressed by *Carl Stumpf* [30.186, p. 250], according to which *the same interval represents, with rising pitch up to about* c^3 [C_6 in US octave designation]*, an increasing distance in sensation*. However, Stumpf considered this but an assumption, and was rather critical of so-called *distance comparisons* of pairs of tones [30.186, pp. 247 ff.] since he had found that observers with a musical background would almost *go crazy* if forced to estimate *tone distances* while at the same trying to suppress perception of musical intervals. As an example, he said that three pure tones having frequencies of 71 : 97 : 111 Hz would be perceived as a somewhat detuned fourth (71 : 97 ≈ 540 cent) and as an enlarged whole tone (97 : 111 ≈ 233 cent), and not as tone distances per se.

The mel scale was issued in a revised version [30.113] based on additional experiments where subjects had to segment large frequency distances (200–6500 Hz, 40–1000 Hz, 3000–12 000 Hz) into *four equal pitch distances*. This is a complicated task since, for example, the interval 200–6500 Hz in fact spans more than five octaves (Table 30.4). The task thus implies dividing an interval of 6027 cent into four equal parts. If this interval is segmented into four *tone distances*, each such distance would be ≈ 1507 cent wide, and each pair of tones would need a frequency ratio of $f_i/f_j = 2.388$. The five tones framing four *equal tone distances* thus would have frequencies close to

200 477.61 1140.56 2723.71 6500 Hz

As can be expected, the data obtained from subjects show a huge amount of variability caused by the experimental design, and even averages are far from the correct values. For Stevens, the mel scale was meant to demonstrate that the sensation of pitch (of pure tones) is a *function* of frequency, yet cannot be simply equated with frequency. As an anchor point, Stevens chose the sensation of pitch corresponding to a sine tone of 1000 Hz presented at 40 dB SPL to have a *melodic pitch* of 1000 mel. The pitch function for the mel scale is shown in Fig. 30.8.

Here, frequency is kept linear to indicate how large an *octave* expressed in mels would be in corresponding frequency. For conversion of mel to frequency, several algorithms have been proposed [30.187, 188], like

$$\text{mel} = 1127.01048 \times \log\left(\frac{f}{700+1}\right);$$

$$\text{also mel} = \left(\frac{1000}{\log(2)}\right)\left(\log\left(\frac{f}{1000+1}\right)\right)$$

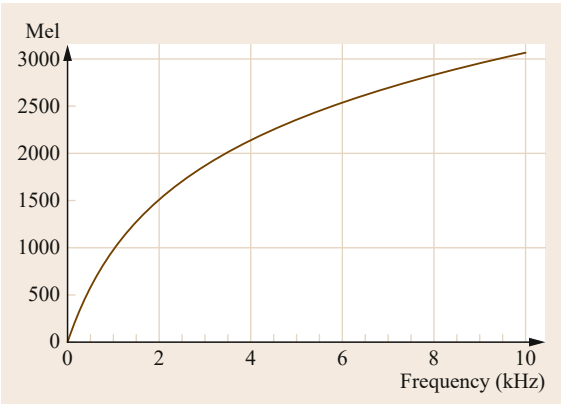


Fig. 30.8 *Mel pitch* as a function of frequency (Hz)

For example, doubling 800–1600 mel corresponds to a *tone distance* of 723.57–2195.07 Hz, which is much more than an octave, namely, in excess of an eleventh (expressed on a logarithmic scale relevant to psychoacoustics and music, the *tone distance* equals 1921.27 cent). Conversely, the octave 800–1600 Hz would yield a *tone distance* expressed in mels like 853.94–1340.674, that is, a ratio of merely 1 : 1.57, which is much less than a musical octave (the ratio 1.57 equals 781 cent, implying a musical interval of a minor sixth). With frequency given on a logarithmic scale, a graph for the mel pitch function results as shown in Fig. 30.9.

After using fractionation and equisection, Stevens later extended his experiments to magnitude estimation of pitch distances. He admitted that *many of the observers registered some degree of protest* [30.104, p. 407] when called to make such estimates. Also, experiments involving piano tones as well as pure tones showed the influence of musical intervals, in particular of the octave, on pitch magnitude estimates [30.189].

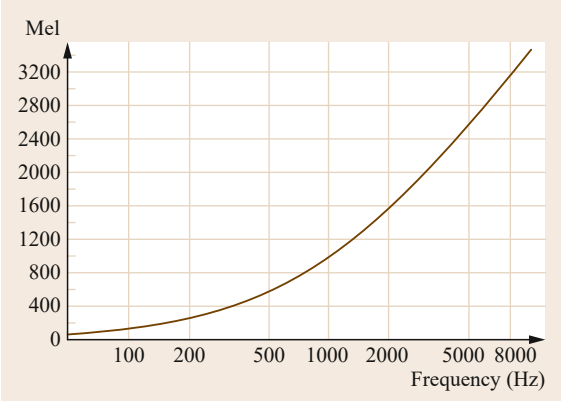


Fig. 30.9 Mel pitch function for a frequency range from 50 Hz to 10 kHz

Table 30.4 Structure of frequency distance 200–6500 Hz

200		400		800		1600		3200		6400		6500	Hz
	1200		1200		1200		1200		1200		27		Cent

Indelible effects of musical interval perception were discussed in terms of *response biases*, but empirical data indicated that the mel pitch function would not remain invariant. However, it was held that the mel scale was *fairly well established* from empirical data, and moreover was supported by *the simple relationship it bears to JNDs, critical bands, and other auditory parameters* [...] *as well as to anatomical and physiological findings* [30.151, p. 47]. Stevens was of the opinion that the mel scale of *subjective pitch* or *apparent pitch* [30.88, p. 165 ff.] had deep anatomical and physiological foundations, in that it mapped frequency to the BM, and was suited to describe basic psychoacoustic concepts such as the CB and the JND in terms of the *mel* unit. According to *Stevens* [30.88, pp. 165–67], there is a correspondence between *equal distances along the BM and equal pitch differences in mels* so that *the pitch function [in mel] provides a linear map of the BM*. Also, he believed the smallest resolvable difference between two pitches to be *about one mel*. In comparison, *Zwislocki* [30.151, p. 50] had calculated 1 JND $\approx$ 4.5 mel at $\approx$ 80 dB SPL. Finally, the size of the CB was assumed to comprise 100–180 mel (depending on frequency range), with an average somewhere around 137 mel [30.190, p. 557] or, rather, 100 mel [30.88, p. 167]. Ideas expressed by Stevens on CBs relating to the BM and data published by *Georg von Békésy* [30.191] and other researchers relating stimulus frequency to place of maximum excitation on the CP led to a cochlear map and a formula proposed by *Greenwood* [30.192]. From the hypothesis that CBs (without any overlap) represent equal distances on the BM, the position x of a frequency f on the BM (apex to base) can be calculated like

$$f = A \times (10^{ax} - k) \,,$$

in which A, k, and a are specific for certain species, among them humans [30.192, 193]. The graph of the pitch function (Fig. 31.4 and Fig. 31.5) shows that, *over most of the BM, and most of the frequency range, log frequency versus basilar position is nearly a straight line* [30.193, p. 2602]. However, this is not the case for low frequencies ($f < 500$ Hz) where the curve flattens markedly towards the apex (Fig. 31.4).

In regard to an anatomical and physiological interpretation of the mel scale, *Zwislocki* [30.151, p. 55] concluded *that equal distances on the pitch scale correspond to equal numbers of peripheral neurons*. He gave the following correspondences

$$1 \text{ mel} \mathrel{\widehat{=}} 12 \text{ neurons},$$
$$1 \text{ JND} \mathrel{\widehat{=}} 52 \text{ neurons},$$
$$1 \text{ CB} \mathrel{\widehat{=}} 1300 \text{ neurons}.$$

Zwicker et al., who had carried out experiments on *ratio pitch* independent of Stevens, yet had employed the method of fractionation (*half as ...*) and established a scale centered at 125 Hz where ratio pitch nearly converges with frequency up to about 400 Hz. For higher frequencies, the increase in frequency to produce same ratios is similar to the *mel scale* [30.109, pp. 57 ff.]. *Zwicker* and *Fastl* [30.152, p. 162, Table 6.2] relate the scale of 24 CBs (expressed in Bark(z)) to the length of the human BM (about 32 mm), the number of JNDs, mels, and the number of hair cells. From this table, the following correspondences are given

$$24 \text{ Bark} \mathrel{\widehat{=}} 32 \text{ mm} \mathrel{\widehat{=}} 640 \text{ JND} \mathrel{\widehat{=}} 2400 \text{ mel} \mathrel{\widehat{=}} 3600 \text{ hair cells}$$
$$1 \text{ Bark} \mathrel{\widehat{=}} 1.3 \text{ mm} \mathrel{\widehat{=}} 27 \text{ JND} \mathrel{\widehat{=}} 100 \text{ mel} \mathrel{\widehat{=}} 150 \text{ hair cells}$$
$$0.01 \text{ Bark} \mathrel{\widehat{=}} 13\,\mu\text{m} \mathrel{\widehat{=}} 1 \text{ mel}$$
$$1 \text{ JND} \mathrel{\widehat{=}} 3.8 \text{ mel}$$
$$1 \text{ mm of BM} \mathrel{\widehat{=}} 20 \text{ JND} \mathrel{\widehat{=}} 75 \text{ mel}$$

Questions concerning the validity of the *mel scale* must have started when observations from the experiments of *Stevens* and *Volkmann* [30.113] could not be replicated in another laboratory [30.194]. In 1956, experiments performed at Stevens' lab led to effects that gave even more concern in that observers, when asked to divide a *tone distance* of 400–7000 Hz into four equal appearing parts, did not choose identical frequencies in ascending and in descending order (different from what Stevens had expected; [30.120]). Also, hypotheses Stevens had formulated with respect to equal pitch differences (in mels) corresponding to equal distances on the BM could not be confirmed, while the actual data supported models of the cochlea map where distance on the BM relates to log frequency as well as to musical intervals ([30.120] and Sect. 31.3). *Stevens* [30.88, p. 168] maintained that musical intervals based on frequency ratios such as 2 : 1 or 3 : 2 would not be perceived as identical in different registers, and instead *increase in perceived pitch extent with increasing*

stimulus frequency. Seeing that musicians tune their instruments (e.g., the strings of a violin) to musical intervals, Stevens dispraised them for tuning *to the musical scale, not to the scale of subjective pitch*.

The claim that a scale of *subjective* or *apparent* pitch would exist independent of musical pitch scales has been refuted, from various angles. First, there is sufficient empirical evidence to conclude that, at least for subjects with some musical background, *equal ratios of frequency give rise to approximately equal intervals of pitch* [30.195, p. 412]. Second, experiments involving transposition of tonal patterns showed that both subjects with and without musical training carry out transpositions in a log frequency medium, which is suited to represent as well as to preserve melodic and other tonal patterns as invariant under transposition [30.196].

While the musical scale based on log frequency is form-bearing (*morphophoric*), in that melodies can be shifted up and down the scale, the *mel scale* as proposed by *Stevens* and *Volkmann* [30.113] is not. The very term *melodic scale* thus is a misnomer. The only finding that may support ideas connected with the *mel scale* is that, above 5000 Hz, in particular subjects lacking musical training tend to perform *tone distances* (when asked to realize a tonal pattern) that do not conform to the size of musical intervals. Explanations for such performance include mechanisms of pitch perception where an upper limit of temporal coding seems to exist at about 4–5 kHz (Sect. 31.4) as well as unfamiliarity with the sensation of pure tones or even small noise bands [30.109, pp. 557 ff.], which were used for pitch production in a high frequency range in order to avoid musical allusions. *Stevens* has pointed to difficulties reported by observers to deal with *those high pitchless tones* [30.104, p. 408]. A third argument against the *mel* concept was that, in a cross-cultural perspective including *Western* music, the logarithmic musical scale accounts for octave identity (a *universal* or at least *near-universal* auditory *and* musical phenomenon; Sect. 31.1.2) as well as for the representation of modal structures and melodic contours based on a hierarchy of pitches [30.197]. Finally, the inadequacy of the *mel scale* becomes fully apparent if a piece of music is realized with this scale (for compelling melodic and harmonic examples, listen to demonstration no. 49 provided by [30.198]).

Though *Licklider* [30.199, p. 1003] had characterized the *mel scale* as representing *the ordinary type of pitch perception*, this is not quite correct. The experiments conducted by Stevens (as well as by Zwicker) evidently are directed to but one of the dimensions constitutive of pitch perception, namely *tone height* while *tone chroma* (the property of tones to be part of a scale organized in intervals that can be identified as meaningful entities; see Sect. 31.1) was disregarded. Stevens himself wasn't even sure whether *subjective pitch* would conform to a metathetic continuum, or if the low end (frequencies below, perhaps, 200 Hz) should be viewed as prothetic [30.104, pp. 406–409]. He saw clearly that *pitch* is a sensory attribute that does not lend so easily to scaling outside and independent of musical pitch [30.88, pp. 164–69]. Though concepts of *ratio pitch* and the *mel scale* continued to be outlined in books on psychophysics [30.152, Chap. 2] and [30.1, pp. 354–356] and [30.4, pp. 295–297], and the *mel pitch function* also has been used in signal processing of speech (where cepstral coefficients are calculated on the basis of the *warped* mel scale [30.200]), the attraction such constructs as the mel scale had in the heyday of *operationism* is certainly lost.

Though scaling of sensory attributes has continued (and is inevitable, if only for practical reasons such as sound and noise control), attempts at establishing *the* psychophysical law applicable to a range of modalities and sensory attributes might be met with some reluctance. However, there have been alternative approaches to formulating general laws with respect to perception and cognition. One such law evolved from concepts of multidimensional scaling and was directed to perceptual generalization, viewed as *a cognitive act, not merely a failure of sensory discrimination* [30.201, p. 1319]. The concept of similarity, basic to our experience and, therefore, at the core of various methods of unidimensional and multidimensional scaling, has been employed in formulating a general recognition theory [30.202, 203].

Another approach stems from information theory. Regarding perception as a choice between alternatives, *Norwich* [30.204] has offered a *fundamental equation*

$$F = kH\,,$$

where F is a perceptual variable, $k > 0$ is a constant, and H is the entropy. From the *fundamental equation*, an equation called *the entropy law*

$$F = \frac{1}{2}k\ln(1+\gamma I^n)$$

is derived with I = intensity and γ being a constant (the value for γ as well as for the exponent n need to be determined from experimental data), which is said to embrace both the *laws* of Weber and Fechner as well as of *Plateau* and *Stevens* [30.204, Chap. 10]. The entropy function can be fitted to empirical data from, for example, experiments on loudness.

Sensations Briefly Re-viewed

Concluding the chapters on measurement of sensation, an issue that has been brought up since Fechner will

be briefly re-viewed, namely, the nature of *sensations*. It should be recalled that, in the 19th century, psychology as a discipline largely developed from two angles, one being sensory physiology, the other philosophy (for much of the historical background, see [30.144]). Sensations had to be treated, on the one hand, in regard to sensory organs, their anatomy, physiology, and specific functions. In addition, characteristics of physical stimuli had to be investigated to see how peculiar sensations would relate to properties of the stimulus. On the other hand, the *content* of the sensations as providing the substrate for perceptions was of interest. In this regard, sensations have been viewed as *elementary processes of consciousness* [30.49, p. 30] dependent on the stimulus input as well as the function of the respective sensory channel from periphery to center. The concept of sensation in this way comprises aspects beyond the physiological level, in that a sensation can convey a *quality* or express an *intensity*, and leads to further categorization of the input involving previous perceptual experience and memory. This is the framework that underlies much of Helmholtz's *Tonempfindungen* (as well as his writings on physiological optics).

From an empiricist and realist perspective, it can be argued that sensory data, which result from peripheral stimulation such as sound waves and a transduction process of physical energy into neural code, contain information that is analyzed at a number of stages of the (auditory) pathway up to the cortex, to yield percepts that can be further processed in various ways. A functional view thus adopted of course needs to go into (anatomical, physiological, psychophysical) detail provided, for the most part, by empirical research. Such research reveals that sensory input is analyzed, in the auditory system (as well as in other sensory channels) with respect to stimulus features, with the purpose of identifying *what* the input represents [30.45]. Parallel, and in specific neural pathways [30.8, Chap. 6], the input is analyzed as to *where* it comes from. Information from both streams (identification, localization) then is combined into some code representing an (auditory or visual) *object* at higher levels where further perceptual processing as well as motor responses may be started.

The nature of sensation and perception has also been an issue in philosophy where, for example, extensive debates concerning the status and purposes of *sense-data* have taken place [30.205, Chap. 7] long before major camps in philosophy took a monistic turn in regard to the time-honored mind-body problem. From a monistic point of view inspired by neuroscience, there is no question that musical sound we hear is transformed into *sequences of electrochemical pulses streaming up the axons of neurons* [30.15, p. 49], and that, even though these pulse sequences as such are *toneless signals*, they can be analyzed in the CNS so that we perceive a sound by noticing at least some of its attributes. What we perceive by hearing a sound (say, the opening chord of an electric guitar in *A Hard Day's Night*) of course is not the *toneless signal* of neural pulse trains but a representation of the physical stimulus. Representation in this respect means that spectral and temporal features of the sound are transformed into an auditory image (Sect. 31.5), which permits further analysis into constituents that are perceived as phenomenal attributes. From a logical point of view, the relation between sensation and perception in principal is hierarchical (having a *sensation* is a necessary condition for *perception*; [30.205, Chap. 7]) though a gradual transition is possible (Sect. 30.1.2). For example, it can be said that, on the level of sensation, a set of tones of nearly equal sound level and almost identical onset time is heard as a complex sound while, on the level of perception, it is categorized as a chord or sonority [30.206, p. 208]. Even a phenomenologist like *Husserl* [30.207] acknowledged that comprehension often starts from *outer perception* (conceived as a basically receptive mode of becoming aware of things), but is soon turned into acts of judgment and predication, from which generalized empirical concepts (such as the *type*, perceived as a recurrent combination of certain features) and finally notions on higher levels of abstraction result. Hence, sensing things in the *world out there* can be regarded the initial step in a process leading to perception that includes more detailed categorization of the sensory input; still one step further, a subject may possibly achieve apperception (apprehension), which, in general, calls for conscious analytical access [30.206].

Part of the argument leveled against Fechner had been that measurement of the intensity of sensations is impossible since this would require measurement of intensities inside the organism. In that *inner psychophysics* could not be pursued to much effect in those days, the opposition concluded that *outer psychophysics* only performed measurement of the physical stimulus, and that Fechner's approach to *indirect* measurement of sensation in fact meant not measuring sensation at all. By the time Stevens set to work on psychophysics, factual knowledge of anatomical structures and physiological processes relevant to hearing had grown significantly [30.146]. Correlations between properties of auditory stimuli and response patterns in the AN had been reported [30.208]. In this respect, *inner psychophysics* including measurement of intensities on the neural level had become feasible. Some findings obtained from electrophysiological recordings of cortical potentials yielded a correspondence

of stimulus intensity and amplitude of potentials that seemed to corroborate several of Stevens' power law exponents [30.209, 210]. The amplitude of potentials recorded as a function of stimulus intensity was interpreted as *strength of sensation*. Notwithstanding such attempts at introducing objective methods into psychophysics from neurophysiology, much of the research kept a focus on behavioral studies where responses to stimuli were obtained from methods such as fractionation, magnitude estimation, and cross-modal matching. The main reason why psychophysics often was constrained to investigations of (in many experiments, unidimensional) stimulus–response patterns perhaps is that, in an era when behaviorism and operationism were leading paradigms, observation of *public* behavior was deemed scientific. Even in his final book, *Stevens* reconfirmed this position:

> *What science means by sensation is a construct, a conception built upon the objective operations of stimulation and reaction. We study the responses of organisms, not some nonphysical mental stuff that by definition defies objective test.* [30.88, p. 51]

Sensations thus were taken as *behavioral reactions*, not the psychic processes Fechner seems to have envisaged.

For the purpose of this chapter, sensations are regarded, in general, as processes induced by physical stimuli that lead to specific patterns of excitation of sensory organs as well as to corresponding specific patterns of neural coding (for a recent overview of relevant facts and models, see [30.45]). For example, sensation of tonal roughness and beats has been explained, from *Helmholtz* [30.211] to the present, with respect to spectral and temporal features of sound stimuli giving rise to peculiar BM excitation patterns that in turn are transformed into neural code reflecting modulation. By inference, a certain code pattern found at some of the relays of the AuP indicates a corresponding *sensation* is likely to occur in the organism. From a similar perspective, current research seeks to trace certain stimulus features as encoded in neural spike trains along the AuP up to the cortical level (Chap. 31). According to rate-place concepts of neural organization, activation of certain areas of the brain, and in particular of cortical areas, would be viewed as a token for *perception* to take place.

30.2 Types of Sound and Sound Features Relevant for Hearing and Music Perception

Music comprises a broad range of sounds that can be characterized with respect to their temporal and spectral features. Both interact to yield dynamic *shapes* observable for single sounds as well as for compounds of several sounds that are layered (as in many types of actual music). The production of complex sounds can be effected by means of musical instruments including the human voice; of course, complex sounds are also generated and sensed by other mammals as well as by various vertebrates and invertebrates (for an overview, see articles in [30.212, 213]). In regard to sound generation and sound radiation, there are (physical and biological) systems capable of undergoing periodic or nonperiodic vibration, which can result in audible sound under suitable conditions. A detailed account of relevant facts and models (which entails stochastic and deterministic vibration, linear and nonlinear vibration, etc.) is found in textbooks such as [30.4, 125, 126, 214–217]. For the purpose of the present chapter (where only deterministic vibration is of interest), it suffices to distinguish between (a) periodic vibration, (b) quasiperiodic vibration, and (c) nonperiodic vibration. Periodic vibration, in a physical system such as the undamped oscillator, results in constant period length, $T(s)$, so that $f(t) = f(t+T)$. A single periodic vibration defined as a real-valued sine or cosine function with (peak) amplitude, A, and zero phase angle, φ

$$f(t) = A\sin(2\pi ft + \varphi),\ f(t) = A\cos(2\pi ft + \varphi)$$

can be represented in an amplitude spectrum by a line of length A at frequency f. Several periodic functions (which we assume to represent vibrations of a mechanical or electronic system) can be superposed so as to yield, according to Fourier's theorem [30.4, 217], a complex vibration pattern, where amplitudes, frequencies, and phase angles can be chosen at will. If individual vibrations have different phase angles, these can be plotted in a phase spectrum in addition to the amplitude spectrum. In the following example (Fig. 30.10) five simple vibrations defined by real-valued cosine functions have been superposed.

Note that the frequencies form a harmonic series where $f_n = n \times f_1$ (for $n = 1, 2, 3, \cdots, k$) and that the amplitudes decrease from one component to the next like $A = 1/n$. Also, all components are phase-locked at $\varphi = 0$. It is easy to see that Fig. 30.10 contains ten periods of vibration covering 0.1 s; hence, $T = 10$ ms, which is the inverse of the frequency with which the complex pattern of vibration repeats. Since $T = 1/f$,

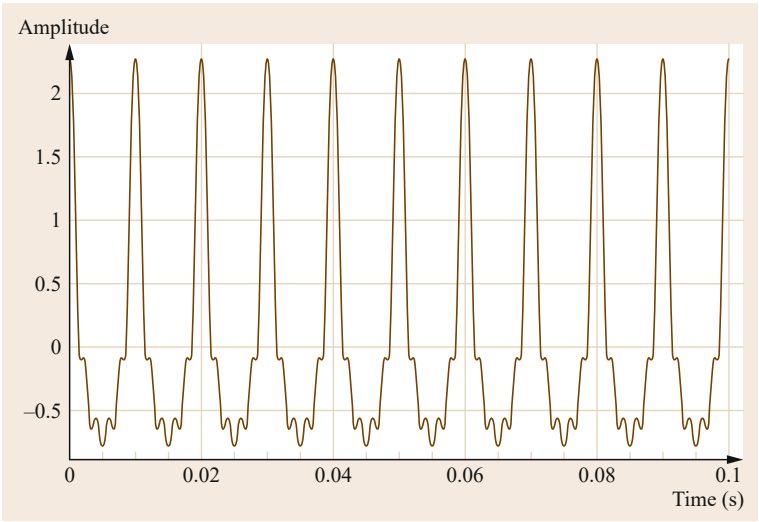


Fig. 30.10 Periodic vibration composed from five harmonic components Cos[2Pi100t] + 0.5Cos[2Pi200t] + 0.33Cos[2Pi300t] + 0.25Cos[2Pi400t] + 0.2Cos[2Pi500t]

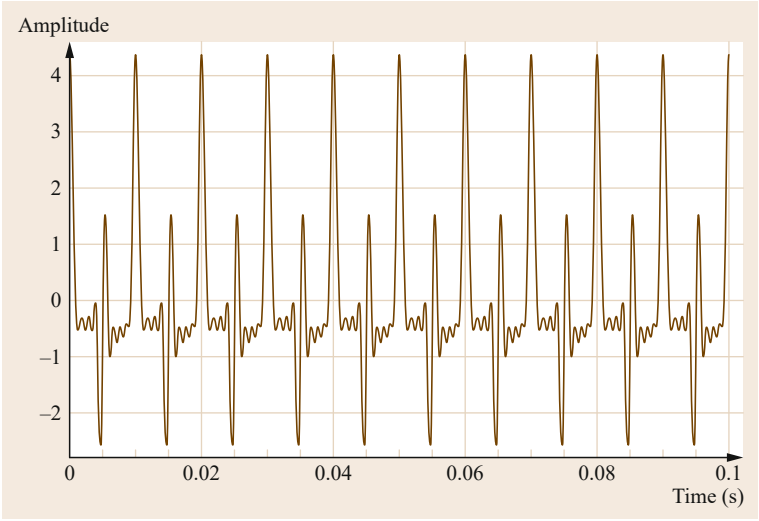


Fig. 30.11 Superposition of nine harmonic components Cos[2Pi100t] + 0.9Cos[2Pi200t − 10Degree] + 0.8Cos[2Pi300t + 20Degree] + 0.7Cos[2Pi400t − 30Degree] + 0.6Cos[2Pi500t + 40Degree] + 0.5Cos[2Pi600t − 50Degree] + 0.4Cos[2Pi700t + 60Degree] + 0.3Cos[2Pi800t − 70Degree] + 0.2Cos[2Pi900t + 80Degree]

also $f = 1/T$ is obtained. The frequency with which the complex vibration or waveform repeats per time unit in this chapter will be indicated as f_0. Since f_0 refers to a complex vibration or sound (as might be displayed on an oscilloscope), it has also been labeled the *oscillatory frequency* [30.3, p. 87]. In the case of the complex vibration shown in Fig. 30.10, f_0 equals the frequency of the lowest harmonic, $f_1 = 100\,\text{Hz}$ so that here $f_0 = f_1$. If the signs $(+, -)$ and the phase angles are used to manipulate the relations between harmonics, a more complex pattern can result. Figure 30.11 shows superposition of nine harmonics, however, the amplitudes are now decreasing like $\{1, 0.9, 0.8, 0.7, \cdots, 0.2\}$, and all components have different phase angles.

Inspection of the graph reveals that there are still ten periods of $T = 10\,\text{ms}$ defined by the strong peaks, however, a second periodicity is now visible with smaller peaks in between the main peaks. From a comparison of Fig. 30.10 and Fig. 30.11, one may infer that the phase relations between harmonic (and also inharmonic) components have a significant effect on the resulting shape of a complex vibration. The patterns shown in Fig. 30.10 and Fig. 30.11 of course can be put into digital sound synthesis so that the waveshapes become audible as complex tones.

Nearly periodic vibration can be observed in the waveshapes of many sounds generated in analogue and digital synthesizers but also in aerophones such as organ flue pipes. For example, the tone C_4 played with the stop Prinzipal 8′ on the historic organ of Hollern (near Hamburg, built by Arp Schnitger 1688–1690, fully restored 2010, tuned to $A_4 = 462.65\,\text{Hz}$) in its steady

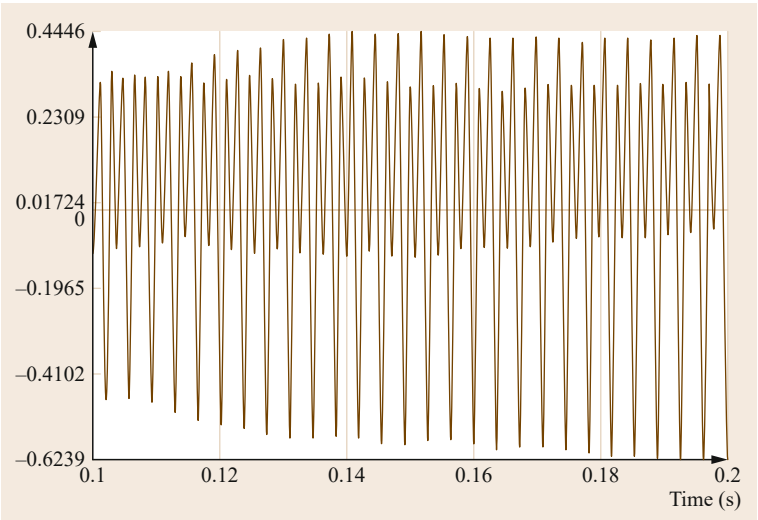


Fig. 30.12 Schnitger organ at Hollern, Prinzipal 8′, tone C_4, periodic waveshape, 100 ms

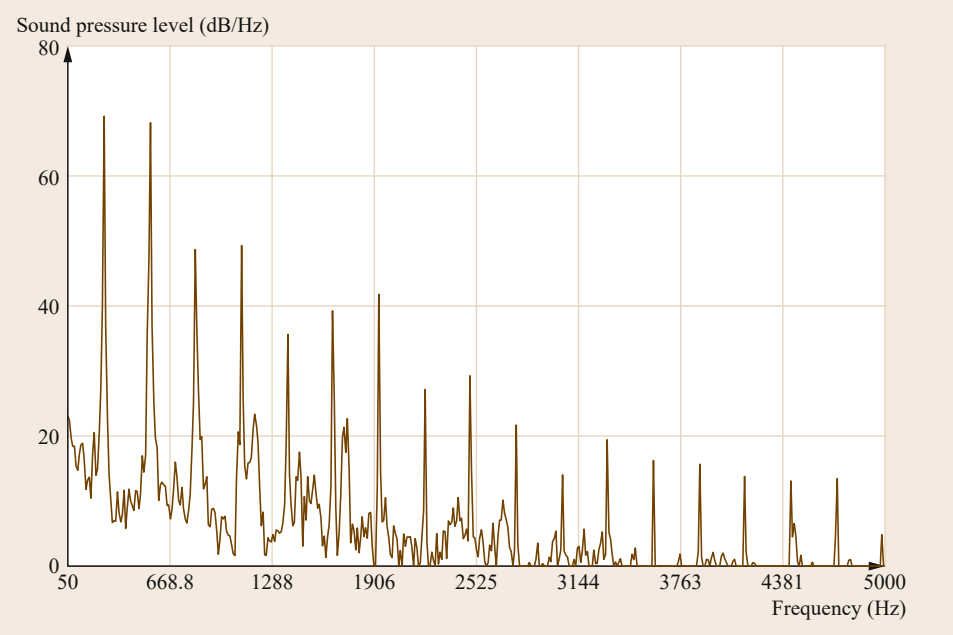


Fig. 30.13 Spectrum (LTAS), organ sound (Prinzipal 8′, tone C4, $f_1 \approx 276.5$ Hz)

state (after transients) has an almost perfect periodic waveshape, of which 100 ms (section 0.1–0.2 s of the sound) are displayed (Fig. 30.12).

Single periods have durations of ≈ 3.63 ms corresponding to ≈ 276.5 Hz, which is the fundamental frequency (f_1) of the first harmonic in the spectrum; 17 harmonic components are shown in the long-term average spectrum (LTAS) in Fig. 30.13.

The correspondence between a periodic (or nearly so) vibration or waveshape and a harmonic spectrum is covered by the Wiener–Khintchine theorem, named after seminal works published by *Wiener* [30.218] on generalized harmonic analysis and by *Khintchine* [30.219] on correlational analysis of stationary random processes. Khintchine proved that the correlation function, $R(t)$, of a stationary random process $x(t)$ can be spectrally decomposed, and that a mean exists for such a correlation function. Wiener demonstrated that, in a time series $s(t)$, the single statistical independent parameter is the autocorrelation function (ACF), $\Phi(\tau)$, which permits calculation of an expected mean [30.220, 221]. The practical implication is that a period or at least a *quasiperiod* of length T can be determined for time series where a high rate of fluctuation is evident

and periodicity is unclear because the *quasiperiod* – if it exists for a given time series – can be much extended (there of course is a limiting value at $T \to \infty$; [30.219]) or *hidden* in data due to a noisy and/or inharmonic signal. Areas where the approach outlined by *Wiener* and by *Khintchine* can be applied include extraction of periodicities from noisy signals such as EEG recordings with which Wiener worked later on to explore possibilities of ACF techniques [30.221]. In regard to music, there are idiophones such as some of the gongs found in Javanese and Balinese *gamelan* where periodicities in the sound signal are difficult to determine due to spectral inharmonicity and spectral density; analyzing such sounds with ACF can help to reveal quasiperiods.

The Wiener–Khintchine theorem, which plays an important role in signal processing [30.4, Chap. 14], [30.222], states that the power spectrum $W(\omega)$ of a stationary time function $f(t)$ equals the Fourier transform of its ACF, $\Phi(\tau)$. The Wiener–Khintchine theorem thus substantiates the role of two principles fundamental to acoustics and hearing: periodicity and harmonicity. ACF analysis (performed with the algorithm for *enhanced ACF*; [30.223]) of the organ sound displayed in Fig. 30.12 and Fig. 30.13, the tone C_4 recorded from the Prinzipal 8′ stop, yields the fundamental, f_1 (3.6 ms ≈ 276 Hz) and shows strong peaks also at lags (ms) corresponding to subharmonics of f_1 (e.g., 10.8 ms ≈ 92 Hz). The fact that the fundamental of a harmonic complex sound as well as its subharmonics can be determined by means of ACF is important also for pitch perception (Sects. 31.4–31.5).

Sounds from natural instruments as well as from the human voice are not always strictly periodic, though. Deviations from strict periodicity can be caused by various factors, among which are:

1) Fluctuations resulting from playing or singing techniques, for instance use of vibrato as is common in violin, cello and flute playing as well as in belcanto singing style.
2) Spectral inharmonicity which corresponds to (often complex) patterns of vibration.
3) Transient onset behavior as observable in sounds of many musical instruments.

In the following, some examples for these three groups will be discussed.

1) Vibrato implies a modulation of period length and, correspondingly, of frequency (FM). The notion *frequency*, in this context, changes its meaning. According to $f = 1/T$, the frequency of a stationary sinusoidal indicates the number of periods per second. However, if the period length is continuously varied with time (whereby, in the most simple case, the variation in the parameter again is periodic and corresponds to a sine or cosine function), so is the frequency of a sinusoidal, which thus becomes the *instantaneous frequency*. In terms of signal and systems theory [30.217, Chap. 6], [30.4, Chap. 19], the instantaneous frequency in an frequency modulation process can be calculated from the instantaneous or local phase $\varphi(t)$, which is a real-valued function for the complex-valued signal $\boldsymbol{x}(t)$. The instantaneous angular frequency can be determined as the derivative of the phase:

$$\omega(t) = \varphi'(t) = \frac{\mathrm{d}}{\mathrm{d}t}\varphi(t) \; ;$$

the instantaneous frequency then is

$$f(t) = \frac{1}{2\pi}\varphi'(t) \; .$$

In a practical perspective applicable to musical sound, one can say that the instantaneous frequency is that calculated for any single period of a quasiperiodic signal fluctuating more or less regularly up and down in pitch (see [30.224], with an example from Near-Eastern melismatic singing). For a number of periods, each of which differs in length (or duration as measured in *ms*), a mean period length can be calculated, to which a mean frequency corresponds as its inverse. Again, in a practical perspective, such quasiperiodic signals where the period length of any period deviates from the mean within certain limits, can be expected to evoke an *average pitch* corresponding somehow to the mean period length as well as to the mean frequency calculated for an interval relevant to auditory perception (Chap. 31).

One can easily build FM sounds from sine and cosine functions using the basic equation

$$y(t) = \sin(2\pi f_c t + \Delta f \cos(2\pi f_m t)) \; ,$$

where f_c is the carrier frequency, Δf is the peak deviation from the carrier and f_m is the modulation frequency. For example, in Mathematica code, the input

```
Play[Sin[2Pi 500t + (20 Sin[2Pi 5t]/5)], {t, 0, 10},
    SampleDepth → 16, SampleRate → 44100]
```

results in a FM sound object with audible pitch shifts up and down (similar in quality to singing a vowel such as /a/ with vibrato). If represented as a spectrum (Fig. 30.14), this FM sound reveals why the concept of *frequency* for such signals needs further consideration.

It is obvious that there are major spectral components symmetric to the center frequency, which is

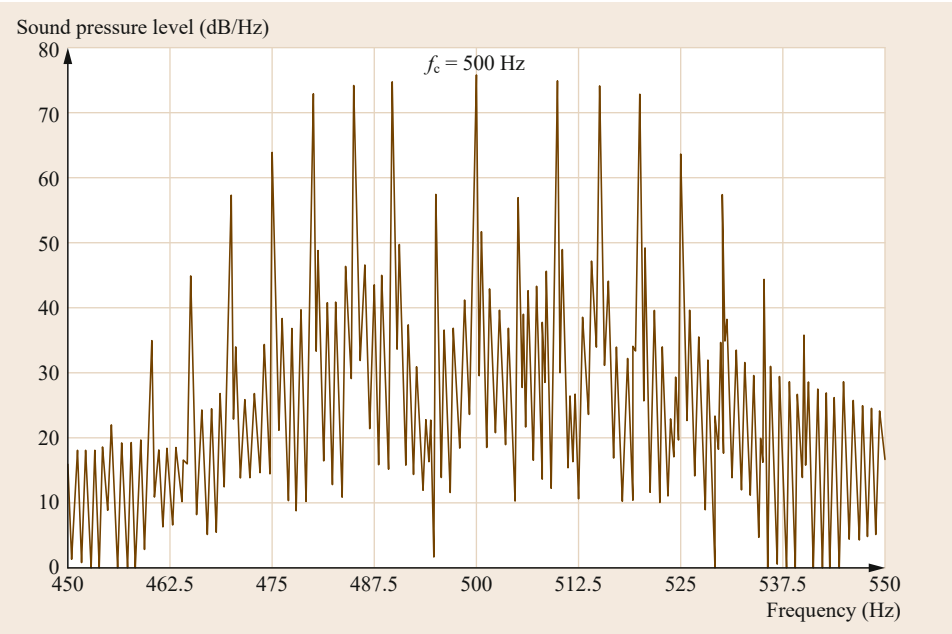


Fig. 30.14 Spectrum for FM sound Sin[2Pi500*t* + (20Sin[2Pi5*t*]/5)]

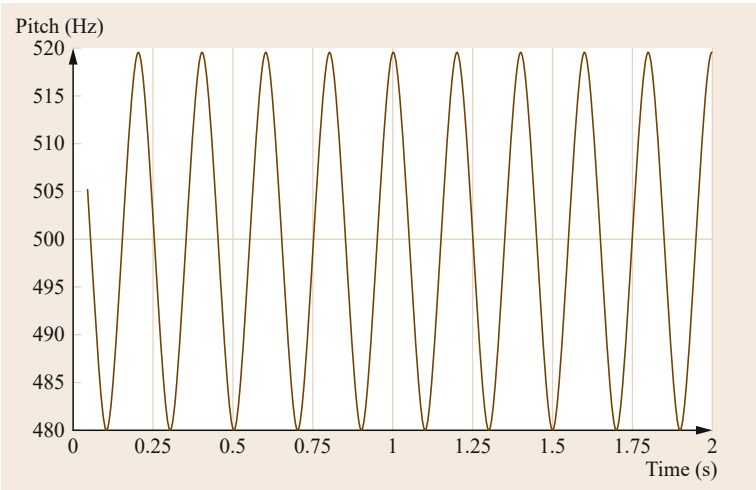


Fig. 30.15 Pitch track (f_0) for FM signal calculated from ACF

$f_c = 500\,\text{Hz}$ and represents the carrier of the FM sound. The other spectral components are at multiples of the modulation frequency, $f_m = 5\,\text{Hz}$ relative to f_c. The frequency spectrum, because of averaging frequencies over time for a certain interval (defined by the length of the fast Fourier transform (FFT) window; [30.4, 217, 222]) does not reveal specific temporal information any longer as to when exactly the period of the signal did correspond to one of the frequencies shown as peaks in the spectrum. Conversely, when the pitch track representing f_0 of the signal is calculated from the ACF (Fig. 30.15), it becomes evident that f_0 varies steadily over time; consequently, there is an instantaneous frequency for any time point. From a practical perspective, one can take a frequency f_0 corresponding to each single period to find the pitch track for a musical signal that includes FM such as a piece played on a violin *con espressione*, that is, with finger vibrato (Chap. 32).

2) Another factor responsible for the existence of quasiperiodic sounds is inharmonicity of spectral components. Inharmonic spectra are found, due to vibration patterns and wave propagation in solids [30.116, 215, 225], in almost all idiophones such as xylophones of West and Central Africa and metallophones (gongs, gong chimes, keyed metallophones comprising thick

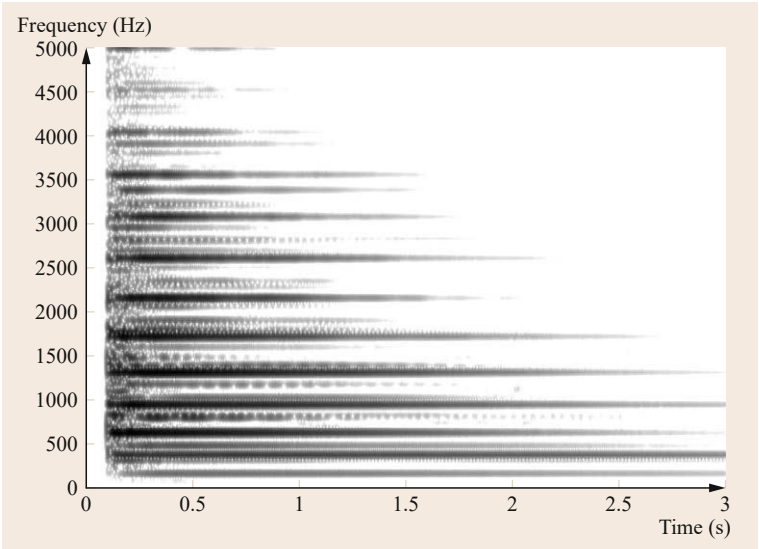


Fig. 30.16 Bell no. 7, Brugge carillon, spectral components 0–5 kHz

bronze or iron bars coupled to resonators) prominent in ensembles of South East Asia (e.g., Javanese and Balinese *gamelan*). Spectra of sounds recorded from vibration in solids are inharmonic because phase velocity of transversal as well as axial bending waves is dependent on frequency, that is, roughly, $c_B \propto \sqrt{f}$. Spectral inharmonicity of xylophone bars can be reduced by optimizing the geometry as well as by the use of synthetic material such as *palisono* (which resembles wood in certain features but is much more homogeneous and isotropic in structure). Similarly, the body of swinging and carillon bells can be shaped so that an almost harmonic spectrum results instead of the traditional spectral structure of *Western* bells where, typically, a minor third is found in one octave and a major third in the next (the frequency ratios of strong bell partials within the lower octaves are, typically, close to 1 : 2 : 2.4 : 3 : 4 : 5 : 6 : 8.1). Because of this peculiar inharmonicity (and additional inharmonic components), also the time series of sound signals recorded from bells is, at best, quasiperiodic; typically, there is no simple repetition pattern observable so that the length of the period τ is difficult to determine (for instance, by ACF methods). Sounds recorded from bells are often quite complex in regard to temporal and spectral features (attack, buildup of modes, sometimes also AM due to inharmonicity and spectral density [30.226, 227]), and hence are not easy to analyze by ear. As an example for a fairly complex spectral structure, bell no. 7 of the famous carillon of Brugge (cast by J. Dumery in 1744), shall be examined. Figure 30.16 is the spectrogram of the bell sound (displayed is the frequency range from 0–5 kHz for 3 s) where at least a dozen or so strong components can be identified plus a number of components carrying smaller amounts of energy. The main spectral components of this bell sound in the range up to 1 kHz are at the frequencies presented in Table 30.5 (for which approx. ratios and cents are given).

Interval (1) expresses the relations from one component to the next while Interval (2) indicates only large framing intervals relative to the lowest component labeled *fundamental* in this scheme (bell founders and campanologists like to use a nomenclature based on *bell partials*). From the frequency ratios it is evident that, except for the minor third, most of the components are nearly harmonic relative to the fundamental; however, the cents reveal that, already in the second octave above the fundamental, the major third is too sharp (to

Table 30.5 Brugge carillon, bell no. 7, main spectral components and bell partials

No.	Hz	Ratio	Cents	Cum.	Interval (1)	Interval (2)	Bell partial
1	153	1		0	Fundamental	Fundamental	Hum
2	306	2	1200	1200	Octave	Octave	Prime
3	367	2.4	315	1515	Minor third		Tierce
4	470	3.07	428	1943	Major third	Twelfth	Quint
5	617	4.03	471	2414	Fourth	Double octave	Nominal
6	792	5.18	432	2864	Major third		
7	935	6.11	287	3151	Minor third		

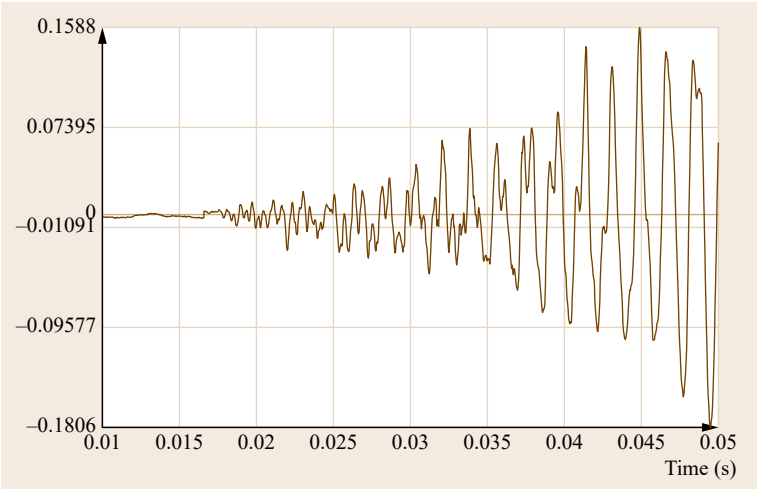


Fig. 30.17 Transient part of organ sound, Prinzipal 8′, C_4, time series for 50 ms

the effect that the twelfth relative to the fundamental is sharp) while the fourth completing the double octave is flat. Such deviations rather tend to increase in higher octaves. Moreover, in bell spectra the main partials often are found interspersed with smaller inharmonic components suited to complicate the waveshape resulting from the sum of all components, on the one hand, and to diminish its periodicity, to the other. The overall perceptual effect of spectral inharmonicity in bells is ambiguity of *pitch* (Sect. 32.2.4).

3) As a third group of sound phenomena lacking clear periodicity, transients have to be mentioned. One has to see that sound production in many instruments is effected by means of an impulse (like in percussion instruments as well as in plucked strings), or by a sequence of pulses like in aerophones where either a valve (for instance, a double reed as in the oboe or bassoon, or two lips pressed against each other, as is done by players of brass instruments) opens and shuts (completely or partly, depending on blowing pressure), or a jet stream is deflected and interrupted when hitting the sharp edge in the mouthpiece of a flute or in the labium of an organ flue pipe (for details, see [30.215]). Both the valve mechanisms (reeds, lips, also the vocal folds of the glottis) and the edge tone generator operate in a basically periodic process (or duty cycle). However, viewed in detail, these processes are quite intricate, and include nonlinearities in several, if not many parameters (for details, see [30.228, Chap. 7] and Part A of this volume). To understand fundamentals of the transient process of sound generation in an organ flue pipe, it is sufficient to see that excitation of modes of vibration in the mass of air enclosed in the pipe must be achieved against the input impedance of the system and the inertia of the mass before standing waves are established, and a steady-state sound is radiated from the pipe. Before that, with a particular key of the organ pressed, a valve opens and air (*wind* in organ terminology) starts streaming from a reservoir into the pipe. Before a certain pressure level is reached and the edge tone generator comes into action, *wind* in fact streams through the length of the pipe (in long and wide pipes such as the 16′ or even 32′ diapason pipes, this process can last for more than 50 or even more than 100 ms). Vibration then sets in for individual modes but, typically, there are several cycles where first only a few modes are established that are not yet well synchronized, leading to a pattern of vibration that lacks clear periodicity. The onset for individual modes can be very soft (with pressure amplitudes rising smoothly) or more sudden, resulting in a somewhat plosive transient (a sound phenomenon known as *spitting*). The transient sound from organ flue pipes therefore often is noisy and only partly tonal (in regard to pitch and timbre, both viewed according to the Wiener–Khintchine principle, that is, temporal periodicity corresponding to spectral harmonicity).

The transient portion of the organ sound recorded from the Prinzipal 8′ (tone C_4) is shown in Fig. 30.17. It takes about 30 ms to establish periodic vibration firmly where the overall amplitude rises due to increasing locking of modes and, accordingly, more effective superposition of harmonics.

Looking into the first 200 ms of the same organ sound with short-time Fourier transform (STFT) spectrum analysis so that the frames (2048 pts, Hanning) highly overlap (Fig. 30.18), one can follow the transient process from one spectrum to the next. It is obvious that partials 1 and 2 appear first, however, with rather flat spectral lobes that become narrower and steeper with time. Partial no. 3 starts at about 40 ms after onset, and higher partials even later. Also, there are considerable fluctuations in partial frequencies within the first

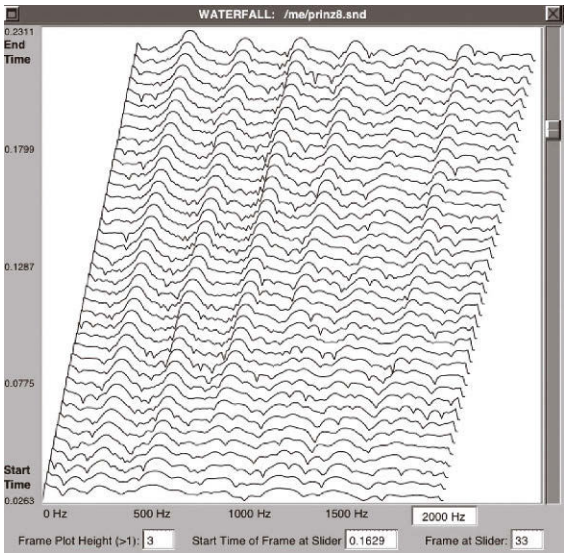


Fig. 30.18 Organ (Hollern), Prinzipal 8′, tone C_4, transient onset, 39 spectra representing ≈ 205 ms of sound

150 ms. Typically one finds the frequencies and amplitudes for each spectral component representing one of the modes of vibration to change from one frame of analysis (FFT window) to the next. Spectral flux (SF) is a parameter to measure the magnitude of such change, which can be expressed as an onset function, *SF* ([30.229] and Chap. 32).

Another method suited to analyze transients is the harmonic-to-noise ratio (HNR), which measures the degree of periodicity in the signal by either ACF or cross-correlation function (CCF) [30.230]. In the measurement presented here, the parameter actually calculated by CCF is periodicity-to-noise. The ratio is expressed in dB (a HNR of 0 dB means there is equal energy in the harmonics of a complex tone and in the noise). For the organ flue pipe, the HNR measured for one second is shown in Fig. 30.19.

As a consequence of the noisy transient, the spectral centroid (calculated with a phase vocoder algorithm; [30.9, pp. 45 ff.]) jumps to ≈ 7 kHz immediately after onset, where it peaks for ≈ 20 ms and then rapidly falls to 700 Hz and stays there for the time the sound is in its steady state (Fig. 30.20). The spectral centroid usually is calculated according to

$$C = \frac{\sum_{k=1}^{K} A_k f(t)_k}{\sum_{k=1}^{K} A_k}$$

where A_k are the partial amplitudes and $f(t)_k$ are the partial frequencies.

The spectral centroid [30.228, pp. 27 ff.] indicates the point (expressed as a frequency in Hz) of the spectrum where half of the energy is above this point, and the other half is below. The spectral centroid can be calculated for each frame of analysis (also labeled *window* in FFT algorithms; the size of the window typically is 2^n sample points, where $n = \cdots, 8, 9, 10, 11, 12, \cdots$). Taking the centroid for each frame, and having frames

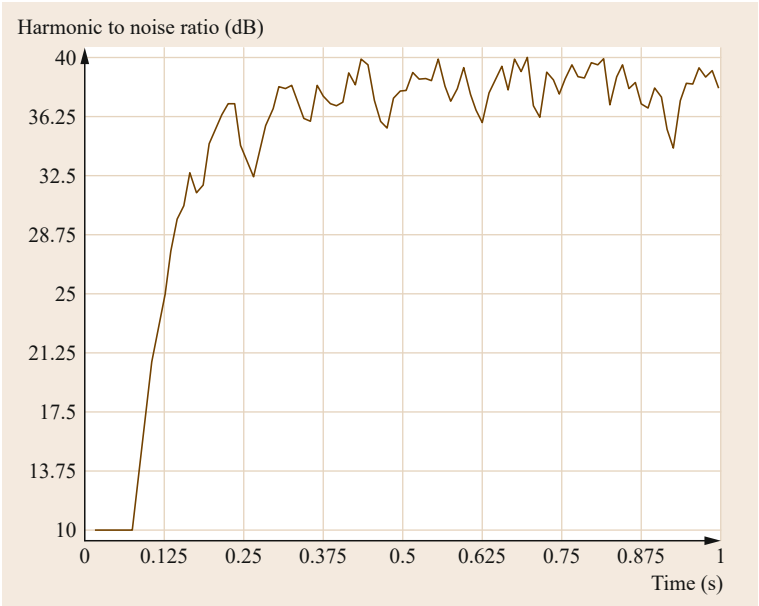


Fig. 30.19 Harmonic-to-noise ratio (HNR), onset of organ flue pipe, 1 s

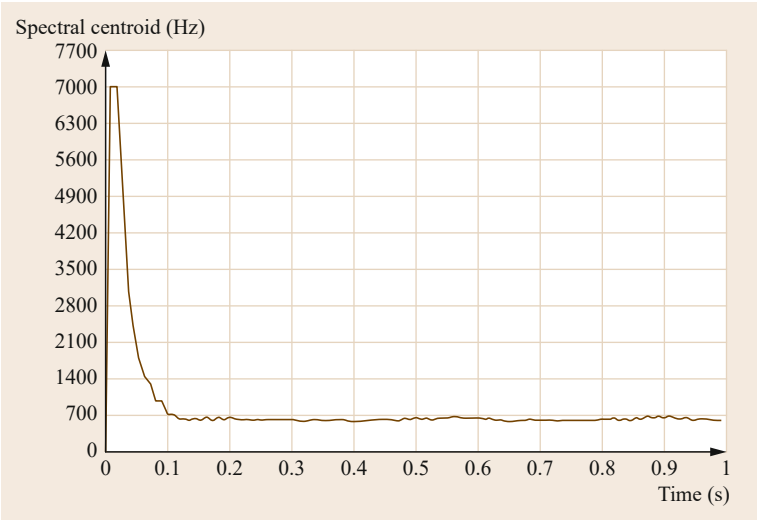


Fig. 30.20 Graph of spectral centroid plotted for 1 s of organ sound, Prinzipal 8′

overlap by at least half of the window size, the spectral centroid then can be plotted as a (quasicontinuous) function of time.

Such curves are of interest since the spectral centroid is a signal parameter that is closely related to the psychoacoustic dimension of *brightness*. Increasing spectral centroid means a shift of energy in the direction of higher frequencies (towards the *treble range*).

A useful parameter to describe the temporal evolution of a particular sound is its envelope. Musicians know to shape the envelope of sounds on a synthesizer by setting values for attack time, decay time, sustain and release relative to level (so-called ADSR envelope; Fig. 30.21).

As a scheme, an ADSR envelope looks like Fig. 30.21; one often can find a similarity between the envelope of a sound derived from intensity (as a measure of energy flux) over time and the curve of its spectral centroid. If the envelope starts with a rapid attack and rises to high amplitude, it is very likely that the sound has been elicited by an impulse or by some other rapid transfer of energy into the system that leads to a similar effect. Concentration of energy at the onset means that, for example in a *natural* instrument such as a swinging bell where the energy is transmitted from the clapper to the bell in, on average, 1 ms or less [30.231], very many eigenmodes are elicited almost at once. Consequently, the overall amplitude of the sound rises fast to a peak but in general also soon decays because many of the higher modes die away from damping. A similar behavior can be observed in, for example, sounds recorded from single strings of harpsichords, which are plucked with considerable elongation of the string (relative to its rest position) followed by a sudden release of the string from the plectrum. Both the bell and the thin

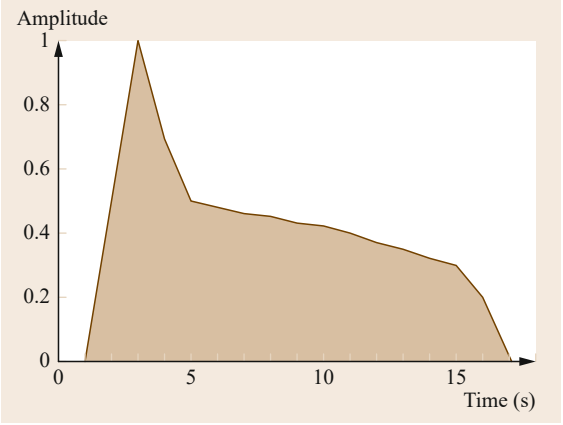


Fig. 30.21 Schematic shape of ADSR envelope (amplitude or level over time)

strings of the harpsichord exhibit a rapid attack, a fast decay, and a long sustain.

To conclude this section, there are sounds from idiophones (such as gongs, xylophones, cymbals) and membranophones (various types of drums), which are quite short and percussive, and which mostly also have an inharmonic spectrum. A number of instruments may even exhibit a quasicontinuous spectrum within a certain bandwidth similar to a noise signal. For example, if a snare drum is played with the snares applied, the modes of the upper head are mixed with the noisy sound of the lower (resonant) head to which the snares are attached. Of course, the actual sound produced depends on the type of drum as well as on the heads and snares used, the tension of both membranes and the force applied to the upper head, the place of excitation, etc. Figure 30.22 is a spectrogram of a snare drum played

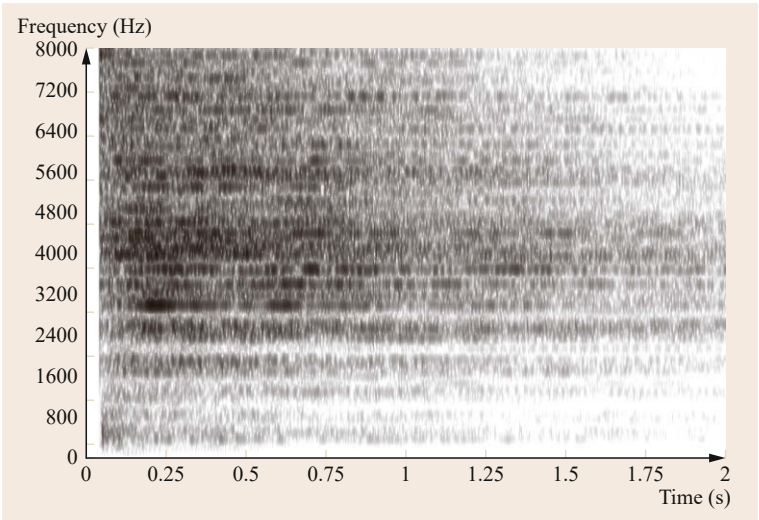


Fig. 30.22 Snare drum, single stroke with stick, spectrogram 0–8 kHz, 2 s

with a single stick for one stroke suited to illustrate the case where spectral energy distribution is close to quasicontinuous. The sound was recorded with a single condenser microphone placed in a position where a player sitting at a drum set would be expected to have his or her left ear.

The spectrogram reveals a fast attack, with much of the energy radiated from the drum within the first second. In fact, the SPL within 0.7 s drops by about 40 dB.

In many musical contexts, a range of complex sounds occur at the same time (or nearly so). This can lead to extremely variable waveshapes and, correspondingly, to abrupt changes in spectral energy distribution within short spans of time (consider, for example, symphonic works by Mahler or Stravinsky's *Sacre du printemps*). Perception of musical sound, therefore, requires that our sense of hearing is capable of fast adaptation to changing sound patterns from which relevant features must be detected.

30.3 Some Basics of Sound in a Sound Field

A sound field is an (ideally) homogeneous medium filled with molecules (*particles*); in a medium such as air, sound waves propagate at a speed, c, which in air at a temperature of 20 °C is about 343.2 m/s. Sound waves are generated from forces leading to movements of particles around their resting positions (leaving aside Brownian molecular movement). Field quantities relevant for sound waves and wave propagation in a given sound field are wave speed, c, pressure, p, particle velocity, v, and density, ρ. Also, temperature can be of relevance (since density and, consequently, wave speed change with temperature). Sound waves [30.126, Chap. 6] can be viewed as disturbances changing the equilibrium in the sound field. Changes of pressure, p, of particle velocity, $v = \partial\xi/\partial t$ (ξ = displacement) and density, ρ in a given medium are closely interrelated as expressed by Euler's law

$$-\frac{\partial p}{\partial x} = \rho_0 \frac{\partial v}{\partial t} \,.$$

A local change in pressure thus leads to particle motion. Assuming pressure and density in a medium like air to consist of a normal component (p_0, ρ_0) and an alternating component ($p_\sim$, $\rho_\sim$), we have $p = p_0 \pm p_\sim$ and $\rho = \rho_0 \pm \rho_\sim$. Pressure is a scalar while velocity is a vector, $\boldsymbol{v}_\sim$. According to the principle of continuity, density of air (which at sea level and 15 °C temperature is $\approx 1.2041\,\mathrm{kg/m^3}$) can be assumed to change but little in a sound field. The acoustic impedance,

$$Z = \frac{p(x,t)}{\boldsymbol{v}(x,t)} = \rho_0 c \,,$$

in air at 20 °C is $413.3\,\mathrm{N\,s\,m^{-3}}$. The magnitude of changes occurring in a sound field appears to be quite small [30.232, Chap. 2]. A SPL of 100 dB means the rms pressure amplitude is $2\,\mathrm{N/m^2}$ as compared to an average resting pressure at about $100\,000\,\mathrm{N/m^2}$ (or 1000 hPa). In a plane wave propagating in air (density $\approx 1.2\,\mathrm{kg/m^3}$) at 340 m/s the rms particle velocity

($\boldsymbol{v} = \mathrm{d}\xi/\mathrm{d}t$) is 5 mm/s at 100 dB, and the displacement, ξ, of a particle from resting position for a pure tone of 1000 Hz is just 1 μm. However, acceleration of particles can be impressively high, since, for a SPL of 100 dB and a pure tone of 1000 Hz, one obtains an acceleration of $\approx 30\,\mathrm{m/s^2}$, which is the threefold of acceleration of gravity.

The modified Euler equation (which expresses the conservation of impulse) valid for gases is

$$\rho_0 \frac{\partial \boldsymbol{v}_0}{\partial t} = -\frac{\partial p_0}{\partial x} .$$

The equation of continuity (which expresses the conservation of mass) in the gas is

$$\rho_0 \frac{\partial \boldsymbol{v}_0}{\partial x} = -\frac{\partial \rho_0}{\partial t} = \frac{1}{c^2}\frac{\partial p_0}{\partial t} \text{ where } c^2 = \frac{\mathrm{d}p}{\mathrm{d}\rho}(\rho_0)$$

is the speed of sound in air and ρ_0 is the normal density at rest. The wave equation for a plane wave traveling into one direction (along the x coordinate) can be derived thus

$$\frac{\partial^2 p_0}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} .$$

Solutions to the wave equation in one dimension are of the form $p(x,t) = f_1(x-ct) + f_2(x+ct)$ where f1 and f2 in general are functions that can be differentiated twice. The two functions express that two plane waves are propagating from the point of excitation in opposite directions along the coordinate x as in a string conceived as a one-dimensional continuum. For undamped harmonic plane waves, a possible solution is of the form

$$p(x,t) = A\cos\left[2\pi\left(\frac{x}{\lambda} \pm ft\right) + \varphi\right],$$

where A denotes the peak amplitude, λ denotes the wavelength and φ is the phase angle. Such a simple wave would have a spatial period of $\lambda = c/f$ and a temporal period of $T = 1/f$. In the harmonic wave described by the formula, sound pressure changes periodically as a function of place and time. The solution for spherical waves can be simplified to that for plane waves, except for the near field defined by the wavelength where both velocity and wave impedance are more complicated terms [30.216, pp. 100–102].

A wave transports energy through a given medium. The intensity of a sound, I, represents the mean density of the energy transported in a wave

$$I = \frac{1}{T}\int_0^T p\, v \sin^2\left[\omega\left(t - \frac{x}{c}\right)\right]\mathrm{d}t$$
$$= \frac{1}{2} p\, v = p_{\mathrm{eff}}\, v_{\mathrm{eff}} = \rho\, c\, v^2 .$$

In this formula, p and v denote peak values, and p_{eff} and v_{eff} indicate rms values. Energy density

$$w = \frac{p^2}{2\,\rho_0\, c^2} + \frac{\rho_0\, v^2}{2}$$

sums potential energy (first term on the right side of the equation) and kinetic energy (second term); w is a periodic function for both time and place. The sound energy contained in a volume V of a medium then is the potential and the kinetic energy integrated over that volume. A simple harmonic wave of wavelength λ, traveling in one direction (along the coordinate x) with a frequency of f at a phase velocity of c contains a certain amount of energy per period. If the wave goes through a plane, S (the size of which may be taken as $1\,\mathrm{m^2}$), the energy flow in a short time interval Δt will be

$$\Delta W = wSc\Delta t .$$

The ratio $\Delta W/\Delta t = J$ is the power (or strength) of the energy flux, and the ratio $J/S = j$ is the flux density, which again is a periodic function of time, implying that a wave traveling through a plane S in fact carries a sequence of energy pulses. It should be mentioned that the frequency of energy pulses is twice the frequency of a given wave (there are two energy concentrations per period as expressed by wavelength). As is evident from Table 30.1, the energy ($\mathrm{W/m^2}$) at the threshold of hearing is extremely small, indicating the fine sensitivity of our hearing organ. However, small changes in local pressure in the air are in fact the signal parameter with which sound waves are transmitted to the ears of mammals. All relevant information is contained in the fluctuations of the pressure amplitude that can be registered over time with a (suitably sensitive) microphone, and can be stored as for analysis in an analogue or digital format (e.g., as a time series of samples).

References

30.1 S. Gelfand: *Hearing. An Introduction to Psychological and Physiological Acoustics*, 4th edn. (Dekker, New York 2004)

30.2 B. Moore: Frequency analysis and masking. In: *Hearing*, ed. by B. Moore (Academic, San Diego 1995) pp. 161–205

30.3 E. Terhardt: *Akustische Kommunikation* (Springer, Berlin 1998)

30.4 W. Hartmann: *Signals, Sound and Sensation* (Springer, New York 1998)

30.5 J. Neuhoff (Ed.): *Ecological Psychoacoustics* (Elsevier Academic, San Diego 2004)

30.6 C. Plack, A. Oxenham: The psychophysics of pitch. In: *Pitch. Neural Coding and Perception*, ed. by C. Plack, A. Oxenham, R. Fay, A. Popper (Springer, New York 2005) pp. 7–55

30.7 B. Moore: *Introduction to the Psychology of Hearing*, 6th edn. (Emerald, Bingley 2012)

30.8 J.O. Pickles: *Introduction to the Physiology of Hearing*, 3rd edn. (Emerald, Binkley 2008)

30.9 J. Beauchamp: Analysis and synthesis of musical instrument sounds. In: *Analysis, Synthesis, and Perception of Musical Sounds*, ed. by J. Beauchamp (Springer, New York 2007) pp. 1–89

30.10 H. Fastl, E. Zwicker: *Psychoacoustics, Facts and Models*, 3rd edn. (Springer, Berlin 2007)

30.11 R. Meddis, E. Lopez-Poveda: Auditory periphery: From pinna to auditory nerve. In: *Computational Models of the Auditory System*, ed. by R. Meddis, E. Lopez-Poveda, R. Fay, A. Popper (Springer, New York 2010) pp. 7–38

30.12 M. Florentine (Ed.): *Loudness* (Springer, New York 2011)

30.13 R. Fay, A. Popper (Eds.): *Comparative Hearing: Mammals* (Springer, New York 1994)

30.14 Y. Cohen, A. Popper, R. Fay (Eds.): *Neural Correlates of Auditory Cognition* (Springer, New York 2013)

30.15 D. Dennett: *Consciousness Explained* (Penguin, London 1993)

30.16 P.S. Churchland: *Neurophilosophy. Toward a Unified Science of the Mind-Brain* (MIT Press, Cambridge 1986)

30.17 P.S. Churchland, T. Sejnowski: *The Computational Brain* (MIT Press, Cambridge 1992)

30.18 G. Roth: *Das Gehirn und seine Wirklichkeit. Kognitive Neurobiologie und ihre Philosophischen Konsequenzen*, 5th edn. (Suhrkamp, Frankfurt 1996)

30.19 G. Botterill, P. Carruthers: *The Philosophy of Psychology* (Cambridge Univ. Press, Cambridge 1999)

30.20 P. Carruthers: *Phenomenal Consciousness. A Naturalistic Theory* (Cambridge Univ. Press, Cambridge 2000)

30.21 A. Clark: *Mindware. An Introduction to the Philosophy of Cognitive Science* (Oxford Univ. Press, Oxford 2001)

30.22 T. Metzinger: *Being No-One. The Self-Model Theory of Subjectivity* (MIT Press, Cambridge 2003)

30.23 V. Gadenne: *Philosophie der Psychologie* (Huber, Toronto 2004)

30.24 M. Gazzaniga (Ed.): *The Cognitive Neurosciences*, 4th edn. (MIT Press, Cambridge 2009)

30.25 N. Rescher: *Nature and Understanding. Metaphysics and Method of Science* (Clarendon, Oxford 2000)

30.26 E. Rosch: Principles of categorization. In: *Cognition and Categorization*, ed. by E. Rosch, B. Lloyd (Erlbaum, Hillsdale 1978) pp. 27–48

30.27 K. Popper, J. Eccles: *The Self and Its Brain* (Springer, Berlin 1981)

30.28 E. Oeser, F. Seitelberger: *Gehirn, Bewußtsein und Erkenntnis*, 2nd edn. (Wiss. Buchges, Darmstadt 1995)

30.29 J. Searle: *Intentionality. An Essay in the Philosophy of Mind* (Cambridge Univ. Press, Cambridge 1983)

30.30 F. Brentano: *Psychologie vom empirischen Standpunkt* (Meiner, Leipzig 1924), ed. by O. Kraus

30.31 E. Husserl: *Phänomenologische Psychologie* (Meiner, Hamburg 2003)

30.32 P.S. Churchland: Epistemology in the age of neuroscience, J. Philos. **83**, 544–553 (1987)

30.33 H. Gardner: *The Mind's New Science* (Basic, New York 1987)

30.34 G. Gigerenzer: *Adaptive Thinking* (Oxford Univ. Press, Oxford 2000)

30.35 G. Edelman: *Bright Air, Brilliant Fire: On the Matter of the Mind* (Basic, New York 1992)

30.36 A. Damasio: *Descartes' Error: Emotion, Reason and the Human Brain* (Avon, New York 1995)

30.37 P. Godfrey-Smith: *Complexity and the Function of Mind in Nature* (Cambridge Univ. Press, Cambridge 1996)

30.38 H. von Helmholtz: Die Thatsachen in der Wahrnehmung. In: *Vorträge und Reden*, Vol. 2, ed. by H. von Helmholtz (Vieweg, Braunschweig 1906) pp. 213–247

30.39 H. von Helmholtz: Ueber den Ursprung der richtigen Deutung unserer Sinneseindrücke, Z. Psychol. **7**, 81–96 (1894)

30.40 G. Gigerenzer, D. Murray: *Cognition as Intuitive Statistics* (Erlbaum, Hillsdale 1987)

30.41 L. Marks: "What thin partitions sense from thought divide": Toward a new cognitive psychophysics. In: *Psychophysical Approaches to Cognition*, ed. by D. Algom (North-Holland, Amsterdam 1992) pp. 115–186

30.42 W. Wundt: Über psychologische Methoden, Philos. Stud. **1**, 1–38 (1882)

30.43 E. Boring: *Sensation and Perception in the History of Experimental Psychology* (Appleton-Century-Crofts, New York 1942)

30.44 St Coren: Sensation and perception. In: *History of Psychology*, Vol. I, ed. by D.K. Freedheim (Wiley, Hoboken 2003) pp. 85–108

30.45 T. Bossomaier: *Introduction to the Senses. From Biology to Computer Science* (Cambridge Univ. Press, Cambridge 2012)

30.46 N. Birbaumer, R.F. Schmidt: *Biologische Psychologie*, 7th edn. (Springer, Heidelberg 2010)

30.47 F.J. McGuigan: *Biological Psychology. A Cybernetic Science* (Prentice, Englewood Cliffs 1994)
30.48 W. Wundt: *Grundzüge der Physiologischen Psychologie*, 5th edn. (Engelmann, Leipzig 1903)
30.49 O. Külpe: *Grundriss der Psychologie auf Experimenteller Grundlage* (Engelmann, Leipzig 1893)
30.50 D.M. Armstrong: *Perception and the Physical World* (Routlege and Kegan Paul, London 1961)
30.51 E. Brunswik: *Wahrnehmung und Gegenstandswelt. Grundlegung einer Psychologie vom Gegenstand her* (F. Deuticke, Leipzig, Wien 1934)
30.52 W. Wundt: *Grundriß der Psychologie* (Engelmann, Leipzig 1896), 15th edn. 1928
30.53 M. Vogel: *On the Relations of Tone* (Verlag für Syst. Musikwiss., Bonn 1993)
30.54 E. Scheerer: Mentale Rrepräsentation in interdisziplinärer Perspektive, Z. Psychol. **201**, 136–196 (1992)
30.55 A. Schneider, R.I. Godøy: Perspectives and challenges of musical imagery. In: *Musical Imagery*, ed. by R.I. Godøy, H. Jørgensen (Swets Zeitlinger, Lisse 2001) pp. 5–26
30.56 A.J. Watkins, M.C. Dyson: On the perceptual organization of tone sequences. In: *Musical Structure and Cognition*, ed. by P. Howell, I. Cross, R. West (Academic, Orlando 1985) pp. 71–120
30.57 A. Paivio: *Images in Mind. The Evolution of a Theory* (Harvester Wheatsheaf, New York 1991)
30.58 P. Zimbardo: *Psychologie*, 5th edn. (Springer, Berlin 1992)
30.59 W.D. Keidel: *Biokybernetik des Menschen* (Wissenschaftl. Buchges., Darmstadt 1989)
30.60 G.S. Halford, W.H. Wilson, S. Phillips: Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology, Behav. Brain Sci. **21**, 803–865 (1998)
30.61 M. Leman: *Music and Schema Theory* (Springer, Berlin 1995)
30.62 M. Leman: A Model of retroactive tone-center perception, Music Percept. **12**, 439–471 (1995)
30.63 M. Drobisch: *Erste Grundlehren der Mathematischen Psychologie* (Voss, Leipzig 1850)
30.64 E.H. Weber: Der Tastsinn und das Gemeingefühl. In: *Handwörterbuch der Physiologie*, Vol. III, ed. by R. Wagner (Vieweg, Braunschweig 1846) pp. 481–588
30.65 G.T. Fechner: *Elemente der Psychophysik*, Vol. 1, 2, 3rd edn. (Breitkopf Haertel, Leipzig 1907)
30.66 H. Gundlach: *Entstehung und Gegenstand der Psychophysik* (Springer, Berlin 1993)
30.67 S. Horst: The role of phenomenology in psychophysics. In: *Handbook of Phenomenology and Cognitive Science*, ed. by G. Gallagher, D. Schmicking (Springer, Berlin, Heidelberg 2010) pp. 447–469
30.68 I. Kant: *Kritik der Reinen Vernunft*, 2nd edn. (Riga, Hartknoch 1787)
30.69 G.T. Fechner: *Revision der Hauptpunkte der Psychophysik* (Breitkopf Haertel, Leipzig 1882)
30.70 D. Laming: *Mathematical Psychology* (Academic, New York 1973)
30.71 J. Michell: *Measurement in Psychology. Critical History of a Methodological Concept* (Cambridge Univ. Press, Cambridge 1999)
30.72 A. Szabó: *The Beginnings of Greek Mathematics* (Reidel, Dordrecht, Boston 1978)
30.73 T. Sonar: *3000 Jahre Analysis: Geschichte, Kulturen, Menschen* (Springer, Berlin 2011)
30.74 Aristotle: *Metaphysics*, Book V, 1020a/b
30.75 R. Niederée, L. Narens: Axiomatische Meßtheorie. In: *Handbuch Quantitative Methoden*, ed. by E. Erdfelder, R. Mausfeld, T. Meiser, G. Rudinger (Beltz, Weinheim 1996) pp. 369–384
30.76 E. Cassirer: *Substanzbegriff und Funktionsbegriff. Untersuchungen über die Grundfragen der Erkenntniskritik* (Cassirer, Berlin 1910)
30.77 O. Hölder: *Die Axiome der Quantität und die Lehre vom Mass*, Berichte über die Verhandlungen der Königl.-Sächs. Ges. der Wiss, Vol. 53, 1901) pp. 1–64
30.78 H. von Helmholtz: Über den Ursprung und die bedeutung der geometrischen Axiome. In: *Vorträge und Reden*, Vol. 2, ed. by H. von Helmholtz (Vieweg, Braunschweig 1903) pp. 1–31, 381–383 Lecture delivered in 1870 at the University of Heidelberg
30.79 H. von Helmholtz: Zählen und Messen, erkenntnistheoretisch betrachtet. In: *Philosophische Aufsätze, Eduard Zeller zu seinem 50jährigen Doctorjubiläum gewidmet* (Fues, Leipzig 1887) pp. 17–52
30.80 H. von Helmholtz: Zählen und Messen, erkenntnistheoretisch betrachtet. In: *Wissenschaftliche Abhandlungen*, Vol. 3, ed. by H. von Helmholtz (Barth, Leipzig 1895) pp. 356–391
30.81 H. Ebbinghaus: *Grundzüge der Psychologie*, Vol. 2, 4th edn. (Veit, Leipzig 1919), ed. by K. Bühler
30.82 D. Krantz, R. Luce, P. Suppes, A. Tversky: *Foundations of Measurement*, Vol. 1 (Academic, San Diego 1971)
30.83 D. Krantz, R. Luce, P. Suppes, A. Tversky: *Foundations of Measurement*, Vol. 2 (Academic, San Diego 1989)
30.84 D. Krantz, R. Luce, P. Suppes, A. Tversky: *Foundations of Measurement*, Vol. 3 (Academic, San Diego 1990)
30.85 D. Krantz: From indices to mappings: The representational approach to measurement. In: *Frontiers of Mathematical Psychology. Essays in Honor of Clyde Coombs*, ed. by D. Brown, J. Smith (Springer, New York 1991) pp. 1–52
30.86 D. Krantz, R. Luce, P. Suppes, A. Tversky: *Foundations of Measurement*, Vol. 3: *Representation, Axiomatization and Invariance* (Academic, San Diego 1990)
30.87 S. Stevens: On the theory of scales of measurement, Science **103**, 677–680 (1946)
30.88 S. Stevens: *Psychophysics. Introduction to its Perceptual, Neural and Social Prospects*, ed. by G. Stevens (Wiley, New York 1975) repr. by Transaction Books, New Brunswick 2000
30.89 S. Stevens: A Scale for the measurement of a psychological magnitude: Loudness, Psych. Rev. **43**, 405–416 (1936)

30.90 S. Stevens: Sensation and psychological measurement. In: *Foundations of Psychology*, ed. by E. Boring, H. Langfeld, H. Weld (Wiley, New York 1948) pp. 250–268

30.91 S. Stevens: Mathematics, measurement and psychophysics. In: *Handbook of Experimental Psychology*, ed. by S.S. Stevens (Wiley, New York 1951) pp. 1–49

30.92 S. Stevens: On the psychophysical law, Psych. Rev. **64**, 153–181 (1957)

30.93 S. Stevens: Issues in psychophysical measurement, Psych. Rev. **78**, 426–450 (1971)

30.94 S. Stevens: Perceptual magnitude and its measurement. In: *Handbook of Perception*, Vol. II, ed. by E. Carterette, M. Friedman (Academic, New York 1974) pp. 361–389

30.95 C.W. Savage: *The Measurement of Sensation. A Critique of Perceptual Psychophysics* (Univ. of California Press, Berkeley 1970)

30.96 C. Lesche: On psychophysical measurement, Sven. Tidskr. Musikforsk. **53**, 91–106 (1971)

30.97 D. Laming: *The Measurement of Sensation* (Oxford Univ. Press, Oxford 1997)

30.98 D. Benson: *Music: A Mathematical Offering* (Cambridge Univ. Press, Cambridge 2006)

30.99 H.F. Cohen: *Quantifying Music. The Science of Music at the First Stage of the Scientific Revolution, 1580–1650* (Reidel, Dordrecht, Boston 1984)

30.100 W. Kristof: *Untersuchungen zur Theorie Psychologischen Messens* (A. Hain, Meisenheim am Glan 1969)

30.101 G. Gigerenzer: *Messung und Modellbildung in der Psychologie* (Reinhardt, München 1981)

30.102 N. Cliff: What is and isn't measurement. In: *Statistical and Methodological Issues in Psychology*, ed. by G. Keren (Erlbaum, Hillsdale 1982) pp. 3–38

30.103 F. Sixtl: *Meßmethoden der Psychologie. Theoretische Grundlagen und Probleme*, 2nd edn. (Beltz, Basel 1982)

30.104 S. Stevens, E. Galanter: Ratio scales and category scales for a dozen perceptual continua, J. Exp. Psych. **54**, 377–411 (1957)

30.105 J.C. Falmagne: *Elements of Psychophysical Theory* (Oxford Univ. Press, New York 1985)

30.106 R. Carnap: *Physikalische Begriffsbildung* (G. Braun, Karlsruhe 1966)

30.107 J. Bortz, C. Schuster: *Statistik für Sozial- und Humanwissenschaftler*, 7th edn. (Springer, Berlin 2010)

30.108 W.A. Wagenaar: Stevens vs. Fechner: A plea for dismissal of the case, Acta Psychol. **39**, 225–235 (1975)

30.109 E. Zwicker: *Psychoakustik* (Springer, Berlin 1982)

30.110 R. Luce: *Sound and Hearing. A Conceptual Introduction* (Erlbaum, Hillsdale 1993)

30.111 G.T. Fechner: *Revision der Hauptpunkte der Psychophysik* (Breitkopf Haertel, Leipzig 1882)

30.112 J. Michell: *Measurement in Psychology. Critical History of a Methodological Concept* (Cambridge Univ. Press, Cambridge 1999)

30.113 S. Stevens, J. Volkmann: The Relation of pitch to frequency: A revised scale, Am. J. Psych. **53**, 331–353 (1940)

30.114 S. Stevens: The measurement of loudness, J. Acoust. Soc. Am. **27**, 815–829 (1955)

30.115 J. Merkel: Die Abhängigkeit zwischen Reiz und Empfindung, Philos. Studien **4**, 541–594 (1888)

30.116 A. Schneider: *Tonhöhe – Skala – Klang. Akustische, tonometrische und psychoakustische Studien auf vergleichender Grundlage* (Orpheus, Bonn 1997)

30.117 O. Abraham, E.M. von Hornbostel: Zur Psychologie der Tondistanz, Z. Psychol. **98**, 233–249 (1925)

30.118 C. Pratt: Bisection of tonal intervals smaller than an octave, J. Exp. Psych. **6**, 211–222 (1923)

30.119 C. Pratt: Bisection of tonal intervals larger than an octave, J. Exp. Psych. **11**, 11–26 (1928)

30.120 D. Greenwood: The Mel scale's disqualifying bias and a consistency of pitch-difference equisections in 1956 with equal cochlear distances and equal frequency ratios, Hearing Res. **103**, 199–224 (1997)

30.121 R.M. Warren: Measurement of loudness and brightness: Scaling a chimera. In: *Advances in Psychophysics*, ed. by H.G. Geisler, J.M. Zabrodin (Deutscher Verlag der Wiss., Berlin 1976) pp. 215–237

30.122 R.M. Warren: Subjective loudness and its physical correlate, Acustica **37**, 334–346 (1977)

30.123 R. Warren: *Auditory Perception. An Analysis and Synthesis*, 3rd edn. (Cambridge Univ. Press, Cambridge 2008)

30.124 L. Marks, M. Florentine: Measurement of loudness, part I: Methods, problems and pitfalls. In: *Loudness*, ed. by M. Florentine (Springer, New York 2011) pp. 17–56

30.125 P. Morse: *Vibration and Sound* (Acoust. Soc. Am., Woodbury 1991)

30.126 P. Morse, K.U. Ingard: *Theoretical Acoustics* (Princeton Univ. Press, Princeton 1986)

30.127 A. Schneider, A. von Ruschkowski: Techno, decibels, and politics: An empirical study of modern dance music productions, sound pressure levels, and 'loudness perception'. In: *Systematic Musicology: Empirical and Theoretical Studies*, ed. by A. Schneider, A. von Ruschkowski (Lang, Frankfurt 2011) pp. 13–62

30.128 T. Renz, A. Wolf: Versuche über die Unterscheidung differenter Schallstärken. In: *Archiv für physiologische Heilkunde*, Vol. 15 (1856) pp. 185–193

30.129 A. Aberle: *Die Täuschungen in der Wahrnehmung der Entfernung der Tonquellen*, M.D. dissertation, University of Tübingen (H. Laupp, Tübingen 1868)

30.130 S. Weinzierl: Grundlagen. In: *Handbuch der Audiotechnik*, ed. by S. Weinzierl (Springer, Berlin 2008) pp. 1–39

30.131 M. Wien: Ueber die Empfindlichkeit des menschlichen Ohres für Töne verschiedener Höhe, Archiv ges. Physiol. **97**, 1–57 (1903)

30.132 R.L. Wegel: The physical Examination of hearing and binaural aids for the deaf, Proc. Natl. Acad. Sci. **8**, 155–160 (1922)

30.133 H. Fletcher, J.C. Steinberg: The dependence of the loudness of a complex sound upon the energy in the various frequency regions of the sound, Phys. Rev. **24**, 306–317 (1924)

30.134 J.C. Steinberg: The relation between the loudness of a sound and its physical stimulus, Phys. Rev. **26**, 507–523 (1925)

30.135 B.A. Kingsbury: A direct comparison of the loudness of pure tones, Phys. Rev. **29**, 588–600 (1927)

30.136 V.O. Knudsen: The sensitivity of the ear to small differences of intensity and frequency, Phys. Rev. **21**, 84–103 (1923)

30.137 R. Riesz: Differential intensity sensitivity of the ear for pure tones, Phys. Rev. **31**, 867–875 (1928)

30.138 H. Barkhausen: Ein neuer Schallmesser für die Praxis, Z. Tech. Phys. **7**, 599–691 (1926)

30.139 H. Barkhausen, H. Tischner: Die Lautstärke von zusammengesetzten Tönen und Geräuschen, Z. Tech. Phys. **8**, 215–221 (1927)

30.140 L.F. Richardson, J.S. Ross: Loudness and telephone current, J. Gen. Psychol. **3**, 288–306 (1930)

30.141 L.B. Ham, J.S. Parkinson: Loudness and intensity relations, J. Acoust. Soc. Am. **3**, 511–534 (1932)

30.142 H. Fletcher, W. Munson: Loudness, its definition, measurement and calculation, J. Acoust. Soc. Am. **5**, 82–108 (1933)

30.143 A. Schick: *Schallbewertung. Grundlagen der Lärmforschung* (Springer, Berlin 1990)

30.144 E. Boring: *A History of Experimental Psychology*, 2nd edn. (Appleton-Century-Crofts, New York 1950)

30.145 G.L. Hardcastle: S.S. Stevens and the origins of operationism, Phil. Sci. **62**, 404–424 (1995)

30.146 S. Stevens, H. Davis: *Hearing. Its Psychology and Physiology* (Wiley, New York 1938)

30.147 S. Stevens: Perceived levels of noise by Mark VII decibels, J. Acoust. Soc. Am. **51**, 575–601 (1972)

30.148 M. Mathews: What is loudness? In: *Music, Cognition and Computerized Sound an Introduction to Psychoacoustics*, ed. by P. Cook (MIT Press, Cambridge 1999) pp. 73–78

30.149 M. Sader: *Lautheit und Lärm. Gehörpsychologische Fragen der Schall-Intensität* (Hogrefe, Göttingen 1966)

30.150 S.S. Stevens: The direct estimation of sensory magnitudes – loudness, Am. J. Psych. **69**, 1–25 (1956)

30.151 J. Zwislocki: Analysis of some auditory characteristics. In: *Handbook of Mathematical Psychology*, Vol. III, ed. by R. Luce, R. Bush, E. Galanter (Wiley, New York 1965) pp. 1–97

30.152 E. Zwicker, H. Fastl: *Psychoacoustics, Facts and Models*, 2nd edn. (Springer, Berlin 1999)

30.153 B. Moore: *Introduction to the Psychology of Hearing*, 5th edn. (Academic, Amsterdam 2007)

30.154 E. de Boer: Mechanics of the cochlea: Modeling efforts. In: *The Cochlea*, ed. by P. Dallos, A. Popper, R. Fay (Springer, New York 1996) pp. 258–317

30.155 R. Patuzzi: Cochlear micromechanics and macromechanics. In: *The Cochlea*, ed. by P. Dallos, A. Popper, R. Fay (Springer, New York 1996) pp. 186–257

30.156 M.J. Epstein: Correlates of loudness. In: *Loudness*, ed. by M. Florentine (Springer, New York 2011) pp. 89–107

30.157 M. Chatterjee, J. Zwislocki: Cochlear mechanics of frequency and intensity coding. II. Dynamic range and the code for loudness, Hearing Res. **124**, 170–181 (1998)

30.158 I. Sigalovsky, J. Melcher: Effects of sound level on fMRI activation in human brainstem, thalamic and cortical centers, Hearing Res. **215**, 62–76 (2006)

30.159 D. Langers, W. Backes, P. van Dijk: Brain activation in relation to sound intensity and loudness. In: *Hearing – From Sensory Processing to Perception*, ed. by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, J. Verhey (Springer, Berlin 2007) pp. 227–234

30.160 M. Röhl, B. Kollmeier, St Uppenkamp: Spectral loudness summation takes place in the primary auditory cortex, Hum. Brain Mapp. **32**, 1483–1496 (2011)

30.161 L. Marks: *Sensory Processes. The New Psychophysics* (Academic, New York 1974)

30.162 R. Luce: On the possible psychophysical laws, Psych. Rev. **66**, 81–95 (1959)

30.163 A. von Ruschkowski: *Lautheit von Musik. Eine Empirische Untersuchung zum Einfluss von Organismusvariablem auf die Lautstärkewahrnehmung von Musik*, Ph.D. Thesis (Univ. of Hamburg, Systematic Musicology 2013), Electronic resource available at http://www.sub.uni-hamburg.de

30.164 M.H. Birnbaum, R. Elmasian: Loudness "ratios" and "differences" involve the same psychophysical operation, Percept. Psychophys. **22**, 383–391 (1977)

30.165 W.S. Torgerson: Distances and ratios in psychophysical scaling, Acta Psychol. **19**, 201–205 (1961)

30.166 E.B. Newman: The validity of the just noticeable difference as a unit of psychological magnitude, Trans. Kansas Acad. Sci. **36**, 172–175 (1933)

30.167 M. Florentine, M. Epstein: To honor Stevens and to repeal his law (for the auditory system). In: *Fechner Day 2006*, ed. by D. Kornbrot, M. Msetfi, A.W. MacRae (Univ. Hertfordshire Press, Hatfield 2006) pp. 37–41

30.168 W. Jesteadt, L. Seibold: Loudness in the laboratory, part I: Steady-state sounds. In: *Loudness*, ed. by M. Florentine (Springer, New York 2011) pp. 109–144

30.169 J. Marozeau: Models of loudness. In: *Loudness*, ed. by M. Florentine (Springer, New York 2011) pp. 261–284

30.170 S. Stevens: The volume and intensity of tones, Am. J. Psych. **46**, 397–408 (1934)

30.171 S. Stevens: The attributes of tones, Proc. Natl. Acad. Sci. **20**, 457–459 (1934)

30.172 W.R. Garner: *The Processing of Information and Structure* (Erlbaum, Potomac 1974)

30.173 J. Grau, D.N. Kemler: The distinction between integral and separable dimensions: Evidence for the integrality of pitch and loudness, J. Exp. Psy-

chol.: Gen. **117**, 347–370 (1988)
30.174 L. Marks: On cross-modal similarity: the perceptual structure of pitch, loudness and brightness, J. Exp. Psychol.: Hum. Percept. Perform. **15**, 586–602 (1989)
30.175 N.H. Anderson: Integration psychophysics and cognition. In: *Psychophysical Approaches to Cognition*, ed. by D. Algom (North Holland, Amsterdam 1991) pp. 13–113
30.176 D. Green, J. Swets: *Signal Detection Theory and Psychophysics* (Wiley, New York 1966)
30.177 R. Luce: Thurstone and sensory scaling: Then and now, Psych. Rev. **101**, 271–277 (1994)
30.178 J. Zwislocki: *Sensory Neuroscience. Four Laws of Psychophysics* (Springer, New York 2009)
30.179 H. Fletcher: A space-time pattern theory of hearing, J. Acoust. Soc. Am. **1**, 311–343 (1928)
30.180 S. Stevens, J. Volkmann, E. Newman: A scale for the measurement of the psychological magnitude pitch, J. Acoust. Soc. Am. **8**, 185–190 (1937)
30.181 A. Houtsma: Pitch perception. In: *Hearing*, ed. by B. Moore (Academic, San Diego 1995) pp. 267–295
30.182 H.P. Hesse: *Die Wahrnehmung von Tonhöhe und Klangfarbe als Probleme der Hörtheorie* (Gerig, Köln 1972)
30.183 E. de Boer: On the 'residue' and auditory pitch perception. In: *Handbook of Sensory Physiology*, Vol. 3, ed. by W.D. Keidel, W.D. Neff (Springer, Berlin 1976) pp. 479–583
30.184 R. Plomp: *Aspects of Tone Sensation* (Academic, London 1976)
30.185 A. de Cheveigné: Pitch perception models. In: *Pitch. Neural Coding and Perception*, ed. by C. Plack, J. Oxenham, R. Fay, A. Popper (Springer, New York 2005) pp. 169–233
30.186 C. Stumpf: *Tonpsychologie*, Vol. 1 (Barth, Leipzig 1883)
30.187 L. Beranek: *Acoustic Measurements* (Wiley, New York 1949)
30.188 G. Fant: Analysis and synthesis of speech processes. In: *Manual of phonetics*, ed. by B. Malmberg (North Holland, Amsterdam 1968) pp. 173–177
30.189 J. Beck, W. Shaw: Magnitude estimations of pitch, J. Acoust. Soc. Am. **34**, 92–98 (1962)
30.190 E. Zwicker, G. Flottorp, S. Stevens: Critical bands with loudness summation, J. Acoust. Soc. Am. **29**, 548–557 (1957)
30.191 G. von Békésy: *Experiments in Hearing* (Wiley, New York 1960)
30.192 D. Greenwood: Critical bandwidth and the frequency coordinates of the Basilar membrane, J. Acoust. Soc. Am. **33**, 1344–1356 (1961)
30.193 D. Greenwood: A cochlear frequency-position function for several species – 29 years later, J. Acoust. Soc. Am. **87**, 2592–2605 (1990)
30.194 D. Lewis: Pitch scales, J. Acoust. Soc. Am. **14**, 127 (1942)
30.195 D. Ward: Musical perception. In: *Foundations of Modern Auditory Theory*, Vol. 1, ed. by J. Tobias (Academic, New York 1970)
30.196 F. Attneave, R. Olson: Pitch as a medium: A new approach to psychophysical scaling, Am. J. Psychol. **84**, 147–166 (1971)
30.197 W.J. Dowling: Musical scales and psychophysical scales: Their psychological reality. In: *Cross-Cultural Perspectives on Music*, ed. by R. Falck, T. Rice (Univ. Toronto Press, Toronto 1982) pp. 20–28
30.198 R. Shepard: Pitch perception and measurement. In: *Music, Cognition, and Computerized Sound. An Introduction to Psychoacoustics*, ed. by P. Cook (MIT Press, Cambridge 1999) pp. 149–165
30.199 J. Licklider: Basic correlates of the auditory stimulus. In: *Handbook of Experimental Psychology*, ed. by S.S. Stevens (Wiley, New York 1951) pp. 985–1039
30.200 A. Srinivasan: Speaker identification and verification using vector quantization and mel frequency cepstral coefficients, Res. J. Appl. Sci. Eng. Technol. **4**, 33–40 (2012)
30.201 R. Shepard: Toward a universal law of generalization for psychological science, Science **237**, 1317–1323 (1987)
30.202 F.G. Ashby, N. Perrin: Toward a unified theory of similarity and recognition, Psychol. Rev. **95**, 124–150 (1988)
30.203 N. Perrin: Uniting identification, similarity and preference: General recognition theory. In: *Multidimensional Models of Perception and Cognition*, ed. by F. Ashby (Erlbaum, Hillsdale 1992) pp. 123–145
30.204 K. Norwich: *Information, Sensation and Perception* (Academic, San Diego 1993)
30.205 G. Ryle: *The Concept of Mind* (Hutchinson, New York 1949)
30.206 C. Stumpf: *Erkenntnislehre*, Vol. 1 (Barth, Leipzig 1939)
30.207 E. Husserl: *Erfahrung und Urteil. Untersuchungen zur Genealogie der Logik*, 5th edn. (Meiner, Hamburg 1976), ed. by L. Landgrebe
30.208 E. Wever, Ch Bray: Action currents in the auditory nerve in response to acoustical stimulation, Proc. Natl. Acad. Sci. **16**, 344–350 (1930)
30.209 W.D. Keidel, M. Spreng: Neurophysiological evidence for the Stevens power function in man, J. Acoust. Soc. Am. **38**, 191–195 (1965)
30.210 M. Spreng: Langsame Rindenpotentiale, objektive Audiometrie und Psychoakustik. In: *Physiologie des Gehörs. Akustische Informationsverarbeitung*, ed. by W.D. Keidel (Thieme, Stuttgart 1975) pp. 277–358
30.211 H. von Helmholtz: *Die Lehre von den Tonempfindungen als Physiologische Grundlage für die Theorie der Musik*, 6th edn. (Vieweg, Braunschweig 1913)
30.212 B. Lewis (Ed.): *Bioacoustics. A Comparative Approach* (Academic, London 1983)
30.213 G. Manley, A.N. Popper, R. Fay (Eds.): *Evolution of the Vertebrate Auditory System* (Springer, New York 2004)
30.214 M. Cartmell: *Introduction to Linear, Parametric and Nonlinear Vibrations* (Chapman Hall, London 1990)

30.215 N. Fletcher, T. Rossing: *The Physics of Musical Instruments*, 2nd edn. (Springer, New York 2000)

30.216 N. Fletcher: *Acoustic Systems in Biology* (Oxford Univ. Press, New York 1992)

30.217 S. Tempelaars: *Signal Processing, Speech and Music* (Swets and Zeitlinger, Lisse 1996)

30.218 N. Wiener: Generalized harmonic analysis, Acta Math. **55**, 117–258 (1930)

30.219 A. Khintchine: Korrelationstheorie der stationären stochastischen Prozesse, Math. Annalen **109**, 604–615 (1934)

30.220 N. Wiener: *Extrapolation, Intrapolation and Smoothing of Stationary Time Series* (MIT Press, Cambridge 1949)

30.221 N. Wiener: *Cybernetics or Control and Communication in the Animal and in the Machine*, 2nd edn. (MIT Press, Cambridge 1961)

30.222 S.L. Marple: *Digital Spectral Analysis with Applications* (Prentice Hall, Englewood Cliffs 1987)

30.223 T. Tolonen, M. Karjalainen: A computationally efficient multipitch analysis model, IEEE Trans. Speech Audio Process. **8**, 708–716 (2000)

30.224 A. Schneider: Change and continuity in sound analysis: A review of concepts in regard to musical acoustics, music perception and transcription. In: *Sound – Perception – Performance*, ed. by R. Bader (Springer, Berlin, Heidelberg 2013) pp. 71–111

30.225 R. Bader: Additional modes in a Balinese gender plate due to its trapezoid shape. In: *Concepts, Experiments and Fieldwork: Studies in Systematic Musicology and Ethnomusicology*, ed. by R. Bader, C. Neuhaus, U. Morgenstern (Lang, Frankfurt 2010) pp. 95–111

30.226 A. Schneider, D. Müllensiefen: Musikpsychologie in Hamburg. Ein Forschungsbericht, Syst. Musikwiss. – Systematic Musicol. **7**, 59–89 (2000)

30.227 A. Schneider, M. Leman: Sonological and psychoacoustic characteristics of carillon bells. In: *The Quality of Bells, Proc. of the 16th Meeting of the FWO Res. Soc. on Foundations of Music Research*, ed. by M. Leman (IPEM, Univ. Ghent, Ghent 2002)

30.228 R. Bader: *Nonlinearities and Synchronization in Musical Acoustics and Music Psychology* (Springer, Berlin 2013)

30.229 S. Dixon: Onset detection revisited. In: *Proc. 9th Intern. Conf. on Digital Audio Effects (DAFx-06), Montreal* (2006) pp. 133–137

30.230 P. Boersma: Accurate short-term analysis of the fundamental frequency and the harmonic-to-noise ratio of a sampled sound. In: *Proc. Inst. phonetic Sci. Univ. Amsterdam*, Vol. 17 (1993) pp. 97–110

30.231 B. Lau, R. Bader, A. Schneider, P. Wriggers: Finite element transient calculation of a bell struck by its clapper. In: *Concepts, Experiments and Fieldwork: Studies in Systematic Musicology and Ethnomusicology*, ed. by R. Bader, C. Neuhaus, U. Morgenstern (Lang, Frankfurt 2010) pp. 137–156

30.232 M. Möser: *Technische Akustik*, 8th edn. (Springer, Berlin 2009)

31. Pitch and Pitch Perception

Albrecht Schneider

This chapter addresses sensation and perception of pitch mainly from a functional perspective. Anatomical and physiological facts concerning the auditory pathway are provided to the extent necessary to understand excitation processes resulting from sound energy in the middle ear as well as within the cochlea. Place coding and temporal coding of sound features is viewed in regard to frequency and period as two parameters relevant for pitch perception. The Wiener–Khintchine theorem is taken as a basis to explain the correspondence between temporal periodicity and spectral harmonicity as two principles fundamental to perception of pitch and timbre. The basics of some models of the auditory periphery suited to extracting pitch from complex sounds either in the time or in the frequency domain will be outlined along with examples demonstrating how such models work for certain sounds. Sections of this chapter also address tone height and tonal quality as components of pitch as well as the rather dubious nature of the so-called tone chroma. Issues such as isolating tone quality from height (as in *Shepard tones*) and an alleged preference of subjects for stretched octaves are covered in a critical assessment. A subchapter on psychophysics includes just-noticeable difference (JND) and difference limen (DL) for pitch, the concept of auditory filters known as critical bands, the sensation of roughness and dissonance as well as special pitch phenomena (the *residue* and the missing fundamental, the concept of virtual pitch, combination tones). Another section covers spectral fusion, Stumpf's concept of *Verschmelzung*, and the sensation of consonance. Further, there are sections on categorical pitch perception as well as on absolute and relative pitch followed by a brief survey of scale types, tone systems and intonation. The chapter closes with a section on geometric pitch models and some basic features of tonality in music.

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_31

31.1 Pitch as Elementary Sensation and as Perceptual Quality

The English-American term *pitch* that is used, most of all, in physics and engineering as well as in psychoacoustics and music has many meanings. A major aspect of *pitch* semantics is to denote a spatial relation such as a distance, an interval, a rise or inclination. In regard to certain types of sound (such as described in Sect. 30.2 above), sensory processing triggered by the acoustic input leads to perception of a quality that is labeled *pitch* in English, *hauteur du son* in French, *Tonhöhe* in German, *altura del tono* in Spanish, *Wysota tóna* in Russian, and *yingāo* in Chinese. The terms used in French, German, Spanish, Russian (and many more languages) are equivalent in that they all express a particular sound quality in regard to a spatial dimension ranging from low to high (in some cultures, the direction rather is from high to low, but still referring to a dimension imagined as vertical in space). It seems that, at least for a range of natural sounds found in the environment (such as animal calls, birdsong, speech) human listeners almost involuntarily tend to categorize such signals with respect to qualities pertaining to *pitch* as well as to *timbre*. Pitch hereby includes at least two different spatial and also two different qualitative aspects. In regard to spatial ordering, subjects usually can distinguish various sounds as they appear to have a different height relative to a dimension low–high. Of two sounds, one can be judged to be higher or lower than the other, or both may appear equal in height (some may appear indiscriminable in this respect due to ambiguous sound structures; Sect. 31.1.2). If sounds appear different in height this must not imply that they also have a definite pitch since such differences can be evoked, for example by bands of filtered noise distributed around different center frequencies. If the noise bandwidth is wide enough, two or more such stimuli presented one after another give rise to sensations of only a relative difference in *height* that, typically, goes along with a relative difference in *brightness* (these two attributes are inextricably connected in normal sensation and can be dissolved only by some manipulation in the lab; Sect. 31.1.2). However, one can narrow the bandwidth to such an extent that each sound attains an almost *tonal* quality so that one can even play simple melodies making use of narrow-band noise with different center frequencies [31.1]. Already *Külpe* [31.2, pp. 108 ff.] argued that, besides a sensation of *tone height* resulting from periodic tones, a corresponding sensation of *noise height* can be experienced. Narrowing such noise bands further and further in the end would leave just one spectral component per sound (typically, the center frequency), and playing the same simple melody (say, *Frère Jacques*) then would result in sensations of definite pitches. *Definite* in this respect means the tones now occupy a very small segment or even a precise spot on the dimension low–high, which can be defined in an appropriate physical parameter (like log frequency). The shift from relative distances as produced by a sequence of noise bands to tones each of which are defined (in the most simple case) by exactly one frequency position also activates a peculiar perceptual quality, namely that of distinct intervals between several tones. Perceiving a sequence of tones in regard to pitch involves an estimate of their relationships, which in part are spatial (in terms of distances and intervals between them and/or a *contour* such tones establish in a pitch/time space) and in part have to do with *affinity* and *similarity* (Sects. 31.7, 31.8, 31.9, 32.2). *Houtsma* [31.3], exploring pitch perception with a range of different sounds (pure tones, harmonic *residue* complexes presented either diotically or dichotically, amplitude-modulated broadband noise, clicks), concluded that noise bands can evoke sensations suited to discriminate them on a dimension low–high in an ordinal way, whereas genuine pitch perception is restricted to those sounds that realize intervals as ratios that can be correctly identified. To be sure, noise signals that can be discriminated as to their relative *height* when presented successively may still suggest kind of a rough scale (up or down) or a melodic contour though lacking definite, identifiable pitch ratios [31.3]. Hence, pitch perception in a strict sense depends on sound signals that are quasiperiodic and comprise a more or less harmonic spectrum. As has been confirmed in a range of experiments, resolvable spectral components are essential in pitch perception (apparently predominating temporal information [31.4]). In regard to temporal processing of pitch that is extracted from complex sounds, one has to take reaction time and integration constants of the auditory system into account [31.5, 6], [31.7, Chap. 5]. From data concerning the time needed to establish a stable pitch sensation for pure tones, factors that seem of influence are the period duration and frequency as well as the sound level. Since the periods relevant for music can range from 33 ms (at 30 Hz) to 0.067 ms (at 15 kHz), the time needed to reliably detect a pitch (where an estimate of the pitch obtained from a sample of subjects reaches a percentage of correct answers clearly beyond chance) should be roughly proportional to the stimulus period and frequency. Some early investigations found 60−100 ms for very low frequencies (50 Hz), about 30 ms for 300 Hz and ≈ 15 ms for the frequency range from 0.5−5 kHz [31.8]. An average value of 66 ms for arbitrary pitch estimates has been suggested more recently [31.9, p. 205], with

possibly less time (20–40 ms) needed if subjects have information beforehand about the likely pitch range of the stimuli. For complex tones with only resolvable harmonics a time window of 20 ms has been suggested [31.10] while integration time is probably extended for complex sounds comprising only unresolved higher harmonics. From published data one can assume that about 5–8 periods of a periodic sound stimulus are sufficient for a stable pitch sensation.

The most basic stimulus known to elicit a perception of a distinct pitch is the sine tone. It can be taken as the reference for pitch perception [31.11, Chap. 11.1.2] since a sine tone of medium intensity, at least in the midrange from about 200 to 1500 Hz, usually evokes a single pitch, which is regarded as both salient and unambiguous. The main reason for this effect is that such a sine tone constant in frequency stimulates in a precise way a very small section of the basilar membrane (BM) and leads to a neural response pattern in the auditory nerve (AN) that reflects both the frequency (place code) and the period (time code) of the acoustic stimulus (Sects. 31.3 and 31.4). In a frequency range well suited to human listeners (say, between 250–1000 Hz), a sine tone presented at a level just 3–6 dB above threshold already evokes a clear pitch. It has been observed in various experiments that subjects usually can match a sine tone in frequency to a test sound such as a harmonic complex tone with but little variance if the fundamental of the complex tone is in the range of many sounds known from speech and music (that is, between about 200 and 800 Hz), and the sound level is also convenient (40–60 dB). Also, many subjects are able to sing or hum a sound at, or close to, the same audible pitch after listening first to a sine tone of, say, 440 Hz, presented for 2 s at a sound pressure level (SPL) of 50 or 60 dB by means of loudspeakers to their ears at a distance of 1–3 m. *Pitch strength* as measured by magnitude estimation increases for sine tones at medium frequencies with SPL (from 20 to 80 dB) as well as with duration (up to 300 ms [31.12, Chap. 5.7]). However, a sine tone is hardly a natural stimulus while many sounds we hear in the environment (from birdsong to human speech and, of course, music) have harmonic spectra comprising several or many partials. It has been found in some mammalian species that the DL for harmonic complex tones was lower than for sine tones presented at the same frequency [31.13, p. 63]. If a harmonic complex tone is presented to human listeners in a range where f_1 is between $\approx$ 130 and 800 Hz (roughly C_3–G_5), there are several partials besides f_1 that can be *resolved* by the BM–CP filter bank (Sect. 31.3). Pitch information can be derived in the frequency domain from f_1 as well as from such partials that directly reinforce f_1 (the octaves at f_2 and f_4). Other resolvable partials also contribute to the perceived pitch of harmonic complexes [31.4]. Moreover, the periodicity of such a signal as encoded in the complex waveshape (Figs. 30.10 and 31.6a) provides another cue. Therefore, a harmonic complex with partials phase-locked (as sine or cosine components) and with amplitudes decreasing with harmonic number offers several cues (or clues) for perceptual determination of pitch. There are auditory models that derive pitch estimates either from matching partials to a harmonic template or from periodicity analysis (Sect. 31.5).

In general, sounds with harmonic spectra are stronger or more *salient* in regard to pitch perception than are sounds with inharmonic spectra; sounds comprising continuous spectra such as noise bands usually give rise to weak or ambiguous sensations of pitch. Pitch sensations can also be evoked from amplitude modulation (AM) of noise bands and even from repetitions of a sound reflected from hard surfaces with a period of several ms between reflections indicating that periodicity is an important factor. However, a sensation of pitch can be elicited also from sharp edges in spectra in broadband signals [31.4, 14] as well as from a range of other phenomena (briefly reviewed in [31.15]). From a large number of experiments and observations it is clear that there are both spectral and temporal cues relevant for sensation and perception of pitch (for overviews see [31.9, 11, 16]).

31.1.1 Pitch as Dependent on Frequency and Period: A Brief Review

Historically, the concept of *pitch* as connected to that of *frequency* dates back to the late 16th and the first half of the 17th century (with contributions notably from both Vincenzo and Galileo Galilei, Isaac Beeckman, and Marin Mersenne [31.17, 18]). Mersenne came close to calculating the absolute frequency of a vibrating string according to

$$f = K\frac{1}{l}\sqrt{\frac{P}{\rho}}\,,$$

where K is a constant, l is the string length, P is the tension and ρ is the density of the string. Mersenne succeeded in finding, in a fair approximation to correct numerical values, fundamental frequencies for both strings and organ pipes. He and some of his contemporaries also noticed that the sound of a vibrating string of given length contains harmonic partials (an issue investigated in detail by Sauveur around 1700). In early acoustical theory, vibration of strings was modeled after the pendulum, whereby each very small segment Δl of a string can be viewed to perform the same motion as

that of a pendulum [31.18]. Taking a string segment Δl as a point mass accelerated by some force, its displacement from equilibrium over time describes a sinusoidal trajectory per period of vibration. Insofar as pitch is considered a function of the frequency of vibration, it is also a function of the number of vibration cycles conceived as periods completed per time unit since $f = 1/T$ and $T = 1/f$. Knowing that the basic pattern of vibration in strings is sinusoidal, one could surmise that sensation of pitch is dependent on sinusoidal vibrations. However, there were also experiments carried out by Robert Hooke (around 1681) and much later by Félix Savart where they found that a pitch sensation can be evoked from a periodic process such as pressing a strip of brass or card against the teeth of a revolving toothed brass wheel [31.19, p. 42]. This was also an elegant method to determine the absolute frequency (which equals the number of teeth on the wheel times the number of revolutions of the wheel per second [31.18, p. 32]). By about 1820, Charles Cagniard de la Tour invented the siren consisting of rotating discs perforated with equally spaced holes [31.20, p. 30]. Such a siren produces regular sequences of puffs of air suited to evoking a pitch sensation. *August Seebeck* [31.21] followed these experiments, using a siren as sound generator. In a number of conditions he found that a periodic time signal composed of isochronous pulses is sufficient to produce a salient pitch. *Seebeck* [31.22, p. 453] concluded that a (steady) tone results from the repetition of patterns of vibration whereby the duration of the period defines the *height* (German: *Höhe*) of the tone. Seebeck was severely criticized by *Ohm* [31.23] who claimed that, for a steady tone to become audible and to have a certain *pitch*, a vibration of the form $A \sin 2\pi(mt + p)$ would be necessary either in a direct form (i. e., a sine wave of frequency mt as stimulus) or as a *real component* that can be *precipitated* from a regular sequence of impulses (the German text has *Eindrücke* for impulses). Though the Seebeck–Ohm dispute is fairly well known [31.16, 24–26] and several of Seebeck's experiments have been explained as well as repeated by *Schouten* [31.24], it is still worth recalling some of Seebeck's observations and the arguments that followed from both sides. *Seebeck* [31.22, 27] reported in particular that for the condition where the holes on the rotating disc of the siren were spaced equidistantly, there was one clear tone (a *Hauptton* heard as a salient pitch) plus a number of harmonic partials (addressed as either *Beitöne* or *Obertöne*), which he perceived as *very weak*. This statement is repeated several times [31.22, p. 454, 456, 463], [31.27, p. 362]. *Ohm* did not deny that a siren could produce a series of impulses; he [31.23, p. 518] insisted that such impulses had to follow one another at intervals of the length $1/m$, and that p had to be constant for each pair of successive intervals to allow a steady tone to become audible. Ohm's argument in part addressed the ratio of the duration of each impulse relative to the pulse period (the pulse duty factor τ/T, where τ is the duration of the pulse and T is the pulse period). By assuming τ to cover almost T, Ohm imagined a sequence with broad pulses that could have resulted in one *real* sinusoidal at frequency $(mt + p)$ plus a few weak *Beitöne* (which Ohm admitted could be included in the complex sound). *Seebeck* [31.22, pp. 455–57] rejected Ohm's interpretation as unrealistic since, under the condition of equidistant holes fed with air by one pipe, τ had to be much smaller than T, and certainly like $\tau < 1/2\,T$. *Seebeck* [31.27] has stressed that Ohm's concept, which relates pitch to a fundamental frequency f_1 strong enough to be heard as a single *tone*, was just the narrow formulation of a much wider principle; like Seebeck, Ohm of course saw the reciprocity of frequency and period, to conclude that the repetition of a vibration at a period of $T = 1/f$ is the fundamental condition for steady tones to occur. However, while Ohm demanded a single strong sinusoidal vibration $A \sin 2\pi(mt + p)$ or a complex reducible to such a pattern to be present in each period, *Seebeck* [31.22, 27] stressed the importance of the period $T = 1/f$ as such, and argued that periods could be filled with various waveshapes of the general form $f(\cos 2\pi mt)$, that is, a superposition of cosine terms. He [31.22, pp. 474 ff.] went even further when arguing that, on the basis of series expansion applied to building complexes from series of cosine terms, complex tones might have a f_1 partial with a rather weak amplitude, a_1, so that this component $(a_1 f_1)$ could not be the sole source of the clearly audible tone (with a pitch sensation corresponding to f_1). Rather, the tone sensation might result from the combined effect of successive partials ($a_2 f_2$, $a_3 f_3$, $a_4 f_4$, etc.) because these partials, when *taken together*, *result in the same period* as that covered by $a_1 f_1$. Thereby, a weak $a_1 f_1$ component would be reinforced (in hearing) by a group of successive higher partials. Seebeck even saw the possibility (from his mathematical considerations rather than from his experience) that the first terms in a series of harmonic partials might be so weak in amplitude that they could not contribute in any significant way to a complex tone while a number of higher partials *combined* could express the motion pattern equivalent to $A \cos 2\pi(mt + p)$. Finally, *Seebeck* [31.22, p. 479] speculated that several partials of the form $a_s \cos 2\pi(smt + p)$ could produce a tone (in sensation, a pitch at f_1) even *if no term* of the form $a \cos 2\pi(mt + p)$ *is present*. This of course is the case of the *missing fundamental*. Hence, the important factor Seebeck brought into play was that various waveshapes, representing different spectra, can share

the same period, and that their repetition at $f = 1/T$ can give rise to a very similar or even identical sensation of *tone height* (*Höhe des Tones*). In this respect, Seebeck advanced a concept later known as *repetition pitch*, *periodicity pitch*, or *low pitch* (including the pitch of complex harmonic sounds with a *missing fundamental* [31.16, 24, 25]).

Helmholtz [31.28, Chap. 4] advanced the concept that our ears will sense only a *pendulum-like* (i. e., sinusoidal) vibration as a simple tone, and will dissolve complex sounds into a series of sinusoidals that are sensed as a series of tones. This concept (following Fourier and Ohm, respectively) must be viewed in regard to Helmholtz's model of the BM as a chain of tuned resonators; from anatomical data he assumed 600 *radial fiber* resonators to be available for each of the seven musically relevant octaves, which makes 50 for each semitone [31.28, pp. 241–242]. A mechanical resonator of course will be excited if a spectral component carries energy at the resonance frequency. Hence, it was natural for Helmholtz to postulate that sensation of a simple tone is effected by a sinusoidal whose frequency matches that of a resonator. Harmonic partials thereby would engage different resonators, and would give rise to respective sensations, which were combined into one percept representing a complex tone (e.g., the note G_4 played on a violin) on a higher neural level. According to Helmholtz, we become aware of only the percept (while sensations may trigger *subconscious inferences*, see Sect. 30.1.1).

With the Ohm–Seebeck dispute and Helmholtz's BM resonator model, the stage was set for pitch theories that had a focus either on the ear as frequency analyzer [31.29–31] or on the hearing system viewed as performing periodicity detection [31.24, 25, 32–35]. To be sure, both aspects have been combined in various ways to account for actual observations obtained from both physiological investigations and psychophysical experiments (for overviews see [31.9, 16, 36]). Also, both frequency analysis and periodicity detection have been implemented into a range of auditory models. Essential features of such models will be discussed in Sect. 31.5 (below) following a sketch of the auditory pathway (AuP) in which some functional characteristics relevant for pitch processing are outlined (Sects. 31.2 and 31.4).

31.1.2 Tone *Height* and Tone *Quality* as Components of Pitch and the Dubious Nature of *Tone Chroma*

In the previous sections, pitch sensations were considered to be largely dependent on either frequency, or period, or both. The characteristic of pitch to be the perceptual correlate of simple or complex tones that differ in either f_1 or f_0 refers to one dimension usually termed *tone height* (see above) that has two poles at *low* and *high*. Using a tone generator that outputs a sine wave into an amplifier and headphones (of suitable quality), one can easily explore both the range of audible frequencies and the smallest frequency differences that can be detected as pitch JNDs. However, when sweeping in a quasilinear fashion through a wide frequency range, one will notice changes in *brightness*, which increases with rising and decreases with falling frequency. Evidently, the sensory quality experienced as brightness covaries with frequency as well as with *tone height*. Hence, in a first approximation, both *tone height* and *brightness* can be viewed as a function of frequency. For this reason, it has been argued that sequences of pure tones especially in high frequency regions elicit sensations of different degrees of brightness rather than evoking *pitches* (the latter imply assignment of tones to locations sufficiently *distinct* to permit an estimate of intervals between them). The *mel scale* (see above, Sect. 30.1.4), which did not consider the perceptually relevant *octave equivalence* of pairs of pure or complex tones that conform to a frequency ratio of 2 : 1 on the one hand, and stretched into high frequency regions far beyond musical register on the other, has thus been classified as a *brightness scale* [31.26, p. 158].

A factor most relevant for music perception is that most subjects report strong *similarities* between pitches evoked when frequencies have doubled or halved in rising or falling direction respectively. Two tones presented at frequency ratios of 2 : 1, 4 : 1, or 0.5 : 1 between their fundamentals are perceived as equivalent not in height but in their quality; a tone an octave above or below the reference tone appears as a repetition in regard to quality [31.37]. *Rameau* [31.38, p. 16], in discussing the nature of complex harmonic tones, had already remarked that we know from experience that the octave is nothing but the replica of a certain tone, and that we easily can confuse two tones an octave apart (*je sçavois que l'Octave n'est qu'une replique, combien il y a d'identité entre les sons et leurs répliques, et combien il est facile de prendre l'un pour l'autre, ces sons même se confondant à l'oreille.*). The effect is usually addressed as *octave equivalence* whereby a musical interval spanning eight tones (whole steps) is taken to express the perceptual similarity. However, besides musical experience there are facts from acoustics and neurophysiology suited to address the phenomenon.

In regard to acoustical features, two sine tones of frequencies 200 and 400 Hz and equal amplitude, when presented simultaneously and with synchronous onset, match perfectly as two periods of the higher tone equal

one period of the lower tone; together, they result in a simple symmetric waveshape that repeats steadily at a period of 5 ms. If the same octave is formed from two harmonic complex tones (each having six partials with amplitudes $1/n$), the waveshape in Fig. 31.1 results.

It is easy to see that the main period is at 5 ms and a second periodicity is established at half of that value. Subjects may hear two tones clearly separated by an octave; however, with the 5 ms period much stronger in amplitude, the complex with f_1 at 200 Hz is prevalent. *Helmholtz* [31.28, Chap. 10] advanced an explanation for *octave equivalence* in which he emphasized two harmonic complex tones like those represented in Fig. 31.1 have partials that coincide with their respective frequencies, and no partials to interfere with each other:

200 400 600 800 1000 1200
400 800 1200 1600 2000 2400

Hence, several pairs of partials reinforce each other while roughness of such a two-tone interval is negligible in regard to *critical bands* (CB, Sect. 31.6.2 below). Also, one can consider the partials from both tones as belonging to one harmonic series starting at $f_1 = 200$ Hz. This would explain the high degree of uniformity that octaves such as depicted in Fig. 31.1 elicit in perception. It has been reported already by *Stumpf* [31.39, 40] that subjects lacking musical training (and not used to analytical listening) tend to take two tones presented at an octave as one.

Besides acoustical cues for what *Helmholtz* [31.28] labeled *Klangverwandtschaft* (affinity of complex tones due to common harmonic partials), there are such from neurophysiology. As is well known, pure tones and complex tones evoke neural discharge patterns that, in single AN fibers, can be measured with microelectrodes and usually are represented as either a poststimulus time (PST) histogram or as an interspike interval histogram (ISIH). In such diagrams (abscissa: time in ms, ordinate: rate of intervals measured) one sees peaks not only at the period of the pure tone but also at multiples [31.41]. In addition, ISIH often exhibit a slight dispersion around spike periods (ms) corresponding to the period of the acoustic stimulus (and its multiples). A possible explanation is the probabilistic nature of neural response patterns (including spontaneous activity and relative as well as absolute refractory period of neurons); in a single AN fiber, spikes must not occur at every period of the stimulus and also phase-locking of discharges is not always as strict as one may expect from the stimulus (e.g., steep wave crests at the onset of each period as in Fig. 31.1). In regard to ISIH being composed of multiples of the period corresponding to the stimulus, two pure tones an octave apart can give rise to neural responses that are quite similar. The phenomenon that two tones an octave apart appear similar and consonant has been attributed to common stimulus and neural periodicities already by *von Hornbostel* [31.42, p. 713] from observations on binaural (dichotic) hearing. More recently, experiments involving dichotic fusion indicated evidence for internal octave templates [31.43]. For complex tones, consonant intervals show few competing periodicities and have clear peaks in ISIH obtained from AN fibers whereas dissonances show many competing periodicities [31.44–46]. The most coherent periodicity pattern is that of the octave.

The factors referred to will be less efficient or even missing when two tones an octave apart are presented in succession. If two pure tones are one or even two

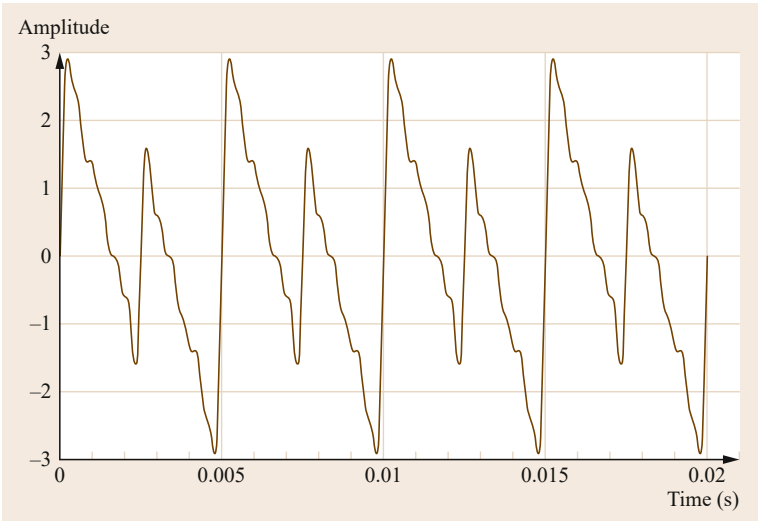


Fig. 31.1 Octave 200 : 400 Hz formed of complex tones (six partials each, $A = 1/n$)

octaves apart, there of course is no coincidence of partials, yet still most subjects with some musical experience will identify the interval correctly by making use of echoic and short-term memory as well as interval categories stored in long-term memory (LTM) (Sect. 31.7). Even in a condition where subjects listen to a slow sine sweep starting at a certain frequency (say, 100 Hz), they will take the starting point as an *anchor* and will register whenever the sweep reaches octaves at 200, 400, 800 Hz, and probably also at 1600 and 3200 Hz, respectively. At frequencies above ≈ 4 kHz, interval recognition deteriorates due to neurophysiological limits of periodicity coding (Sect 31.4) though some well-trained musicians manage to identify intervals at low error rates even above that limit [31.47]. However, the effect of increasing error rates for tones and intervals presented above 4 kHz has been reported for subjects with so-called *absolute pitch* (AP [31.48] and below, Sect. 31.7).

Perceptual octave equivalence is a basic condition for converting the dimension low–high from relative distances (sensations differ with respect to *tone height* and *brightness*) to musical scales and intervals. If tones that are an octave apart are perceived as having the same *quality*, this constitutes a cycle that repeats every octave. Theoretically, the number of elements within (or rather on) the cycle could be infinite. *Brentano* [31.49], pondering the problem of gradation of *tonal qualities*, considered the JND as a possible unit of discretization of the octave. In regard to the very small JND for frequency differences in the middle octaves, he dismissed these *simple elements* in favor of *saturated elements* (his reasoning was by drawing parallels to vision and color perception, see below), which can be identified as the tones in a musical scale. The point of course is that *tonal qualities* must not only be discriminable (as are JND) but identifiable as such to become meaningful units in regard to perception as well as musical composition and performance. Hence the number of *saturated elements* within one octave must be considerably smaller than the number of JNDs that fall into the same interval. The number of tones per octave used in music and the number of *pitches* relevant for performance as well as listening differs among cultures. In *Western* music theory, the number of tones usable per octave according to notation (taking sharps, flats, double sharps, double flats into account as they occur in many works) is 31−35 [31.50, pp. 183 ff.]. In systems of just intonation (JI), the number of tones may be as high as 53 [31.51] or even 171 per octave [31.52] in order to implement a huge set of diatonic, chromatic, and enharmonic tonal relations with deviations no greater than the DL for pitch in humans. However, the number of pitch categories per octave that can be identified in perception and cognition is much smaller than the number of pitches that can be distinguished both in performance and listening. For example, a shift from a-flat to g-sharp in a chord progression A^b major–E major in JI is well audible, yet it is in doubt whether average music listeners could also identify the two different tones (by naming them while listening) or could explain which pitches/tones have been exchanged in the two chords. Recognizable tonal relations (scale steps, intervals, chords, other sonorities) in part depend on acoustic and psychoacoustic features (offering cues for perception) and in part are internalized, in each musical subject, in a process of learning and practice. In many textbooks for elementary music education and music psychology, the number of so-called *pitch classes* per octave relevant for practice is given as $n = 12$ (by analogy to the number of keys available on a standard *Western* keyboard), which seems a superficial view not only in regard to professional musicians engaged in, for example, madrigal or barbershop singing, or performing in string or saxophone quartets, but also because of confusing concepts. The notion of *pitch class* is applicable to, and has been used in, mathematical models of scales based on set theory where a reduction from n pitches to m *pitch classes* ($m \ll n$) is feasible [31.53]. The notion of a pitch category, though, should be reserved for perceptual and cognitive issues as are relevant for *categorical perception* (Sect. 31.7).

Tones identifiable as pitches within an octave may have a distinct tonal quality; a minimum requirement is that such tones are periodic (or nearly so) and last for enough time to be sensed. Tones having a *tone quality* with their pitch are the basis for perceiving scale steps and intervals, and in turn the function of each tone is largely defined by structural relations within a scale as well as by their *contact* to adjacent tones (for a detailed discussion see [31.54, 55]). The tonal quality of tones identifiable as musical pitches within an octave (and recurring in each octave) has been distinguished from tone height by *Geza Révész* [31.37, 56], who also experimented with separating *tonal quality* from *tone height*. In elaborating on a concept known as *two-componential theory of pitch*, Révész distinguished two principles of similarity relevant for the perception of tones, one being the proximity of tones differing only slightly in frequency and in pitch (as sensed by subjects), the other the similarity of *quality* between tones an octave apart. Contradicting earlier approaches, which had held that similarity in pitch decreases linearly as the frequency distance between two tones increases (implying that the fifth above a given tone would be more *dissimilar* than a minor and a major second), Révész argued that frequency distance between tones alone cannot be a decisive factor for judging their similarity. Obviously,

that distance is small for two tones that differ by only a JND in pitch and still appear quite *similar* in this respect, while the distance is large for two tones an octave apart where similarity again is very high. However, it is a different kind of similarity than that based on distance estimates. *Lotze* [31.57, p. 212], among many scientists, had pointed to the singular experience of *octave equivalence*; he described hearing two tones an octave apart as *an undefined fusion of sameness and raise*. *Lotze* [31.58, p. 465] also assigned the same *tone character* to tones an octave apart. In line with these qualifications, *Révész* [31.37, p. 15 f.] regarded *quality* as a phenomenal characteristic that tones an octave apart have in common (in this respect, all notes with the same tone name but located in different registers have the same *quality*). In his 1913 treatise, he presumes that a scale of tones exists (as he used his piano for some experiments) that repeats within each octave. A more precise yet somewhat different statement is found in *Révész* [31.56, p. 67]: *The 'tone qualities' of the chromatic and any other scale form a recurrent series the members of which are complete within one octave*. Evidently, there is a shift here of *quality* from tones an octave apart to 12 chromatic scale tones. He elaborates further that *interval quality* and *tone quality* are phenomenal attributes *that do not appear isolated since their existence is based on the musical tone series* [31.56, p. 85]. In this respect, the two attributes become distinguishable in perception because musical scales constitute a series of sounds where different tone qualities go along with different degrees of *tone height* (and brightness). In a strict sense, *tone quality* without *height* is an abstraction. *Révész* [31.37, p. 131 f.] argued that, from a phenomenological point of view, one can conceive of a series of pure qualities as well as of a series of pure *heights*. He concedes that, in reality, only a musical scale provides both components in an audible and perceivable form. *Révész* [31.37] took the musical scale (the European chromatic scale in ET12) as given. A separate issue is how such *tone qualities* come into existence, in ontogeny of subjects as well as in a historical and cultural perspective.

Stumpf [31.59], recognizing the merits of Révész's two-componential concept of pitch but seeing also a host of problems involved with this approach, called the 12 chromatic tones of the scale now common in European music *historical qualities*. He [31.59, p. 323] distinguished these from what he labeled *arch qualities* (*Urqualitäten*), the fundamental property of tones within one octave to be sensed as different in pitch. Stumpf agreed to Brentano's view that small increments in pitch (the size of a JND) constitute a change in quality (and not just in *brightness*). He [31.59, pp. 317 ff.] argued the attribute of *tone quality* is lost, as observed in experiments conducted by, in particular, his doctoral student, *Catharina von Maltzew* [31.60], more and more above C_8 (≈ 4186 Hz) while we still sense relative differences in *tone height* (conveyed as *brightness*). Since in this range perception of *qualities* (conceived as discrete, identifiable pitches) deteriorates, so does perception of intervals. This is in line with more previous observations concerning the upper limits of pitch perception as based on periodicity ([31.61] and Sects. 31.4, 31.5).

In his report on current research, *Stumpf* [31.59] addressed the role not only of *tone height*, *brightness*, and *quality*, but also of attributes such as vocality (that had been investigated, in a range of experiments, by Stumpf's doctoral student, *Wolfgang Köhler* [31.62, 63]). *Rich* [31.64] then tried to determine the DLs for those attributes including *quality* (which is called *tonality* in his work). The goal was to manipulate *tonality* independent of *tone height* (which is *pitch* in Rich's study) and brightness, and to let subjects make their respective judgments. The limens given for several subjects, among them the music psychologist, H.P. Weld, differ considerably. Taking Weld's data for limens of *tonality* relative to reference tones at frequencies of 275, 550 and 1100 Hz, respectively, they were 1.44, 2.75, 3.85 Hz and are larger than the limens for *pitch* (tone height) and brightness of the same subject (0.36, 0.41, 0.67 Hz and 0.42, 0.33, 0.74 Hz, respectively). Evidently, the judgments of *tone height* are almost identical with those of brightness. Expressed in cents, the limen for *tonality* still would be just a few cents. In a set of experiments expanding those of Rich, *Straub* [31.65] did seek to determine the JND for *tone height* independently from *tone quality*, finding that the JND for *quality* was about three times the JND for *height*. For a range of frequencies (most reliably 256–480 Hz) the JND for *quality* was ≈ 25–30 cent; however, the design of these experiments in regard to *quality* came close to exploring the range of small pitch variants which, by observers, are assigned to the same scale step (the distributions of judgments Straub obtained for upper and lower limits of single scale steps around a center are quite similar to such from experiments on *categorical perception*, Sect. 31.7). *Hornbostel* [31.42], who labeled the specific quality of tones in a scale *Tonigkeit* (equivalent to *tonality* in *Rich* [31.64]; however, the term *tonality* in English as well as *Tonalität* in German include meanings beyond *tone quality*), explained the emergence of the quality phenomenon thus: If we hear two pure tones in succession that differ slightly in frequency, they appear as equal in regard to *Tonigkeit*. If the frequency of the second tone is changed further, in small increments, there is a certain point where the two tones no longer appear as equal in quality

yet as markedly different. *Hornbostel* [31.42, p. 712] thus described the transition from one *pitch category* to the next. He reported (in line with [31.59, 60]) that *Tonigkeit* was vanishing, even for musicians, in very low and very high registers as well as with tones of very short duration.

Bachem [31.48] labeled the qualitative component of pitch *tone chroma*, thereby making reference to *Brentano* [31.49] who had addressed the structure of tonal organization in hearing by drawing parallels to vision and colors respectively. Taking up some current ideas concerning perception of colors and brightness, *Brentano* [31.49] argued that, within each octave, a series of *saturated elements* returns while the difference from one octave to the next is one of, expressed in visual terms, the ratio of *dark* to *bright* (both are conceived as being combined in all tones audible for humans). In effect, the *dark* component is strong for low register (bass range) and the *bright* for high (treble range). According to Brentano, this unsaturated, continuously variable element of dark and bright (relative darkness and brightness) fuses or intermingles with those *saturated elements* in every octave; going up a scale, the saturated elements retain their specific quality while they can be distinguished from one octave to the next with a decrease of the *dark* and an increase of the *bright* ingredient. In a nutshell, what Brentano proposed was the two-componential theory of pitch (expanded by *Révész* [31.37] who acknowledged the contribution made by Brentano) that comprises the steadily rising or falling *tone height* (confounded with the sensation of tonal brightness) and that series of *tone qualities* forming *saturated elements* comparable to, as Brentano believed, basic colors. The question, then, is how close parallels between vision and hearing could be drawn. As we all know, two colors can fuse or blend to become not just inseparable, but to result in a third color. Two tones forming a harmonic interval may *fuse* (Sect. 31.6.4), but do not change jointly so that the originals disappear and a new tone is created instead. Also, the phenomenon of octave equivalence seems to be restricted to audition, and has not been reported for other modalities. Assuming there is a recurrent cycle of saturated tonal elements that can be identified in every octave, one would have to ask what the nature of saturation is, in particular if these elements are perceived in isolation (as in experiments on so-called *absolute* pitch, AP; see *Miyazaki* [31.66] and Sect. 31.7). There seems to be a salience in the perception of basic chromatic colors (red, green, yellow, blue, orange, purple, brown, pink [31.67]) that relates to the structure of photoreceptors and physiological processes pertaining to the spectral analysis of light sources (for basic mechanisms of vision, see chapters on sensation and perception in [31.68, 69]). In hearing, possible cues for determination of a specific *tone chroma* within the range of an octave would be frequency and period of pure tones as well as the relative brightness conveyed with an *absolute* frequency; for a series of relatively homogeneous complex sounds (i. e., those which have the same spectral pattern and envelope) dispersed over the range of an octave, the spectral centroid (Sect. 30.2) changes accordingly.

The parallels Brentano drew from vision to hearing may seem obvious in regard to brightness, which in fact is an intermodal quality [31.70]. Concerning those saturated elements in terms of a *chroma* series of tones, things are more complex. Historically, Greek music theory already had the term *chroai* ($\chi\rho\omega\alpha\iota$ = colorings), however, this refers to small tuning differences within the stable frame of tetrachords [31.71, pp. 33 f.] and, thus, not to saturated but to *fluid* elements that could be recognized as deviating in pitch from the positions marked by a diatonic or chromatic scale (there are similar phenomena to this day in Near Eastern and Indian music traditions). Also, in modern music theory *chromatic* tones in a scale are often regarded as mere variants or *alterations* of the stable diatonic scale steps [31.55]; the tones of a seven-note diatonic scale would thus count as *saturated elements* of a stable structure that, moreover, can be derived by a progression in pure fifths [31.72]. Of course, one can extend this progression to 12 or even more tones (as in medieval and modern treatises on music theory from Near Eastern traditions [31.73] and Sect. 31.8) which, however, result in a chain of fifths (and not necessarily in a cycle that repeats in each octave). The relatedness of tones by fifths like f–c–g–d–a–e–b is experienced by subjects in listening to music and making music in many music cultures; it is a phenomenon based on acoustic facts and supported by auditory processing characteristics (Sects. 31.4–31.8). In this respect, one might associate the *tonal quality* of particular scale steps and tones with their position and function within a structure defined by basic intervals, most of all, the pure fifth (and its complementary interval, the pure fourth) as the most fundamental consonant interval besides the octave [31.72].

Observations on octave equivalence and tonal quality based on interval relations reflect basic musical experience such as melodies retaining their contour when transposed [31.74, Chap. 5]. *Révész* [31.37, 56] argued that, in addition to melodic constancy, a melody rendered first on one pitch of the scale and then transposed by an octave appears as a *repetition*. There is evidence that some primates, besides humans, are able to recognize simple melodies transposed by an octave as identical [31.75]. Different from the *mel scale*,

a scale based on tone qualities can be conceived as similar to *a solfeggio reference structure, the whole of which is freely transposable (below* 5000 Hz*) in a log frequency medium* [31.76, p. 158]. The reference structure is stored in LTM, however, there are some indications also for a cortical basis of tone quality perception including cyclicality of the *chroma* series as represented within or near the primary auditory cortex (A1) regions [31.77]; see Sect. 31.4.

Typically, tone quality and tone height as well as tonal brightness are combined in perception of musical tones (e.g., tones played on a trumpet or piano). *Révész* [31.37] reported a few *pathological cases* where quality and height had been perceived disjointedly. He [31.56, p. 82 ff.] also suggested that the 12 tone qualities per octave could be created, in a pure form, by playing all the Cs in the audible range (in principle, this would be C_1–C_9) simultaneously, and with equal intensity, then all the C#s, all the Ds, and so on. This, he proposed, would also demonstrate the cyclic nature of those 12 *qualities* (quasi-independent of *height*), at least in ET12. Years later, *Roger Shepard* [31.78, 79] experimented with sound sequences generated on computers where *chroma* was changing seemingly independent from *tone height*, thereby producing a construct known as *the endlessly rising scale* (that has been viewed to constitute an auditory parallel to the illusionary staircases of Maurits Escher). Extending the concept in certain ways, the composer *Jean-Claude Risset* [31.80] produced continuous shifts of both *chroma* and *height* in opposite direction (such sounds might allegorize atomic bombs falling towards their target). Without going much into details of *Shepherd*'s and *Risset*'s constructs [31.78–82], the decisive factor for a *chroma scale* to appear rising *endlessly* is keeping tone height and brightness (as expressed by the spectral centroid, see Sect. 30.2) on a constant level while the 12 *Shepherd tones* consisting of octave-spaced components are repeated over and over again in a chromatic ET12 scale. This is achieved simply by using a bell-shaped (Gaussian) spectral envelope on a log frequency scale; the spectral components – a number of sinusoidals spaced in octaves – are then moved stepwise (from one *chroma* to the next) or slided continuously upwards or downwards under the envelope. The distribution of spectral energy thereby does not change significantly with respect to the centroid [31.81]. To produce the effect, in principle any symmetric envelope with a peak at the centroid will do (a Gaussian envelope works smoothly). In some of Risset's sounds, the *endlessly rising scale* is turned into a quasicontinuous glide [31.83] and in addition the envelope itself is shifted up or down the log frequency scale so that the centroid frequency changes. By running the *endless* glide in one direction (say, upwards) and slowly shifting the centroid of the envelope to the other, a peculiar sound process can be experienced that may be viewed as *illusionary* in that *chroma*, *tone height* and brightness, which normally vary jointly in a rising or falling scale, have been manipulated separately, to some degree.

Finally, an issue known as the *stretched octave* should be briefly discussed. From spectral measurements of thick piano strings (bass strings consisting of a carbon steel core wound with copper wire) it was observed that the frequencies of partials do not exactly match harmonic ratios but are slightly inharmonic due to the stiffness of such strings. The deviation Δf from harmonic ratios ($f_n = nf_1$) increases with harmonic number; the effect is significant yet not dramatic since partial amplitudes in general decrease with harmonic number, meaning higher partials have less weight in the harmonic complex. For example, in a grand piano G_1 string we measured the frequency for $f_{10} = 10.047f_1$, $f_{20} = 20.23f_1$, $f_{30} = 30.76f_1$ ($f_1 =$ 47.425 Hz [31.84]). Piano tuners compensate for the inharmonicity of partials by stretching octaves slightly (as well as by balancing each key within an octave with respect to pitch and timbre). A similar process is observed in the high treble range where the strings are very short.

From observations such as piano tuning, at times a general tendency for stretched octaves in intonation practice of musicians as well as a preference for stretched octaves on the side of listeners has been concluded. Some of the data published from experiments indicate some enlargement of *subjective octaves* (the ratio of fundamental frequencies is somewhat greater than 2 : 1); for example, *Hesse* [31.85] found that musically trained listeners preferred slightly enlarged octaves for successively presented complex tones, with an average (median) size of 1200 + 7 cents. However, in some measurements also octaves with a slightly compressed frequency ratio were found ($f_p/f_q < 2:1$ [31.86, 87] and [31.88, Fig. 1]). One study conducted on intonation practice of eight professional violinists playing the C-major scale over three octaves (C_4–C_7) upwards and downwards showed practically perfect octaves (in upwards motion, the octaves were slightly too large, downwards, a little too small [31.89]). In another study [31.90–92], octaves (tones A_4–A_5) recorded from violinists where correct at 2 : 1 when they used the open A-string while it was slightly too large when they played tones (B_4–B_5) with finger stops, indicating that the sound of the open A_4 was still audible when the violinists played the A_5 and matched the harmonics of the higher tone to those of the lower. Different from these studies, which offered a basic musical setting, most of the earlier lab experiments are difficult

to assess in regard to reliability and validity because the stimuli were presented monaurally per earphones (a condition unsuited for precise pitch judgments) and the sample size often was small with only a few subjects taking part. In experiments carried out simply to check the alleged general tendency for enlargement of *subjective octaves*, this author has used a demonstration provided by *Houtsma* et al. [31.83] (demonstration 15: octave matching) in which (spoken commentary on CD) *a* 500 Hz *tone alternates with a stepwise increasing comparison tone near* 1000 Hz. *Which step seems to represent a correct octave?* The steps are in 5 Hz from 985 to 1035 Hz. Though steps of 5 Hz near 1000 Hz seem a bit coarse in regard to JND and DL data (rather indicating steps of 2 Hz as appropriate), the experimental design seems sufficient for the purpose of controlling the assumption that *stretched octaves* are what subjects prefer. The pairs of tones with which *octaves* of $500 : \{985 \ldots 1035\,\text{Hz}\}$ are realized were presented to samples of subjects (students with either a *major* or a *minor* in systematic musicology) in binaural–diotic condition with audio studio equipment in a room prepared for sound recordings at a convenient SPL (≈ 65 dB). Statistics for four samples (summing to $n = 115$ subjects [31.93, pp. 482–484]) yield a median of 1000 Hz (frequency ratio 2 : 1, 1200 cent) and an aM of 999.043 Hz with SD = 4.4 Hz. The data do not support the hypothesis that subjects *in general* would prefer an enlarged or stretched octave. This issue perhaps needs further investigation in which musically relevant stimuli and conditions should be provided.

31.2 Sketch of the Auditory Pathway (AuP)

From a comparative perspective, auditory processing in mammals has common anatomical, physiological and functional bases [31.94–98]. One has to bear in mind that most of the findings reported on the structure and function of the cochlea and of the AuP up to the cortex were obtained from animal experiments (where animals usually figure as *the chinchilla model*, *the guinea pig model*, etc.). Since the anatomy and physiology as well as specific functions of the auditory system have been reviewed in great detail in a number of volumes of the Springer Handbook of Auditory Research (of which 50 volumes were edited and published from 1992 to 2014) as well as in other publications [31.98–101], a brief summary of facts and hypotheses relevant for sensation and perception of sounds will suffice at this place (see also the review in [31.102]). The auditory system comprises the outer and the middle ear, the cochlea (with the organ of Corti including inner hair cells (IHC) and outer hair cells (OHC) resting on the BM), the spiral ganglion leading to the AN (branch of the VIIIth cranial nerve), and the consecutive relays (nuclei) of the ascending AuP from the cochlear nucleus (CN: nucleus cochlearis ventralis (VCN) and nucleus cochlearis dorsalis (DCN)) to the superior olivary complex (SOC), the N. Lemniscus lateralis (NLL), the inferior colliculus (IC), the medial geniculate body (corpus geniculatum mediale) (CGM), and then to the auditory cortex (AI, AII and adjacent belt regions). A sketch of the ascending AuP can be drawn involving only the main relays and connections (Fig. 31.2).

The afferent AuP is complemented by efferent pathways such as the olivocochlear bundle (OCB [31.98, Chap. 8]), which controls OHC motility and is critical for pitch perception (see below). It should be clear that the AuP has a highly complex system structure incorporating ipsilateral and contralateral communication channels, sequential and parallel processing chains, as well as feedback regulation circuits. In addition, connections of the auditory cortices to other cortical areas, notably in the frontal lobe, must be considered. Apparently, in primates there are dual streams of processing leading from caudal and rostral auditory cortex to prefrontal areas where one circuit may be specialized in object analysis (the *what* channel) and the other in spatial analysis (the *where* channel needed for sound localization; [31.103, 104]). The degree of complexity of the main relays of the AuP such as the IC and the CGM is astounding (for a detailed description see [31.98, 105, 106]).

In addition to Fig. 31.2, a flow chart of the auditory system indicating main stations and functions in regard to processing the sound input can be outlined (Fig.31.3).

Functions as defined in this chart relate to the diagram sketched in Fig. 30.1 above where, after mechanical (middle ear) and then hydrodynamic excitation (cochlea, BM) a transduction stage (IHC, with involvement of OHC) is followed by feature extraction along the AuP, probably up to the level of the IC where, from the evidence available, integration of features seems to begin, which is continued at the level of the CGM. The CGM closely intertwines with the cortex (there are many afferent and efferent projections); because of the high degree of thalamocortical connections [31.107], CGM and AI are often viewed as one large functional module.

The term *auditory object* that appears in the scheme on the cortical level (AI) even in a neurocognitive per-

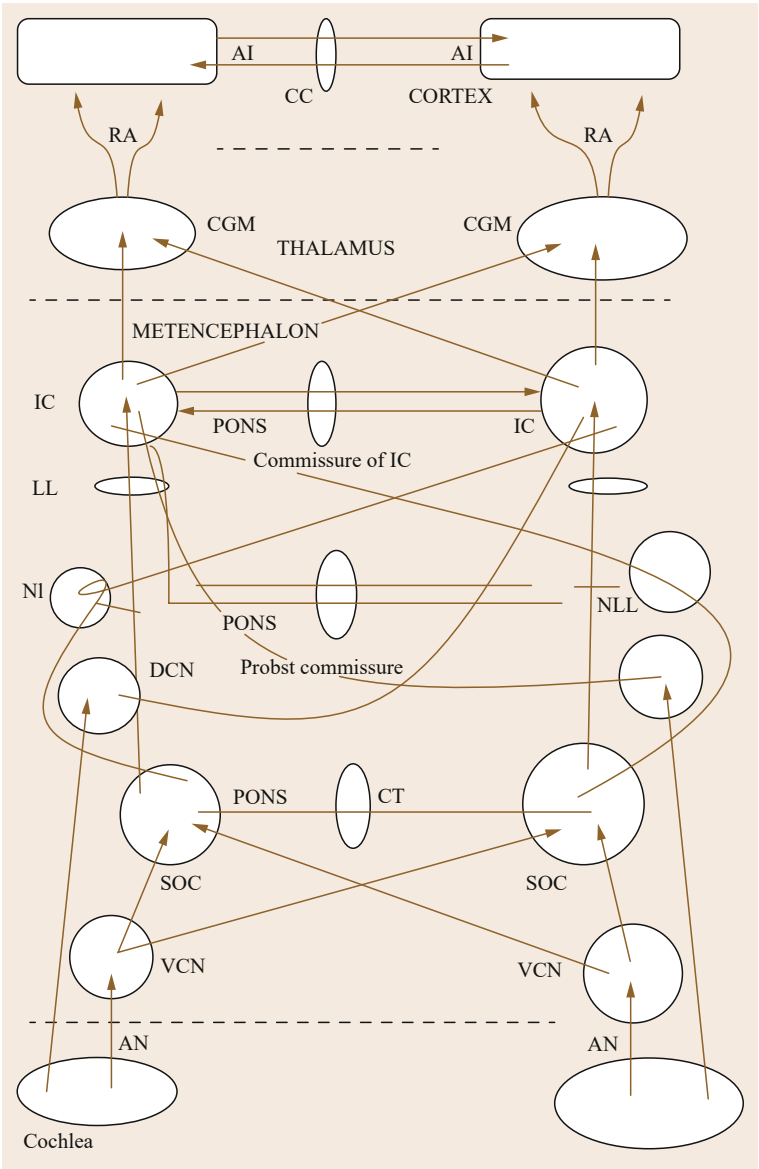


Fig. 31.2 AuP: main stations and connections. Abbreviations: AN = nervus acusticus, VCN = nucleus cochlearis ventralis, DCN = nucleus cochlearis dorsalis, CT = corpus trapezoideum, SOC = nucleus olivia metenceph., NLL = nuclei lemniscus lateralis, LL = laterial lemniscus, IC = colliculus inferior, CGM = corpus geniculatum mediale, RA = radiatio acustica, AI = cortex auditivus (primus), CC = corpus callosum, Nl = nucleus lateralis (superior olivary complex)

spective can have several meanings [31.108]. While sensing an *object* may apply, first of all, to the visual or haptic modality because of the visible or tangible shape of material objects, the term covers auditory phenomena as well. For instance, sensing a harmonic complex tone such as produced from a stringed instrument results from binding components together that may have been extracted from complex sounds in the inner ear and along stations of the ascending AuP. Binding in this case can be based on spectral and temporal cues (harmonicity of partials, common period, concurrent onsets, common envelope) and results in a coherent object that can be perceived as a distinct element (against a background of noise, or in a stream of other such tones). In a more general approach, a fundamental condition for binding features into objects as well as several objects into higher formations (such as several complex harmonic tones into a chord) can be traced to oscillations in the brain (such as the well-known brain waves δ, θ, α, β, γ), which serve to *screen* brain states and to *update* or *refresh* the system in duty cycles. In addition, there seem to be specific oscillations in neural structures of the AuP that might play a role in detecting periodicities in auditory signals relevant for the sensation of pitch and modulation processes. As has been elaborated in detail recently [31.109], various oscilla-

Sound Source(s) in the Environment
↓
Pinna collection of sound energy
↓
Meatus transport of sound energy
↓ impedance matching, linear filter: resonance
Tympanum, ossicles 'conditioning of sound energy'
↓ impedance and gain adaptation
Cochlea: BM frequency dispersion, spectrum analysis
↓ tonotopic frequency representation
Cochlea: IHC mechanoelectrical transduction, low pass filter, halfwave rectifier
↓ coding of signals into neural impulse patterns
Ganglion spirale (AN) spike trains distributed over ca. 30.000 fibers per ear
↓ → connects also to contralateral SOC
CN tonotopic organization, feature detection
↓ → connects to contralateral nucleus
SOC efferent reflex, control of OHC, feature detection
↓ → connects also to NLL and contralateral IC; connects to formatio reticularis
LL/NLL
↓
IC feature detection, spatial hearing, motor responses
↓ → connects to contralateral IC (also connects directly to AI)
CGM feature detection, pattern recognition, integration (1)
↓ → connects to amygdala, limbic system
AI integration (2), formation of 'auditory objects'
→ connects to contralateral AI and to various brain areas (in particular, PFC)

Fig. 31.3 Flow chart of the auditory system (main stations and functions)

tions (which, however, relate to a basic time constant of 0.4 ms that can be attributed to an average synaptic delay) combined into a model of auditory neural processing could account not only for binding of features into objects but also for a correlational analysis suited to detect pitch and sensory consonance.

31.3 Excitation of the Auditory System: From the Tympanum to the BM, the IHC and OHC

Since sound signals including various types of music arrive at the ear of mammals as small disturbances of the normal atmospheric pressure (Sect. 30.3), information relevant to the auditory system is contained in the changes of pressure as a function of time and place. Leaving aside the outer ear channel (which provides a linear filter with a resonance peak around 3.3 kHz; [31.11, pp. 229–230]), the eardrum is the place where sound waves arrive. The tympanum (eardrum) is a membrane of ≈ 0.1 mm thickness and ≈ 8.5 mm diameter covering an area of about 60 mm^2. Elongation of the tympanum for very low tones of 20 Hz just audible is about 0.1 mm, and for tones of 3 kHz the elongation necessary to evoke a faint sensation has been given as 10 picometers (10×10^{-12} m; [31.110, p. 97]), which is less than the diameter of air molecules set to motion.

The main function of the middle ear is to adapt waves from the low impedance of air to the high impedance of cochlear fluid. This is achieved by significantly increasing the pressure in the middle-ear transfer chain. Since the area of the tympanum in humans is about 60 mm^2, and the area of the stapes

footplate working on the oval window (OW, the entrance for sound waves to the cochlea) is $\approx 3-3.2\,\mathrm{mm}^2$, the ratio is about 18 : 1, and the ratio of the two ossicles (malleus, incus) acting as levers is (depending on how the effective length is defined) about 1.5 : 1 up to 2 : 1. The product thus would be in the range from 27–36. Also, the tympanum as a concave membrane increases the pressure delivered to the ossicles and the OW by a factor of two. The transformer ratio in total thus is in the range of $\approx 70:1$ to $80:1$ ([31.111, p. 467] has 80.5 : 1, [31.98, p. 18] has 82.5 : 1, and [31.101, p. 94] has 69.2 : 1, corresponding to 36.8 dB). The middle-ear transfer function and the complex input impedance of the cochlea have been calculated from observations on human temporal bones extracted shortly after death from cadavers [31.112] where the middle-ear sound pressure gain (GME) was 23.5 dB at 1.2 kHz with a slope of $\approx 6\,\mathrm{dB/octave}$ from 0.1 to 1.2 kHz, and about $-6\,\mathrm{dB/octave}$ above 1.2 kHz. The transfer function for ear canal sound pressure to stapes footplate velocity (SVTF) had a maximum at 1 kHz where $0.33\,(\mathrm{mm\,s^{-1}})/\mathrm{Pa}$ is observed. Mean complex impedance Z_c was almost flat between 0.1 and 5 kHz at 21.1 acoustic GΩ MKS. Above 5 kHz, Z_c had a maximum at 6.7 kHz with 49.9 GΩ. Phase of Z_c in the range from ≈ 0.5 to 5 kHz was close to 0°, decreasing to negative values below 0.5 and above 5 kHz. Without the impedance matching in the middle ear, up to 98% of the sound energy would be reflected at the OW because of the high impedance of the cochlea. Stapes movement seems to increase in a linear fashion even for high SPL up to 130 dB [31.98, p. 22], indicating (almost) linear system behavior over a wide dynamic range. However, the lever action of the ossicles has also been regarded as a possible source of some harmonic distortion at higher sound levels, and in particular of difference tones [31.11, p. 231]. Combination tones (sum and difference tones) usually are attributed to cochlear nonlinearities [31.113–115] and [31.98, pp. 147 f.] and below.

The footplate of the stapes working on the OW leads to compression of the fluid in the cochlear duct and to a traveling wave in the scala vestibuli. Patterns of motion in the cochlea and the phenomenon known now as the *traveling wave* were first described by *Georg von Békésy*, in several articles mainly from the 1940s (collected in [31.29, Chaps. 11,12]; for recent research on cochlea functions up to 2000 see [31.115]). Since waves usually travel through a medium at a certain phase speed (dependent on the compressibility and the density of material; in fluids,

$$c = \sqrt{\frac{K}{\rho}}\,,$$

where K is the compression module, and ρ is the density at rest of the medium), there must be something special to this traveling wave, which is to be distinguished from the normal sound wave propagating in a fluid like water. As in fact the viscosity and the density of the cochlear fluids (endolymph, perilymph) are practically identical to water [31.116, 35-3], the phase speed of a sound pressure wave in the cochlea would be expected to be close to 1550 m/s, meaning the distance from the OW to the apex of the human cochlea would be crossed in about 2 µs. Though such a fast wave exists in the cochlea (reflected at the apex/helicotrema to become a standing wave), the *traveling wave* is much slower at about 100 m/s [31.117] and is dispersive in frequency as are bending waves propagating in solids (for a discussion of such analogies, and for fundamental mechanics of the cochlea, see [31.113, 114, 118, 119]). The main reason for the dispersion that has been observed throughout the cochlea is a coupled interaction between fluids in the scalae of the cochlea (for details of cochlea anatomy and physiology see [31.98, 101]) that undergo compression while the BM as a structure is subjected to displacement [31.120]. Dispersion can be attributed in particular to variable longitudinal stiffness of the BM, which decreases from the base to the apex (for a BM length of 35 mm, which is the average from human male and female BM measures) by a ratio of $\approx 100:1$ [31.29]. For the gerbil, a stiffness gradient of 4.43 dB/mm has been reported [31.121]. Since the stiffness decreases along the BM, the compliance increases correspondingly. Dispersion on the BM involves phase delay relative to stapes motion. A number of measurements performed with different methods suggest that the phase delay from the base to the place of a characteristic frequency (CF) in the cochlea is about 1 ms for 10 kHz, 1.5 ms for 5 kHz, ≈ 4 ms for 1 kHz, and close to 6 ms for 0.5 kHz [31.122]. The travel time, however, depends to some extent on SPL; the delay values cited here probably apply for SPL in the range of 50–60 dB. BM displacement depends on species, frequency and SPL. For the chinchilla, displacement at several CF was about 1 nm for a 30 ms tone burst at 30 dB SPL [31.123]. For humans it is of the order of $\approx 3-3.5$ nm at 20 dB SPL, which means ≈ 40 dB cochlear amplifier gain due to OHC activity. From animal experiments (guinea pig) it was estimated that, at a very low SPL of 15 dB, for a CF of 15 kHz the section of the BM that undergoes displacement was just 0.15 mm wide, meaning 14 IHCs and 53 OHCs would be activated [31.124]. In fact, at low excitation levels, the frequency response at the CF equals that of a sharply tuned bandpass, a feature observed also for AN fiber tuning curves.

In some of the early research (reviewed in [31.16, 26]) the BM had been regarded as a chain of mechanical resonators. However, observations on material properties of the BM (which seem to indicate more of a *basilar plate* than a basilar membrane, [31.125, pp. 207 f.]) as well as of phase relations between stapes footplate motion and BM displacement and BM velocity [31.113, 115, 119, 126, 127] are not in line with a simple resonator model. When the traveling wave arrives at the place where its amplitude reaches maximum (indicating maximum pressure and BM displacement respectively), group velocity decreases and the envelope of the traveling wave, viewed instantaneously, attains a characteristic shape as the slope is steep on the side towards the apex, and relatively shallow on the basal side. For a wave from a sine tone of a chosen frequency, the amplitude reaches maximum at a certain place on the BM where the energy contained in the wave dissipates; the wave decays rapidly thereafter. Since the excitation pattern locates a certain frequency at a specific *spot* relative to BM length, such a correlation is termed *tonotopic* or *cochleotopic*. It has been the basis of cochlear maps that have been established for several species [31.128, 129]. Part of the human frequency-place coordinates for the first 20 mm distance from apex is shown in Fig. 31.4 with octaves marked at A_3, A_4, A_5, A_6 {220, 440, 880, 1760 Hz}. To calculate the coordinates, Greenwood's formula $f = A(10^{ax} - k)$ with $A = 165.4$, $a = 0.06$ and $k = 0.87$ has been used.

In a similar vein, distance from apex (expressed as a proportion where 0 = apex and 1 = base near OW/stapes footplate) can be related to log frequency [31.130, p. 222]; the resulting graph (Fig. 31.5) expresses a Fechnerian scaling (Fig. 30.3).

The BM (or, alternatively, the cochlear partition (CP) as the structure between the scala vestibuli and the scala tympani comprising the BM with the organ of Corti, and the scala media up to Reissner's membrane) seems to work, when viewed for single pure tones of various frequencies, as a mechanical low-pass filter with the cutoff frequency decreasing with distance from stapes and OW to the apex/helicotrema. However, for mixtures of complex sounds BM filtering can be understood perhaps more adequately with a model combining a bank of asymmetrical band-pass filters wired in parallel. This is what most models of the auditory periphery based on signal processing realize (for an overview see [31.131]). BM operation, in a mechanical and hydrodynamic approach (as pioneered by von Békésy), can be regarded as *passive* filtering insofar as it was found still operating in the cochleae of human cadavers (where physiological processes bound to metabolism and oxygen supply have more or less ceased). There is evidence, however, that the filter function deteriorates rather soon postmortem so that the quality factor (Q) of tuning/filtering, which relates bandwidth to center or to resonance frequency, is not as good as in vivo condition. Also, there is evidence that the cochlea – a tiny, delicate and highly sensitive organ [31.98, 101] – is extremely vulnerable in a living organism; observations made on (usually anesthetized) animals even under subtle surgical treatment might, therefore, not represent the *true* performance of which cochlear filtering is capable. Moreover, in a living organism, cochlear filtering apparently is supported by OHC motility in what seems to be a feedback loop circuit [31.132] often addressed as the *cochlear amplifier* (current discussion on this issue is summed up in [31.133]). The main function of the OHC (of which humans have about 12 000) thereby seems to amplify the traveling wave in the cochlea, a process particularly useful in regard to low sound levels near threshold. The cochlear amplifier apparently is the main source of cochlear level-dependent nonlinearity that sets in at low SPL (about 30 dB) and shows increasing compression for high SPL (up to ≈ 0.2 dB/dB).

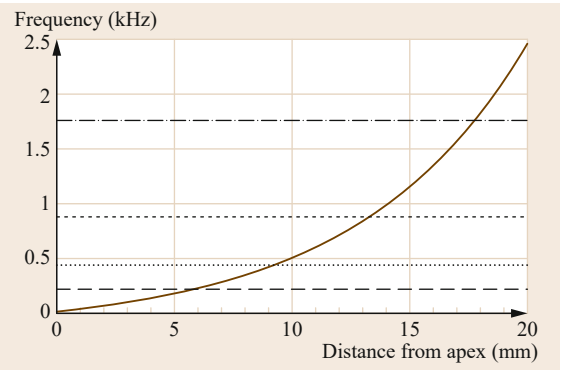


Fig. 31.4 Human BM coordinates: relation of frequency (Hz) to distance (0–20 mm) from apex; four octave markers at 220, 440, 880, 1760 Hz

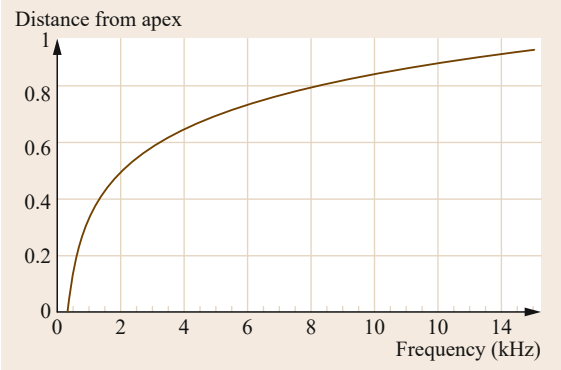


Fig. 31.5 Human BM coordinates: distance from apex (0 = apex, 1 = base)/frequency

Empirical data demonstrate also that *OHC integrity is crucial for sharp tuning of cochlear nerve fibers emanating from the IHC* [31.134, p. 297], confirming the role of the BM-IHC-OHC feedback loop. Results from nonlinear biophysical modeling suggest that the main effect of OHC activity could be undamping of a small area of the BM corresponding to the peak of the traveling wave [31.135]. With a linear (passive) biophysical model of the cochlea (built on cochlear dimensions and fluid dynamics by *Mammano* and *Nobili* [31.136]), the product of time and frequency resolution

$$\Delta t \Delta f \geq \frac{1}{2}$$

known as the *uncertainty principle* (the original formulation by Heisenberg relates to place and momentum in quantum mechanics while a version by Gabor refers also to signal analysis and hearing; [31.137]) was found, for the frequency range above 200 Hz up to 4 kHz, to be close to 0.55, that is, near the theoretical minimum [31.138]. With an active nonlinear biophysical model [31.135], performance still improved, indicating that the BM area elongated by a sine tone under OHC control must be extremely small. Such a mechanism, in which high frequency selectivity seems to be achieved through *lateral* suppression of frequencies in the vicinity of a strong stimulus frequency [31.139], is essential in generating precise information in regard to frequency already in the peripheral sensory organ and transduction process. There are many observations supporting a *cochleotopic* organization of frequency analysis and encoding [31.134]. Also, there are experimental findings according to which correct tonotopic representation of spectral components is essential for pitch perception of complex sounds [31.140]. However, a *peak shift* in excitation dependent on SPL measured in the organ of Corti of animals has been reported [31.141] that might possibly interfere with place theories of frequency coding in the cochlea. According to these observations, increasing the SPL of a stimulating sine tone shifts the peak of excitation on the IHC level to lower (!) frequencies but apparently does *not* affect the apical cutoff on the BM corresponding to a certain frequency. The observation in fact is that [31.126] (in chinchilla), for a given site on the BM corresponding to a CF, the peak responses vary with stimulus level as to frequency. For a CF of 10 kHz, the peak response was obtained for a 10 kHz tone and a level ≤ 50 dB; when the level grew to 60 or 70 dB, the largest response was obtained with an 8 kHz tone. At 80 and 90 dB stimulus level, the largest response resulted from a 7 kHz tone; the shift in maximum sensitivity hence amounts to ≈ 617 cent. The effect, however, cannot be taken as producing a pitch shift in sensation of anything of this size for, in this case, we would experience a significant pitch shift with every change in SPL, and in particular with each crescendo in a concert. As is known from behavioral experiments with humans, influence of SPL on sensation of pitch is slight for pure tones, and even less for complex tones [31.12, p. 114]. However, subjects can experience a pitch shift for sine tones with high SPL applied for a certain time due to fatigue of the BM section that is excited the most (also a temporary or permanent loss in sensitivity may be suffered; Chap. 33).

The transduction process from mechanical excitation to neural spike patterns measured in AN fibers is effected by means of the IHC, of which humans have about 3500 per cochlea. Each hair cell carries about 100 stereocilia [31.142] capable of quite rapid motion, and thus of generating a large number of sensor (or *receptor*) potentials per second. However, IHC intracellular voltage changes in relation to stimulus frequencies decline markedly above 1 kHz (as observed in the guinea pig [31.98, 143]) reflecting cellular membrane low-pass characteristics as well as refractory periods. The distribution of IHC along the BM is somewhat uneven (the number of IHC per mm BM decreases towards the apex, indicating a lesser degree of innervation that could explain reduced frequency sensitivity for $f < 100$ Hz). The AN in humans comprises about 30 000 fibers per cochlea whereby each IHC is connected to several afferent AN neurons. Transduction in IHC from hydromechanical forces into sensor potentials involves mechanical, chemical and electrical processes (reviewed in detail in [31.98, 144]).

31.4 Place Coding and Temporal Coding of Sound Features

On the basis of physiological and psychophysical data, it has been suggested that, in the AN, both time and place information are used for coding pitch [31.36]. In particular, the place information has been assigned to *frequency* and *tone height* (Sect. 31.1.2) while time information, connected with *period*, should account for *low pitch* (including the *missing fundamental*) and *tone chroma*; in effect, pitch processing in this way would occupy two channels [31.145]. In this section, evidence for both principles is reviewed beginning with place.

In the light of the BM-IHC-OHC feedback loop, one can conceive of modules each of which comprises

a very small section of the BM along with the corresponding IHC, OHC as well as the afferent and efferent innervation. In regard to stimulus transfer, each module is selective in that it operates at a characteristic frequency (CF), and also shows maximal output at a so-called best frequency (BF). The basis for sharp tuning apparently lies in BM vibration patterns, which show bandpass characteristics at each CF where signals produce small displacement and velocity at a low threshold [31.126, 127]. In this respect, CF means that a response measurable as output occurs at a low input level; the threshold indicates maximum sensitivity of a module for a certain frequency (or a narrow range of frequencies). The BF is the frequency that produces maximum output of a module, measured as the firing rate of neurons/fibers (above spontaneous discharge rates, and up to saturation) for any given input level. Both CF and BF response are observed at various stations of the afferent AuP up to the cortical level [31.146, 147], indicating a place principle of transduction and neural encoding as well as a transfer function comparable to a bandpass filter. The frequency-threshold curve (FTC [31.143]) of fibers, which represents the neural firing rate for a given input (frequency, level) can be viewed as the inverse of a bandpass with the CF at the minimum. Typically, the FTC for fibers is asymmetric, with a steep rise towards the high frequency side and a soft ascending slope for frequencies below the CF and BF respectively. The relative sharpness of the FTC depends on frequency and level, and is different for units along the AuP. For comparison, FTC are often taken as $Q_{10\,dB}$ (which is the quality or Q-factor of the filter at 10 dB above threshold; $Q_{10\,dB}$ = CF/bandwidth at CF).

The AN of mammals, comprising thousands of fibers, can be viewed as a two-dimensional array, *with the first dimension corresponding to CF (or alternatively, to cochlear place), and the second one to sensitivity or threshold* [31.36, p. 161]. Measurements showed that spectral components of synthesized steady-state vowels can be represented by average discharge rates in a population of fibers with different CF in the AN of the cat [31.148]. However, the spectral envelope of the stimulus fits to discharge rates plotted against the CFs (kHz) only for low SPL. Rate-place representations of stimulus features in the AN [31.149] undergo significant change with level apparently depending on the spontaneous rate (SR) and the threshold of fibers as well as their overall dynamic range and the level where saturation in discharge rate is reached. Since each of the AN fibers seems to have a certain CF and BF that account for a bandpass-like transfer function, a precise evaluation of spectral and dynamic rate-place capabilities of the AN would need synchronous recording of the complete set of fibers engaged in transmitting a broadband signal (such as complex sounds incorporated in speech or music), and possibly signal reconstruction from the neural pattern distributed over many fibers. However, rate-place models on the level of the AN are sufficient for providing neural cues relevant for pitch and timbre perception if such models include *a realistic cochlear frequency map, cochlear filters, suppression, rate-level functions, and threshold distribution* [31.36, p. 204].

As an alternative to rate-place representations, temporal analysis of neural spike patterns in AN fibers and CN units has become a standard method in regard to finding periodicities suited to provide the basis for sensations of various pitch phenomena including pitch of pure and of harmonic complex tones, pitch of harmonic sounds with the fundamental frequency missing, pitches evoked from harmonic and inharmonic AM and frequency modulation (FM) sounds, etc. (for an overview see [31.150]). The beginnings of temporal analysis date back to the 1930s and 1940s when the relation of stimulus frequency to action potentials came into view, which led to hypotheses of neural firing patterns such as the *volley principle* [31.151, 152]. More recent experiments showed that, in the CN, neurons sensitive to onsets fired at highly regular intervals precisely synchronized to the period of a speech signal [31.153], indicating that the periodicity inherent in the stimulus is preserved in neural spike patterns recorded in the AN and the CN.

With the analysis of almost periodic functions set forth by *Wiener* [31.154] and *Khintchine* [31.155], it was possible to study some of the more irregular sound phenomena as well as the periodicities hidden in brain waves. *Licklider* [31.156, 157] expanded the approach to audition by proposing a model of *duplex* pitch perception in which, after initial cochlear filtering of a broadband signal $f(t)$ into a running spectrum, $F(t, x)$, conceived as similar to a spectrogram, a running autocorrelation analysis is performed in each channel on the rectified and smoothed filter output to yield a neural representation in a two-dimensional plane (comprising the length of the BM as x and the lag τ needed for autocorrelation analysis (AC)), which may be fed to higher stations of the AuP. A problem that becomes evident with binaural and dichotic listening is integration and differential processing of temporal information provided from both ears. One model that became quite influential is known as *coincidence detection* applied to neural spike trains [31.158]. The formal operation needed for comparison of input from two channels in regard to estimating the degree of coherence as well as phase effects is cross-correlation function (CCF) [31.159, Chap. 14]. *Licklider* [31.160] formulated a *triplex* theory of pitch perception, tak-

ing into account dichotic phenomena such as the so-called Huggins' pitch (based on interaural phase transition [31.161]) and spatial sound localization where, in addition to autocorrelation function (ACF), CCF is needed to compute binaural interaction with signals fed into both ears. The *triplex* model would represent the three levels of analysis (cochlear place coding of frequencies, temporal coding of periodicities, and information from binaural processing) in a neural architecture of three dimensions.

The *duplex* model developed by Licklider served as the blueprint for a range of computational models to follow [31.131, 162–166]; one part of the neural autocorrelator that became an issue later of course was the delay line needed in particular for low-frequency components of a broadband signal. Given that the lower limit of musical pitch is close to 30 Hz [31.167, 168], the period thus is 33.33 ms, and the delay line necessary to provide the lag of the signal should be of the same order. Though speed of transmission and processing of neural spike trains decreases from AN along the AuP up to cortical AI and belt regions, peaks of the seven prominent brainstem auditory evoked potential (AEP) responses (which are not dependent on the subject's arousal or vigilance) occur within 1–9 ms after stimulus onset [31.169], indicating there is but little *delay* in stations below the thalamus and the cortex. In order to avoid a delay line not supported by anatomical and physiological data, alternative models have been proposed that can still perform the correlational analysis in the time domain [31.109].

Autocorrelation analysis (which draws on the concept of self-similarity with respect to the time dimension) has been applied to neural spike data in regard to various pitch phenomena. Such studies covered, for example, response patterns recorded from the AN and ventral CN for signals comprising either the fourth and fifth harmonic (of a fundamental frequency f_1 that itself is not present) or their odd variants ($4.5f_1$, $5.5f_1$). Autocorrelation analysis of period histograms found peaks representing the period of the input waveform for the harmonic signal and the actual period of the inharmonic signal [31.170]. Extended experiments on AN fibers of the cat resulted in neural spike patterns for a range of stimuli (pure tone, single-formant vowel, click trains, harmonic complexes, AM tone, AM noise) suited to evoke various types of pitch [31.44, 45]. Spike trains, which provide efficient temporal coding [31.171] for natural and artificial sounds can be analyzed according to the time interval between successive spikes (first-order interspike intervals) as well as in regard to successive and nonsuccessive spikes (*all-order* interspike intervals). Analysis of all-order interspike interval histograms (ISIH) is formally equivalent to an autocorrelation analysis. Such histograms were summed up for a population of fibers, and the pitch for each stimulus was estimated from pooled autocorrelation histograms. In addition, the salience of pitch was computed. As can be expected, periodicity inherent in periodic stimuli turns up in the analyses, and is a robust feature not affected by level in a range from 40–80 dB [31.44]. With the analysis based on ISIH, the pitch salience of harmonic complexes was found to be higher than that of a pure tone. This seems plausible because a harmonic complex comprising a series of partials including f_1 offers more clues for pitch extraction than does a pure tone. In particular, temporal analysis of spike patterns accounts for the pitch associated with harmonic complexes lacking f_1 (the *missing fundamental*). In regard to the acoustic input, the period of a harmonic complex does not change when f_1 is missing or even if several low partials are removed (Fig. 31.6a,b).

The harmonic complex shown in Fig. 31.6a comprises eight partials ($f_1 = 100$ Hz) of equal amplitude, superposed as sine tones locked in phase. In Fig. 31.6b, $f_1 = 100$ Hz and $f_2 = 200$ Hz have been removed. As a result, the waveform changes and the amplitude peak at the onset of each period is smaller due to reduced spectral energy; removing f_1 and also f_2 (as in Fig. 31.6b) from a harmonic complex affects the timbre as well as the loudness level perceived, and to some extent also pitch salience, which is weaker for a sound based on the vibration pattern shown in Fig. 31.6b than for a sound that contains a series of harmonic partials including f_1. However, the period in Fig. 31.6b remains at $T = 10$ ms corresponding to the repetition frequency f_0 of the complex waveshape (in a harmonic complex comprising a sufficient number of partials, f_0 equals f_1). Period length (or duration) for a given harmonic complex remains identical as long as several consecutive harmonics are left in the signal, and even with a number of nonconsecutive partials the same period length occurs. Consider, for example, a signal formed by linear superposition of harmonics no. 7, 9 and 11 of $f_1 = 100$ Hz. The resulting waveform is perfectly symmetric and periodic with a period of $T = 10$ ms as shown in Fig. 31.7.

However, one factor affecting pitch salience of a sound composed in this way is that higher harmonics may not be resolved as spectral components by cochlear filtering [31.15, 31] and another factor probably suited to weaken salience and to introduce some ambiguity is the structure of the waveshape, which, different from the selections shown in Fig. 31.6a,b, contains more than one peak per period. Assuming that autocorrelation analysis is apt (at least as a model [31.163–165]), a sound created like Fig. 31.7, when subjected to the *enhanced AC* algorithm [31.172], not only yields the

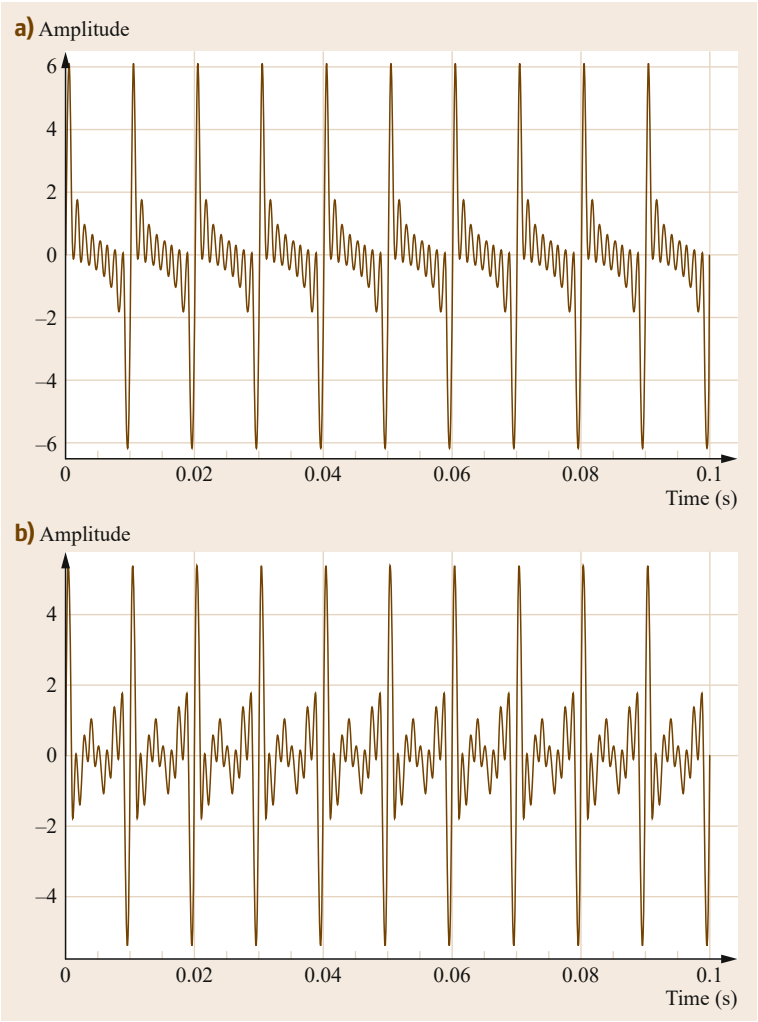


Fig. 31.6 **(a)** Harmonic complex, eight partials, $f_1 = 100$ Hz; **(b)** harmonic complex, partials 3–8 (300–800 Hz)

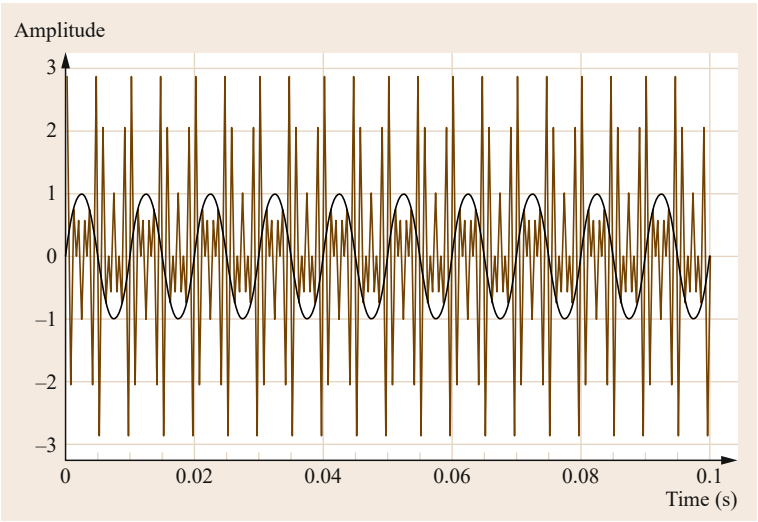


Fig. 31.7 Waveshape from superposition of harmonics 7, 9 and 11 {700, 900, 1100 Hz} of $f_1 = 100$ Hz. The *missing fundamental* f_1 (equal to the repetition frequency f_0 of the complex) is plotted separately (*in black*)

period at $T = 10$ ms but also strong periodicities at 4.5 and 15.5 ms, which means the ACF draws on several peaks contained in the waveshape (and its periodic repetitions) to calculate periods between these peaks. An observation worth reporting, in this circumstance, was that for a stimulus consisting of a pulse train with alternating intervals of 4.7 and 5.3 ms, respectively, recordings of spikes from fibers of the cat's AN subjected to ISIH analysis and autocorrelation histograms revealed periodicities at multiples of 5 ms [31.173], that is, the average of the two alternating intervals ($T \approx 5$ ms) as well as the period resulting from their combination ($T = 10$ ms).

It is of interest to note that spike trains recorded from AN fibers can transmit information representing the periodicity of a harmonic complex (with or without f_1 present [31.44, 45, 174]), thereby providing a strong neural basis for pitch sensation and perception. Neural encoding of periodicities in AN fibers includes musical intervals formed of two complex tones. ISIH analysis performed for several such intervals (minor second 16/15, perfect fourth 4/3, tritone 45/32, perfect fifth 3/2) demonstrates that the periodicities inherent in the acoustic stimulus signal are also found in the neural response where the ISIH peak structure closely mirrors the AC structure of the acoustic input [31.46]. Hence, the temporal input information seems to be fully preserved at the AN level.

The spectral fine structure of complex sounds can also be reconstructed from a Fourier transform of spike trains in the AN collected as period histograms or as interspike interval histograms [31.175], which demonstrates that the cochlear output contains spectral information encoded in temporal sequences. As with rate-place recordings from a rather small number of AN fibers, the Fourier transform of ISIH yields spectra which, for low stimulus SPL, resemble the input spectrum filtered by the CF/BF characteristics of fibers fairly well [31.174].

From a range of investigations of AN fibers as well as of neurons of the CN and of higher stations of the AuP up to the IC there is coherent evidence for both *place* and *time* representations of input signals [31.36, 150, 176]. Since rate-place mechanisms are markedly level-dependent, while purely temporal models fail to account for the greater salience of resolved against unresolved harmonics, a mixed *spatiotemporal* approach has been proposed [31.177]. In regard to place, tonotopic or cochleotopic organization is a feature found in nuclei of the ascending AuP and also in the auditory cortex (A1; [31.146, 147]). Topography of CF gradients and other features differ, however, from one station to the next. In the central nucleus of the IC (ICC), the cochlear frequency map is represented by an array of parallel frequency-band laminae where each lamina corresponds to about 1/3 octave, the width of a CB [31.178]. Also, the functional organization of the ICC (as observed in the cat) seems to indicate there are constant frequency ratios established between corresponding locations on neighboring laminae.

Periodicity encoded in spike trains has been detected in the AN, the CN and still in the ICC [31.106, pp. 270–280]. In the ICC a range of neurons responding strongly to particular modulation frequencies has been found. Such neurons can be characterized by their modulation transfer function, which refers to the discharge rate or the synchronization of units to the envelope of a sound as well as by their best modulation frequency (BMF). It has been suggested, from some observations, that in the auditory midbrain a periodotopic organization exists in ICC laminae where the periodotopic axis seems to extend in a nearly orthogonal pattern to the tonotopic axis [31.179–181]. Though this model as yet may be hypothetical, there is plenty of evidence that periodic stimulus patterns such as the repetition of the envelope of harmonic complexes or regular AM imposed on sounds are encoded in the neural response at the level of the AN, the CN and higher stations up to the IC [31.44, 45, 150, 153, 182–185]. Recordings from single units in the cat's IC show that the beat frequency of 26 Hz corresponding to two pure tones forming a minor second (466, 440 Hz in equal temperament, i. e., ≈ 100 cent) was matched in the discharge rate (relative to the onset of each beat cycle; [31.46]).

A temporal approach also seems apt in regard to binaural (spatial) hearing and dichotic phenomena such as the pitch of the *missing fundamental* created from two successive harmonic partials fed into the left and right ear respectively [31.186]; see Sect. 31.6.3. A temporal model working on ACF and CCF at least explains the perceptual outcome in a straightforward way. As an alternative, two independent sets of neurally encoded place informations from the left and right BM combined into a *central spectrum* have been discussed where harmonic partials are regarded as being matched against harmonic templates from which, in a final step, the fundamental f_0 is derived [31.187–189]. In addition, a mixed model combining place cues and periodicity to derive the *central spectrum* from an additional filtering stage of neural spike trains has been proposed [31.190]. One will have to distinguish the heuristic value of models demonstrating, in a plausible way based on system design and testing, how the pitch extraction process might operate, from the anatomical and neurophysiological evidence needed to validate models.

Moreover, it is obvious from a host of psychophysical experiments that envelope periodicity in many conditions is sufficient to evoke a salient pitch per-

cept. Consequently, neural periodicity analysis must be carried out in stations along the AuP. Factors rather indicating lower stations, up to and including the ICC, are the degenerative nature of phase locking in afferent AuP stations beyond the IC, and increasingly reduced neural firing rates observed in thalamic and cortical processing. In AN fibers, phase locking to pure tones may be as high as 4–5 kHz while the upper limit of neural phase locking to envelope periodicity based on AM is about 2.2 kHz [31.185]. Along the AuP, the upper limit of phase locking decreases significantly, absolute values depending on species investigated as well as on conditions (anesthetized or awake, etc.); a sizable decline is observed between the ICC and the thalamus (CGM), and more so between the CGM and the auditory cortices. In only a small fraction of CGM units, phase locking up to 1–2 kHz was measured while many units synchronized to lower frequencies (125–250 Hz [31.191, pp. 332–334]). In cortical units, phase locking to pure tones may occur up to several 100 Hz while locking to envelopes rarely seems to exceed 100 Hz [31.192]. This, however, is still enough to encode relatively slow AM such as the beat frequencies occurring between two simple or complex tones. Hence, there is evidence for a neural representation of roughness and sensory dissonance based on phase-locked activity at the level of the A1 [31.193]. Further, experiments on perception of the *missing fundamental* showed that patients who had suffered lesions of the right temporal lobe (including Heschl's gyri) performed significantly less consistently than controls [31.194], indicating that detection of envelope periodicity is effected in neural circuits embedded in the right temporal lobe. However, since phase locking declines and processing speed slows down towards cortical areas while the number of neurons involved increases greatly, neural processing of fine-grained temporal information seems less probable (though perhaps possible, depending on the type of code and the distribution of neurons involved [31.150, 195]). As an alternative, transformation of temporal coding into rate-place codes in stations of the AuP has been considered. One problem with rate functions is that, while firing rate usually is found to grow with sound level (subjected to compression increasing with SPL) at lower stations of the AuP, there is evidence of nonmonotonic rate functions from the ICC, the CGM, and the A1 largely due to inhibition [31.146].

The primary auditory cortex (A1) in humans *is found typically in the medial portion of the first anterior transverse temporal gyrus, also known as Heschl's gyrus* [31.196, p. 263]. Detailed anatomical and functional descriptions distinguish between primary auditory cortex, anterior auditory field (AAF), and posterior auditory field (PAF). Topography is also discussed with respect to core, belt, and parabelt regions of cortical areas [31.197]. From electrophysiological as well as imaging studies a tonotopic organization of A1 has been proposed [31.146, 147, 198]. A topography has also been suggested for stimulus intensity (with respect to the cat A1 [31.199]). One frequency gradient for low to high CFs has been identified to run from posterior to anterior A1. Two mirror-imaged tonotopic regions were found with magnetoencephalography (MEG) in the auditory cortex [31.200]. In addition, from MEG of six subjects, it has been inferred that the quasi-orthogonal arrangement of frequency (CF) and periodicity (BMF) suggested earlier for the ICC [31.180] applies also to orthogonal maps in the human auditory cortex [31.178]. Further, neurons in the auditory cortex of the marmoset monkey were reported to respond to pure tones and complex tones with missing fundamentals that have the same f_0 [31.201], that is, where the single frequency f_n (Hz) of the pure tone equals the repetition frequency (f_0) of the complex that has no spectral energy at f_1. Consequently, the same neurons must be capable of detecting either a frequency for which a spectral cue exists or derive the same frequency from a temporal cue (envelope periodicity), thus providing a neural basis for perceptual pitch constancy, i. e., the experience according to which sounds with different spectral and/or temporal structure can evoke *the same pitch* (or nearly so). Since earlier investigations [31.202] had failed to find neural substrates for pitch processing of harmonic complexes with a *missing fundamental* in the auditory cortex of rhesus monkeys, this particular report was met with some skepticism [31.195]. However, there might be neurons in stations of the AuP up to the auditory cortices responding to the same periodicity transmitted either from pure tones or from envelopes of harmonic complex tones lacking f_1.

Applying four functional criteria, namely pitch selectivity, pitch constancy, covariation of response magnitude with pitch salience, and elimination of peripheral phenomena, a study [31.203] combining functional magnetic resonance imaging (fMRI) methodology and psychophysical experiment ($n = 16$ subjects) was undertaken to locate, if possible, a cortical *pitch center*. Five different stimuli (a pure tone plus Gaussian noise masker, a harmonic complex, resolved partials of the same complex band-pass filtered to remove several low partials, again together with Gaussian noise masker, etc.) were presented all of which should evoke *the same pitch* equivalent to that of a pure tone at 200 Hz. Findings included *the fact that pitch-related activation rarely occurred at exactly the same point in the auditory cortex across the different listeners* while *spatial consistency was much more striking within listeners*. Locating the *pitch center*, for seven subjects *it fell in*

different portions of the planum temporale (PT), in one listener it fell in planum polare (PP) and in two listeners it was elsewhere [31.203, pp. 87 f.]. Also, the response magnitude in the pitch center for the different stimuli was unrelated to the pitch salience found in psychophysical experiment. There have been numbers of MEG and fMRI studies directed to finding *the* pitch center (for primates and humans [31.204]), but although they localize distinctive cortical areas active in pitch processing, neither technique converges on microanatomical mapping and assignment nor on the number (one, two) of tonotopic frequency gradients. More recently, again two tonotopic gradients were reported on the basis of fMRI data [31.205] to be located in bilateral cortex on Heschl's gyrus, one on its caudal and one on its rostral bank. From experiments using an adaptation electroencephalogram (EEG) (AEP) design, a circular cortical pitch map has been suggested to account for the circularity of *tone chroma* within the octave [31.77]; Sect. 31.1.2.

Recent discussions of the role of A1 and adjacent cortical areas reflect changes in perspective. Traditionally, the AuP including A1 has been viewed as a basically hierarchical system where certain tasks are performed on each level in regard to feature extraction (see Fig. 31.2 and the flow chart following the graph), with further processing of neural code for pitch and perhaps also timbre and loudness at A1 as well as in adjacent nonprimary areas. Such a view was obtained from, for the most part, electrophysiological recordings where neural response patterns reflecting the acoustical input were traced along the AuP as well as in A1 and other cortical areas. Results did account for a range of stimuli such as pure and complex tones, periodic click trains, AM and FM of pure tones, noise bands, etc. suited to evoke sensations of pitch. In A1 and adjacent areas gradients for frequency and intensity were found. Also, maps with more than one dimension have been proposed with respect to spectral and temporal representations of pitch (for an overview see [31.147, 192, 206]). In order to pinpoint processing of certain features, much of this research employed sound stimuli comprising a small number of dimensions, and single stimuli rather than combinations thereof (for reasons of experimental control and also to minimize context sensitivity). Meanwhile, the use of more complex, real-life stimuli such as animal vocalizations (in animal experiments [31.207]) or speech and musical sound is more widespread. Such studies revealed there is a belt of areas surrounding A1 involved in a more distributed processing reflecting stimulus complexity [31.208, 209]; distributed processing thus seems to distinguish complex from elementary stimuli. Also, animal experiments indicate there are neurons sensitive to particular feature combinations indicative of pitch, timbre, and localization [31.210]. In regard to integration of several features (e.g., tone height, intensity, spectral harmonicity, modulation, roughness), which have been analyzed along stations of the AuP, it has been suggested that a *central* task of the auditory cortices (implied are also belt regions around A1) could be forming auditory objects identifiable in a mixture of sounds [31.211, 212]. From high sensitivity of cortical neurons for certain spectral and/or temporal features on the one hand, and the need to integrate features into objects on the other, a determination for neural coding comprising three types of representation has been proposed [31.213]. These are:

1. Isomorphic representations where firing patterns *faithfully* reflect stimulus features and magnitudes
2. Nonisomorphic transformation of acoustical features into neural code
3. Internal representations of objects derived from a transformation of acoustical features into perceptual dimensions.

Since many real-world sound stimuli are fairly complex in regard to temporal and spectral structure, involvement of cortical areas beyond the acoustical cortices can be expected in order to analyze the *meaning* of distinct objects, and also for binding a sequence of objects into a formation (Sect. 30.1.2). From earlier experiments employing positron emission tomography (PET) it has been suggested that, for perceptual analysis of rather simple melodic phrases, networks of the right prefrontal and right frontal cortex are involved besides areas of the right superior temporal cortex [31.214]; comparison of pitches (in regard to intervals) seems to be carried out in the intraparietal sulcus (IPS [31.196]). Extending the dual-stream model of object (*what*?) and spatial (*where*?) processing, a differentiation between right and left hemispheres has been suggested according to which spectral processing would be predominantly done in the right hemisphere while temporal analysis would be achieved in the left [31.215] to account for often rapidly changing sound structures such as transients. However, given the very fast recognition of musical pitches and timbres in actual listening, the outcome of spectral and of temporal analysis then would have to be integrated instantly for object identification. Such operations, though *automated* in actual perception, draw on and reinforce memory. With respect to pitch perception, a representation of the relevant sound features in short-term memory (STM) seems necessary to allow for more detailed analysis. There are various hypotheses concerning the structure of the working memory as well as involvement of PFC substrates in actually integrating information (there is growing evidence for domain-specific working mem-

ories available for certain tasks [31.216, 217]). As has been observed in many experiments, the STM buffer for pitches and pitch sequences is considerably wider than that for visual stimuli. Musically trained subjects with relative pitch typically have a memory trace suited to reproduce pitches of single tones and undisturbed pitch sequences with only small errors within 1−2 min [31.218]; in general, the error (expressed as SD of pitch deviations) increases markedly with time indicating working memory for auditory input seems to function like a *leaky integrator*. However, in particular musically trained subjects seem to be able to draw on LTM resources and thereby can identify intervals and chords as well as tone sequences ordered along a scale or presented within a melodic Gestalt with higher precision, and also after a greater lapse of time (as was found already by [31.219]; see [31.74, pp. 130 ff.]).

Underpinning cortical plasticity, it has been shown that musical training and learning influences cortical processing of music in perception and performance [31.220, 221]; see also Part C of this book. There are some indications from evoked brainstem responses that musical training has even an effect on phase locking to periodic stimuli and hence on subcortical pitch processing [31.222]. Further, brainstem frequency-following responses (FFR) recorded from musicians and nonmusicians show that musicians synchronize faster to chord arpeggios and respond stronger to notes changed in tuning (as the third in either an A-major or an a-minor chord) than nonmusicians [31.223]. In addition, brainstem responses recorded in nonmusicians were more salient to consonant than to dissonant intervals formed of two complex tones [31.224] suggesting that neural processing makes use of the enhanced periodicity and harmonicity of intervals such as the perfect fifth 3/2 or the perfect fourth 4/3 (as compared against the minor second 16/15 or the major seventh 15/8). Findings established on the AN level of the cat [31.225] thus are corroborated by human brainstem FFR data.

31.5 Auditory Models and Pitch Extraction

On the basis of many empirical findings, obtained from both investigations of the AuP as well as psychophysical experiments, a number of auditory models have been issued since the 1980s (for an overview see [31.9, 131, 226]). Such models typically comprise a number of modules to emulate functions of the outer and middle ear, the BM (or the CP), the IHC and the AN; some include relays of the AuP (CN, IC). Models can be distinguished by either adopting a spectral or a temporal paradigm though they share some features with respect to peripheral processing. Typically, a linear filter is implemented for the transfer function of the ear channel and a bank of slightly overlapping bandpass filters for cochlear tonotopic frequency selectivity. The performance of cochlear models (including Gabor filtering and gammatone filter) has been reviewed recently [31.227]. The spectral approach to pitch extraction models [31.11, 228] operates on a pattern of resolved spectral components (ideally, but not necessarily, partials of a harmonic complex) derived from filtering the input signal. The filter in this case models the cochlear transfer function (CTF) and is similar to the gammatone filter that has been widely used in auditory signal processing [31.229]. The number of filter channels, their center frequencies and their bandwidth can be chosen so as to either match one of the established psychophysical scalings for frequency with respect to CBs or JNDs (Bark scale, ERB scale, SPINC scale; see Sect. 31.6.2), or can be derived directly from the CTF filter characteristics. In Terhardt's model, cochlear filtering results in a number of components that represent candidates for spectral pitches that might be audible as such (since many natural sounds contain spectral components spaced widely enough and strong enough in amplitude to evoke individual pitch percepts). After evaluation of possible masking effects and a weighting of such components that are retained, virtual pitches are extracted by means of a subharmonic matching process. Subharmonic matching means the search for a common denominator fitting a *root* to harmonic or quasiharmonic series (see below). Finally, a pitch-salience pattern is calculated from spectral and virtual pitches. The advantage of Terhardt's model is that it acknowledges the fact that musical sounds (for example, such recorded from swinging or carillon bells) may comprise several clearly audible pitches, which can either correspond to *resolved* harmonic partials or inharmonic spectral components or can result from the interaction of such components giving rise to one or even several virtual pitches [31.228, 230]. In addition to his model of spectral and virtual pitch extraction, *Terhardt* [31.11, 231] has proposed a special kind of Fourier transform called Fourier time transformation (FTT) suited to match spectral and temporal characteristics of peripheral auditory filtering. In certain respects [31.137], FTT analysis can be compared to wavelet analysis (or, in a more general view, to Gabor's concept of *logons* representing analyzable units in time-

frequency planes). Offering constant Q filtering rather than the constant bandwidth of conventional Fourier transform algorithms, FTT offers a better resolution for low frequency spectral components than short-term Fourier transform (STFT) (and a less precise spectral analysis for high frequency bands where harmonic complexes are not *resolved* into individual partials).

Several computational spectral models were designed following concepts of *harmonic templates* or *harmonic sieves* that had been proposed earlier in psychoacoustics on the basis of hearing experiments. Such template models hypothesize that a *central pitch* for complex sounds is established from peripherally resolved harmonics [31.186–188, 232] and that statistics are performed on resolved harmonics for a decision process leading to pitch estimates [31.187, 189]. Statistics based on calculating means and variances for frequency distributions and estimates based on criteria such as maximum likelihood will be needed in the case of hearing several harmonic complexes with overlapping partials as well as in regard to detuned partials or inharmonic complexes where components must be fitted to a template and more than one pitch estimate (expressed as a fundamental period or as f_0 of a complex) per sound may occur. A computational implementation of the *harmonic sieve* concept is the spatial pitch network model (or SPINET [31.233]).

Temporal models of peripheral and central pitch processing usually include a cochlear filter stage. Transduction of filter output into neural spike trains is effected by means of halfwave rectifiers (since IHC fire when stereocilia are deflected in one direction) as well as by implementing gain adaption to simulate nonlinear hair cell compressive behavior. Spike trains in simulated single AN fibers are subjected to AC analysis for channel periodicities, and a sum AC (or correlogram) analysis typically is performed to determine common periodicities across channels, which are fed into a final pitch detection or pitch estimation stage [31.162–164]. The neural activity pattern within and across channels may be represented as auditory images [31.229] that closely reflect the periodic structure of harmonic input signals (such as vowels in speech or sounds from organ flue and reed pipes). Strobed temporal integration may be substituted for AC to account for temporal asymmetry in perception as can result from sounds that are either *damped* or *ramped* with respect to the temporal envelope [31.234, 235]. Also, nonlinear filter types have been substituted for linear gammatone filters to match models even closer to observed BM or CP behavior [31.236]. Computational AN models [31.237] likewise incorporate nonlinearities to account for realistic tuning curves as well as for level-dependent compression and suppression effects.

The basic architecture of early temporal models followed *Licklider*'s [31.157] *duplex* design based on peripheral filtering and subsequent AC analysis. AC analysis can handle a broad range of stimuli and is suited to explain many pitch phenomena [31.163, 164]. It needs, however, a long delay line to deal with low frequencies (see above) for which physiological evidence is lacking. For this reason, alternative systems have been designed where the overall periodicity is determined by means of coincidence detection. The mechanism of coincidence detection has played a role in auditory theory for a long time [31.158] with respect to spatial hearing and sound localization. A model for detection of harmonic intervals in the mammalian CGM based on coincidence of clock cell discharges and spike trains from pure tones forming small integer ratios has been hypothesized [31.238]. To perform periodicity analysis in a physiological plausible model, large assemblies of CN chopper units (for anatomical and physiological background [31.98]) and IC coincidence detectors have been implemented [31.239]. This model extracts periodicity pitch from signals composed of both *resolved* and *unresolved* harmonics and also from a type of inharmonic signal where all harmonic frequencies are shifted by an equal amount, Δf [31.16]. Since the pitch based on the periodicity of a signal can shift if one or several partials are detuned by a certain amount from harmonic ratios, period estimation becomes more demanding than in conventional AC (for a possible solution see [31.240]).

There have been more attempts at building realistic auditory modules including AN, CN (in particular its ventral part, VCN, a main AuP relay), SOC and IC [31.241]. Also, the signal processing approach has been expanded to include a presumptive cortical stage simulating analysis of complex sounds in AI spectrotemporal response fields [31.242]. To this end, peripheral (constant-Q) filtering and nonlinear transduction to neural signal representation is followed by a bank of (linear) modulation filters that yield information about both the temporal and the spectral modulation of the input.

Among the goals for physiologically plausible modeling is to explain, besides pitch perception, some rather uncommon auditory phenomena such as comodulation masking release (CMR; masking is *reduced* by adding *flanker* signals to a probe and a masker, all of which are amplitude-modulated at the same rate; for details see [31.243]). In a model of auditory brainstem processing, CMR has been tentatively explained as *wideband inhibition of narrowband and delayed inhibition* effected by CN circuits [31.244, p. 389]. CMR indicates interaction of spectral components which, adopting the correspondence of CBs to auditory filters (AFs), are

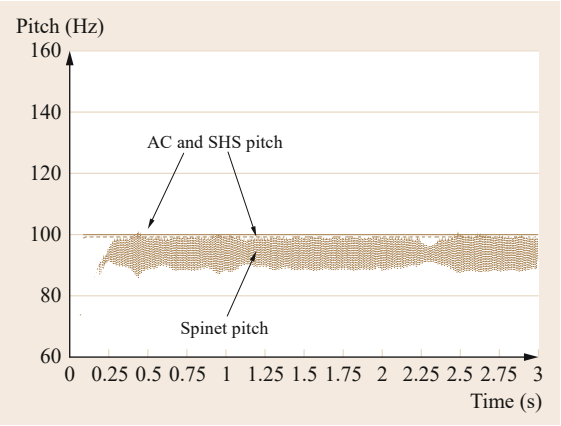


Fig. 31.8 Pitch estimates for a harmonic complex (ten partials, 100−1000 Hz); AC and SHS are at 100 Hz, SPINET offers a *pitch zone* at an average of 94 Hz

falling into separate CBs; CMR seems thereby to be contradicting traditional CB concepts (Sect. 31.6.2).

In order to compare the performance of some of the models mentioned above, a harmonic complex comprising ten partials (100−1000 Hz) in sine phase and a harmonic complex with partials of numbers 4−13 (400−1300 Hz) in sine phase were analyzed by two AC-based algorithms [31.172, 245], one model designed for subharmonic pitch summation (SHS [31.246]; the approach is similar to Terhardt's concept), and one model using the *harmonic template* (SPINET as implemented in the Praat environment; [31.247]). For the harmonic complex, the two AC algorithms and SHS turn out 100 Hz (which is both f_1 and f_0 of this complex) as *fundamental* pitch; the advanced AC algorithm [31.172] in addition offers subharmonics (33, 20 Hz, etc.) as possible pitch candidates while the SPINET model yields a *pitch zone* with a minimum near 86 Hz, maximum close to 101 Hz, and a mean at 94.08 Hz (SD = 3.79 Hz) (Fig. 31.8).

For the harmonic complex comprising partials of numbers 4−13, both AC algorithms again indicate 100 Hz as *fundamental* (in this case, with f_1–f_3 missing, 100 Hz = f_0). SHS fails at finding a single pitch estimate for this sound (yet calculates several f_0 frequencies), and SPINET delivers some spurious output. Similar results are obtained when the two complex sounds (partials 1−10, partials 4−13) consist of alternate sine and cosine partials. The AC algorithms are robust under this condition (owing to the strength of the mathematical principle). Peripheral preprocessors (filter banks as BM/CP models), pitch estimation algorithms and ear models incorporating IHC and AN modules are really put to test when fed with complex sounds of varying spectral inharmonicity such as sounds from swinging and carillon bells or sounds recorded from Javanese and Balinese *gamelan* instruments [31.93, 248–250]. Since increasing spectral inharmonicity implies a decline in periodicity, subharmonic matching and even AC algorithms become less effective, and fail to obtain a single pitch estimate for many complex sounds. To be sure, such outcome reflects perceptual ambiguity as experienced even by musically trained listeners exposed to inharmonic sounds. Music performed on carillons can be regarded a paradigm for composite sounds that are quite intricate to analyze with respect to pitches, in particular when several minor-third bells are played simultaneously [31.251]. One has to remember that *Schouten* [31.24, 32, 33] discussed the concept of the *residue* by pointing to the first ten spectral components of a typical minor-third carillon bell (see Sect. 30.2, Table 30.5, and Fig. 30.16. For details see [31.251, 252]) to argue that the *strike note*, as a virtual pitch, could arise from the interaction of three nearly harmonic components (so that the strike note would equal the bell's prime in pitch but not in timbre). Pitch ambiguity, as is often experienced in bell sounds, can even occur when listening to harmonic complexes such as chords and sonorities where several spectral pitch cues exist and virtual pitches may also arise from an overlay of partials. In principle, almost all complex sounds can give rise to more than one pitch sensation [31.11, Chap. 11]. Also, there are interdependencies between *pitch* and *timbre* (Sect. 32.2); for example, the pitch of a harmonic complex may change by one octave depending on spectral energy distribution [31.253].

31.6 Psychophysics

In this section, a range of phenomena relevant for sensation and perception of pitch will be presented in an overview covering, in particular, JND and DL for *pitch*, critical bands, roughness and dissonance, the *residue* and the missing fundamental, combination tones, fusion and consonance.

31.6.1 JND and DL for Pitch

The JND usually is defined as the smallest increase in stimulus intensity or magnitude detectable by subjects within a sensory modality. Investigations for JNDs relate to a certain experimental technique

in psychophysics (outlined already by [31.254]; see Sect. 30.1.4 above) as well as to a statistical measure in that a JND is established on the basis of responses or observations obtained from subjects indicating a small difference between the magnitudes of two stimuli has been noticed in at least 50% of all trials. The DL often is regarded as the absolute threshold for the discrimination of two stimuli under certain conditions, or as the change in a stimulus that *the average observer* [31.156, p. 998] is capable of detecting. Either referring to an absolute threshold or to an *average* that holds for a population, a DL is a generalized magnitude derived from several or even many JND experiments. The very small JND found in many experiments for pure tone frequency discrimination indicates high sensitivity of hearing for small stimulus changes. On the basis of experiments where frequency-modulated pure tones served as stimuli, the number of JNDs for pure tones in the range of normal hearing has been estimated to comprise 640 steps distributed over 24 frequency groups [31.255], see Sect. 30.1.4, *Loudness Scaling and Scaling of Pitch: Two Illustrative Examples*. The size of the DL ($f/\Delta f$) depends on the reference frequency as well as on conditions (SPL, earphones or free field, monaural or binaural hearing) under which the sounds are perceived; in the range from ≈ 200 Hz to 1 kHz, JND for pitch of pure tones presented at 40 dB SPL in humans was about 1 Hz [31.256]. Both cochlear *place* and neural *time* patterns can convey pitch information in this frequency range. In experiments with a large sample ($n \approx 400$) of subjects, many of which had either professional musical training or were active as amateur musicians, for complex synthetic and natural sounds with f_1 in the range of 250–300 Hz about 90% of the sample recognized frequency shifts of 7 cent [31.257] and still 55% of only 4 cent (4 cent = 1/25 of the size of a half tone (HT) in equal temperament, ET). For the DL of 2 cent (the absolute threshold for pitch deviations that can be detected by humans), the figure was 20% since in particular violinists and sound recording supervisors (German: Tonmeister) were able to deal with such subtle shifts. JNDs for pure tones have also been obtained for nonhuman primates [31.258] where a range of 1 Hz to more than 10 Hz at 1 kHz was found in several studies. It was shown for some species that they can discriminate complex harmonic tones as well [31.13].

31.6.2 Critical Bands (CB), Roughness and Sensation of Dissonance

The concept of CB was formally introduced by *Fletcher* [31.259, 260] who explored the relation of noise masker and pure tone intensities in regard to cochlear excitation patterns and loudness perception. For a pure tone of frequency f_k to be perceived, its intensity must at least equal the average intensity of a noise band of a certain bandwidth. *Fletcher* [31.259, p. 55, Fig. 17], offered a graph where the bandwidth of the noise masker is related to the pure tone frequency; according to Fletcher, the bandwidth of the CB (in Hz)

> *is numerically equal to the ratio of the intensity of the tone masked to the intensity per cycle of the noise producing masking and always corresponds to* 1/2 mm *of length on the BM.*

This ratio has been termed the *critical ratio* (CR). A formal definition [31.261, p. 174] of the CR (in dB) is $\mathrm{CR} = \mathrm{thr} - L_{\mathrm{ps}}$, where thr is the threshold (in dB) of the pure tone, and L_{ps} is the spectrum level of the noise band. The CR as expressed by *Fletcher* [31.259] differs from the bandwidth of the CB found later on in experiments on loudness summation [31.262]; it seems that, at many frequencies, the CR is about 0.4 the value of the CB [31.261, 263]. Characteristic of a CB as described by Fletcher was that, for a given tone of frequency f and intensity I, increasing the bandwidth of the masker beyond a certain value does not change the masking effect. *Fletcher* [31.260] regarded the masker of a certain bandwidth relative to the frequency of the pure tone as activating a constant portion of length of the BM (plus neurons attached; his concept included dividing the BM into 100 *patches of nerve endings*; see Sect. 33.2). In this respect, the CB is the basic unit of the cochlear map that thereby can be described as a chain of filters wired in parallel. Performance features of the auditory filter (AF) have been studied in many experiments where it was found that the bandwidth and the shape of the AF depend, to some degree, on frequency position along the BM as well as on stimulus level (for overviews see [31.263, 264] and [31.159, Chap. 10]). While the AF at low levels is almost symmetric, the low frequency slope of the filter becomes less steep with increasing level. In regard to frequency discrimination and pitch perception, the AF having a bandwidth corresponding to one CB will cover a number of frequency components presented simultaneously falling into this band; hence, such components cannot be *resolved* and perceived as individual constituents. In regard to sensation of loudness, the AF can be viewed as a *bin* that collects the energy or power falling into each AF [31.262, 265]. Both the pitch and the loudness aspect have been discussed with the CB as the basic unit relevant to auditory perception [31.7, 266]. In addition, *Fletcher* [31.259] had speculated that his CB corresponds to 1/2 mm of BM. A similar idea was expressed by *Zwicker* et al. [31.262] who had found CB

bandwidth to be *almost invariant* with level, and had expressed CB bandwidth in mel (Sect. 30.1.4, *Loudness Scaling* ...). From CB data available then as well as from studies *von Békésy* [31.29] had undertaken to establish a cochlear frequency map for several species, *Greenwood* [31.128, p. 1350] derived a formula for calculating cochlear coordinates (Sect. 30.1.4, *Loudness Scaling* ...). From his calculations (and assuming nonoverlapping CBs) he concluded *that one critical band is equal to approx.* 1 mm (of BM), and that about 35 CBs would account for the human range of hearing (20 Hz to 20 kHz). A scale comprising 24 CBs was issued by *Zwicker* [31.255], [31.12, Chap. 6] where for each band the center and the lower and upper limit frequency is given. The CBs are taken as a unit of a critical band rate (z) that spans a frequency range from 100–15 500 Hz; the CB units are termed *Bark* (to honor the contributions of Heinrich Barkhausen (Sect. 30.1.4, *Loudness Scaling* ...) to psychoacoustics). More recently, a refined CB scale was proposed by *Moore* and *Glasberg* [31.267] based on units calculated as *equivalent rectangular bandwidth* (ERB [31.263, 264]). The concept was developed to approximate AF characteristics more closely, and to account for known phenomena such as level dependence of AF bandwidth. The ERB scale is similar in shape to the Bark scale for frequencies above 500 Hz, yet different below where the Bark scale assumed a constant bandwidth of 100 Hz per CB while the ERB decreases in bandwidth toward low frequencies. Also, the ERB scale comprises nearly 40 CBs as compared to the 24 Bark CBs and hence offers a finer grid that is closer to actual frequency selectivity observed in humans. BM sections corresponding to each of the ERBs would be ≈ 0.86 mm wide. As shown in Fig. 31.9 for a range of 50–500 Hz, ERB grows linearly with center frequency.

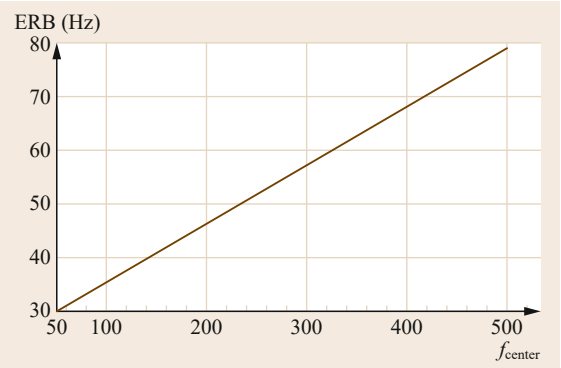


Fig. 31.9 ERB (Hz) as a function of center frequency (f_{center}, 50–500 Hz)

The Bark scale at low frequencies ($f < 500$ Hz) does not conform to listening experience in regard to musical intervals and to the CBs obtained in experiments on roughness sensation. This is quite evident for Bark band 1 (0–100 Hz), which is not well defined (it probably should have used the lower limen for pitch perception at 20 Hz as a reference point) while the next three Bark CBs numerically have the size of an octave, a fifth and a fourth respectively, if expressed in cents, that is, in a perceptually meaningful unit [31.268]. An interval of two sine tones 100 : 150 Hz presented at 50 dB SPL on suitable loudspeakers by most subjects will be heard as a fifth (of 702 cent) formed of two tones that can be identified as such (that is, these two tones are *resolved* and do not fall into one CB). At 200 Hz, two tones forming a minor third (200 : 240 Hz = 316 cent) can be as clearly identified as the interval itself (otherwise, music as we know it would not be possible). Thus, the respective Bark band, assumed to range from 200–300 Hz [31.12, p. 159], is much too wide. The ERB, because of the decrease of bandwidth below 500 Hz, is more appropriate, but still seems a bit wide at very low frequencies, $f < 100$ Hz [31.168]. In experiments on psychophysical tuning curves (PTC) for low frequencies, the *bottom* AF *appears to be located between* 40 *and* 50 Hz [31.269, p. 3179]. In fact, scores obtained for subjects (with some musical training) from interval recognition tasks improved to some degree if the fundamental of the lowest tone was at or above 50 Hz [31.168].

The CB concept is of particular relevance for the sensation of roughness and beats, which in turn are closely involved in sensory dissonance. In his theory of sensory consonance and dissonance, *Helmholtz* [31.28] described sound phenomena where, in pairs of complex tones, fundamental frequencies and higher partials either coincide or deviate from each other so as to cause *disturbances* in sensation. Such disturbances according to Helmholtz are scalable phenomena (on a sensory scale ranging from pleasant to unpleasant) since small deviations from just frequency ratios (e.g., the fifth and fourth in ET) will be tolerated, and a soft modulation may even be sensed as pleasant while there are sonorities that are intermittent in their temporal envelope, to result in a sensation of strong beats. Helmholtz judged such intermittent *creaky* sounds as unpleasant because the neural excitation pattern would reflect the intermittent sound structure (there is empirical evidence for this hypothesis now available [31.225]). He found that sensations of beats and roughness in regard to musical intervals played simultaneously were dependent on two factors, namely (1) the size of the interval and (2) the octave position where the interval is put to sound. For

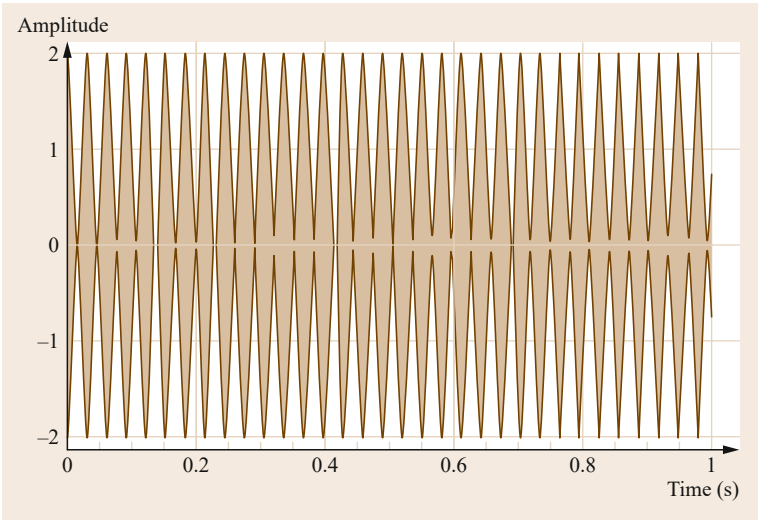


Fig. 31.10 Helmholtz minor second B_4–C_5, 1″ of sound

two pure tones played in the middle range of musical tones, he argued roughness reaches maximum at beat frequencies of 30–40 Hz. As an example, he pointed to the minor second (h′–c″ in Helmholtz's designation, which is B_4–C_5 in USA standard) where the frequency difference amounts to almost 33 Hz (relative to A_4 = 435 Hz). From Helmholtz's description it is clear that by *beats* (German: *Stöße*) he meant the number of intermittent peaks per second (Fig. 31.10).

For most listeners, the beat frequency may be too high to be counted. Also, the two frequency components that form the interval are too close to be discriminable as constituents of the interval, i. e., a minor second. Helmholtz discovered that by keeping one tone constant in frequency and shifting the other slowly from unison upwards or downwards, maximum roughness occurs at the minor second (frequency ratio 16/15, $\approx$ 112 cent) while it is less at the major second (9/8, $\approx$ 204 cent), and about to disappear when reaching the minor third (6/5, $\approx$ 316 cent). He calculated a roughness profile for simultaneous intervals within two octaves (c′–c‴ = C_4–C_6) based on coincidences and interferences between partials of violin tones. While studying consonance and dissonance, Helmholtz had discovered what later became known as the CB, that is, the phenomenon where two pure tones played simultaneously at a narrow interval interfere so as to produce beats and audible roughness while they cannot be identified as single constituents in the sound.

For two sine tones of unity amplitude, one fixed (say, at 200 Hz) and one variable in frequency, their relation changes from unison to slow modulation, then to increasing roughness, which is followed by an *interference zone* where strong beats occur. With the distance (in cents) of the two tones further increasing, roughness in proportion decreases. Finally, two discrete tones become audible at an interval that, in the range from about 300 to 2000 Hz, is larger than a major second and smaller than a minor third (Fig. 31.11). The distance (in cents) where the fixed and the variable tone become audible as two distinct parts of the stimulus can be regarded as the bandwidth of the CB.

The term *interference zone* here is chosen to avoid the term *fusion*, which is often used to denote the phenomenon that two tones cannot be distinguished if too close in frequency. Since the term fusion plays a central role in the theory of consonance (see below) and as two tones beating strongly in fact are sensed as markedly dissonant, terminology should reflect their interference. As was observed already by Helmholtz, the absolute size of the interval where pairs of tones leave the interference zone and become audible as two distinct tones depends to some degree on the frequency range. At low frequencies ($f < 200$ Hz), the bandwidth of the CB seems enlarged as compared to the middle range (where the CB is about 250 cent). For example, taking the minor third A_1–C_2 as played on a grand piano, the fundamental frequencies of the interval in ET are at 55 and 65.4 Hz (= 300 cent) respectively. According to the ERB function, they cannot be sufficiently separated. Assuming AFs to overlap in reality, one would observe some spectral leakage from one CB filter into the next. In fact, listening to such intervals, one can sense roughness due to spectral interference of the two complex tones.

Mayer [31.270] has been regarded as having discovered, almost by chance, the CB [31.128]. In fact he deals with small changes in the size of consonant intervals detectable by trained listeners. For example, the interval 256 : 315 Hz (produced from two tuning

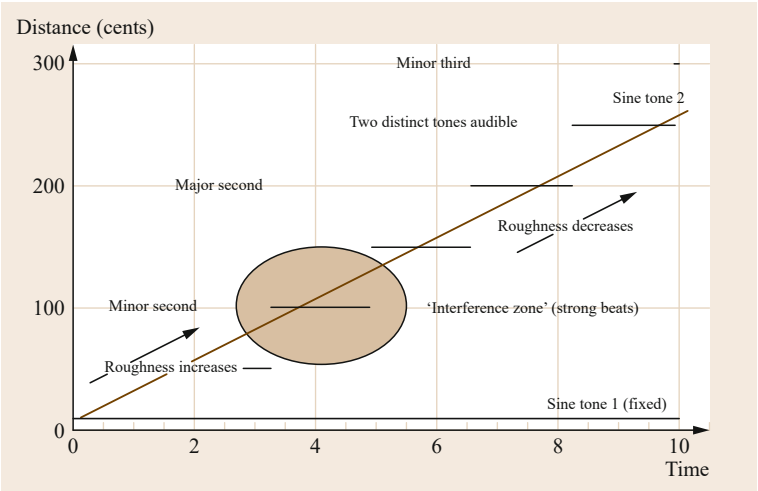


Fig. 31.11 Roughness, beats, CB bandwidth as observed at ≈ 200–500 Hz

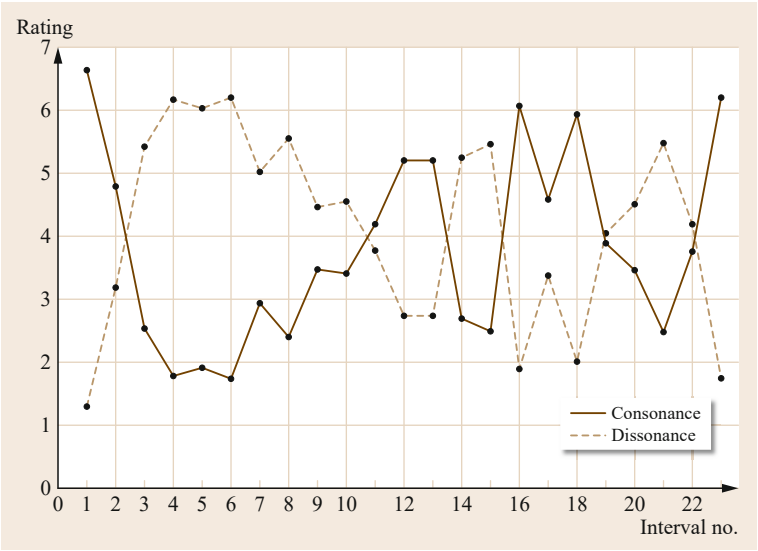


Fig. 31.12 Consonance (*solid line*) and dissonance (*dashed line*) profiles for 23 intervals from unison (no. 1) to octave (no. 23). No. 13 is the pure fourth, no. 16 is the pure fifth, no. 18 is the just major sixth (after [31.268])

forks), representing the tones UT_3–MI_3 was judged as *consonant* by Mayer's observers (among them musical experts). The ratio 256 : 314 Hz appeared as *just consonant* while 256 : 313.5 Hz was said to be *slightly rough* and 256 : 313 Hz *decidedly rough*. To be sure, all four intervals constitute a *neutral third* (about halfway between major and minor third; an interval found in European folk and non-Western music) and have a size of 359, 353.5, 351, and 348 cent respectively. Applying ERB measures, none of these neutral thirds falls within a single CB.

Auditory phenomena related to CBs as concept have far-reaching implications for human psychophysics and also for audio technology [31.159, 271]. Taking the original report *Helmholtz* [31.28] gave on sensations of roughness and dissonance, both are closely related in the experience of musical intervals within one octave (or even within a double octave). Putting a range of 23 intervals within one octave to test, *Schneider* et al. [31.268] obtained judgments from musically trained observers ($n = 51$) on a seven-point scale for *consonance* and *dissonance* of each pair of complex tones. The data were transformed into two complementary profiles for consonance and dissonance (Fig. 31.12). Similar to Helmholtz's observation, dissonance was highest at and around the minor second, at the tritone ($45/32 = 590$ cent) and at the major seventh ($15/8 = 1088$ cent) while consonance was highest at the unison, the pure fifth ($3/2$) and the pure major sixth ($5/3 = 884$ cent).

One should bear in mind that Helmholtz based his observations and conclusions on spectral features and

place analysis rather than on *period* and temporal analysis. While much of the earlier discussions of roughness and dissonance were concerned with spectral structures in regard to *disturbances* or *friction* between narrowly spaced partials on the BM level, there was a shift towards temporal features later on [31.272, 273]. Among the reasons is that, taking the model of the BM as a chain of parallel bandpass filters [31.259], one can argue several narrowly spaced spectral components can be expected to result in one AM signal at the output of each AF. Another aspect is that temporal models usually are phase sensitive. Moreover, signals that give rise to a sensation of roughness can be synthesized in a straightforward manner according to the basic formula

$$y(t) = (A + a \sin [2\pi f_M t]) \sin [2\pi f_T t] ,$$

where A is the mean of the amplitude and a/A is the modulation factor. Applying trigonometric rules, this formula becomes

$$\begin{aligned} y(t) = {} & A \sin [2\pi f_T t] \\ & + \frac{a}{2} \cos [2\pi \, (f_T - f_M) \, t] \\ & - \frac{a}{2} \cos [2\pi \, (f_T + f_M) \, t] . \end{aligned}$$

Hence, the AM signal consists of the carrier at frequency f_T and two sidebands at frequencies $(f_T - f_M)$ and $(f_T + f_M)$. Various AM signals can be generated by manipulating f_T and f_M as well as their amplitudes. As a reference for AM-induced roughness, a signal with $f_T = 1000\,\text{Hz}$ and a modulation frequency $f_M = 70\,\text{Hz}$ played back to observers at 70 dB SPL has been defined as *1 asper* (in Latin *asper* = rough). Taking modulation depth and SPL of (artificial) signals as parameters relevant for roughness sensation into account, it has been suggested that, on the asper scale, about 20 roughness steps could be distinguished [31.12, Chap. 11].

The spectral and the temporal approach to roughness sensation might be regarded as equivalent since both superposition of sinusoidals and AM should result in identical stimulus spectra and temporal envelopes. However, in regard to sensation, to make a sine tone carrier sinusoidally amplitude-modulated at a rate of 40 Hz and two sinusoidal components beating at 40 Hz appear equal in roughness, the modulation factor in the AM signal needs to be set to 0.6 or 0.7, for a wide range of AM carrier frequencies and mean frequencies of the tone pair respectively [31.274, p. 207]. The reason is the temporal envelopes of the two signals are not identical. In music, one typically has to deal with complex spectra where roughness primarily arises from interferences between neighboring spectral components due to dissonant sonorities and/or sounds from idiophones that are clearly inharmonic. A good case to illustrate the problem of roughness from concurrent dissonant sonorities is singing styles of the Balkans where two voices are often singing at an interval smaller than a whole tone, and sometimes two voices go in parallel at a frequency distance of only $\approx 50{-}80$ cent [31.268, 275]. A musical tradition where inharmonicity from idiophones comes into play is carillon music of the Low Countries where frequently pieces written in a major tonality are played on a carillon comprising minor-third bells. Concords (two or even three bells sounding at once) thereby result in marked roughness, besides ambiguity of pitches (the two effects combine perceptually [31.249, 251, 252]). In musical sound analysis, it is essential to take spectral structure into account for modeling roughness [31.276]. In our own experiments, when a range of synthesized stimuli was fed into either computer models based in the time domain or into a (preliminary) model based in the frequency domain making use of an algorithm proposed by *Sethares* [31.277, 278] that in turn takes into account roughness profiles from *Plomp* and *Levelt* [31.279], it was found that the frequency-based model was much more in agreement with behavioral data obtained for the same stimuli [31.268]. Since AF characteristics, in the first *place*, determine to which degree spectral components can be either *resolved* or interfere with each other [31.7, 263, 264], spectral characteristics of signals have to be considered as much as temporal parameters in evaluating roughness phenomena [31.11, Chap. 10.2].

31.6.3 *Residue, Virtual Pitch,* Combination Tones

As has been exposed in Sect. 31.1.1, *Seebeck* [31.22] discussed the phenomenon of the *missing fundamental* (MF), at least as a hypothesis. It was known among organ builders and organists for a long time that two pipes of 16′ and $10\frac{2}{3}'$ length respectively, when sounded together would produce a pitch one octave below the 16′ pipe. Since 16 and $10\frac{2}{3}$ form the ratio 3 : 2 (regarding pipe length) and 2 : 3 in regard to fundamental frequencies (a fifth), the pitch resulting from the combination is the *root* of the series: (1) : 2 : 3. Of the three pitch components, two are real and one is *virtual*. The phenomenon known as *acoustic bass* has been explained either as a difference tone (f_2–f_1 [31.52, p. 59 f.]) or as a *virtual pitch* ([31.11] demonstration 27 on CD accompanying the book). In fact, two organ pipes tuned to A_1 ($f_1 \approx 54.8\,\text{Hz}$) and E_2 ($f_1 \approx 82.3\,\text{Hz}$) respectively, when sounded together generate a pitch that, in an AC-based analysis, is an A_0 at 27.4 Hz (Fig. 31.13). To be

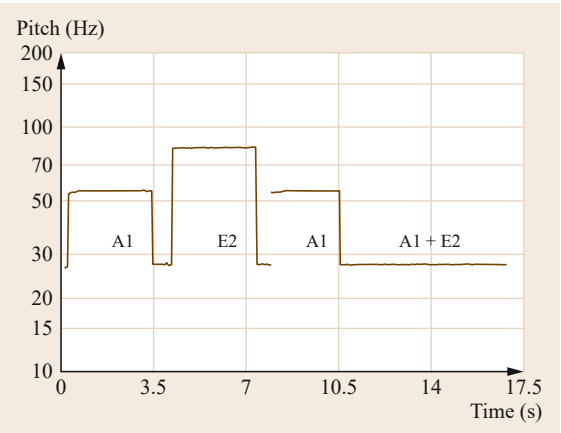


Fig. 31.13 Acoustic bass on pipe organ: A_0 generated from $A_1 + E_2$ played together

sure, such a low pitch, to be produced by one pipe alone, would require a 32′ pipe tuned to A_0.

Further evidence that f_1 of a harmonic complex may not be the sole cue for pitch sensations was found when electric filtering of musical tones demonstrated that, in bowed strings, the fundamental is relatively weak in the spectrum of the lowest tone [31.280]. Notwithstanding this weak f_1 component, the pitch of the bowed violin g-string is registered, by listeners, as G_3. *Schouten* [31.24, 32, 33] developed the concept of the MF to account for a range of sounds and perceptual pitch effects such as the *strike note* of bells. Schouten provided further evidence for a dual mechanism of pitch perception, one making use of f_1 while the *residue* evokes the same *low* pitch at f_0 as conveyed by f_1. *Schouten* [31.24, p. 360] described the residue *as one component with a pitch determined by the periodicity of the collective waveform*, namely, that of higher partials of a harmonic spectrum from which f_1 and possibly several low partials have been removed. According to Schouten's mechanical resonator model (as an analogue for BM filters [31.32, Fig. 9]), groups of successive harmonics will result in excitation patterns that, though different in shape, all represent the same fundamental period (in Schouten's experiments, the period was $T = 5$ ms of a pulse train evoking a pitch of $1/T = 200$ Hz). Experiments later demonstrated that, to evoke residue pitches, a rather small number of lower partials gain *dominance* against a large number of higher partials (if such are present [31.16, 281]). For example, to generate a *residue pitch* at $f_0 =$ 200 Hz, partials from 600 to 1400 Hz (partial nos. 3, 4, 5, 6, 7) appear *dominant* against all higher partials. To be sure, most of these partials can be resolved individually at the BM/CP filter level, which indicates there is *place* information available (Sect. 31.3) besides common periodicity. However, a residue pitch at $f_0 = 200$ Hz might as well be generated from harmonic partials of numbers 15–20 (at 3000–4000 Hz) even though such a sound cannot be resolved into its spectral constituents at given AF/CB characteristics. Moreover, since this sound (partials 15–20 of equal amplitude) carries all the energy in a rather high frequency range, the timbre now is quite *sharp* (Chap. 32) and the pitch is less salient than that of f_0 evoked from low partials. However, even with harmonic complexes consisting of high partials it is still possible to play simple melodic patterns that can be identified, above chance level, by listeners [31.3, 282, 283]. In regard to different distributions of low partials that are resolvable individually by cochlear filtering versus groups of higher partials not resolvable into constituents, there have been considerations whether both phenomena might be operating on the same neural mechanism [31.283].

It is astounding that pairs of successive partials presented at a suitable level and even fed into opposite ears are sufficient to evoke a pitch sensation corresponding to f_0 [31.186]. Also, one can identify melodies played with harmonic complexes in which, for each tone, a different set of partials is used. Since these partials do not include f_1, the melody one perceives consists of a sequence of *residue pitches* [31.11, 83]. Two possible explanations for dichotic residue pitch (one of a pair of successive partials fed into each ear) have been discussed: (1) pitch is derived from the temporal information, which, in the case of two (peripherally resolved) partials presented in dichotic listening, would need a *central* periodicity extraction circuit that combines the periodicities from two ears (probably on the level of the IC; see Fig. 31.2); (2) pitch is derived from a *central spectrum* combining the BM/AN place information into a cochleotopic map (probably on the cortical level) from where a final pitch percept emerges in a harmonic template matching process. In both cases, pitch is derived from a *central* mechanism capable of inferring the *missing fundamental* frequency f_0 from those harmonic components that are present either as *place* or as *time* information. The precision and reliability of the pitch estimate of course are dependent on the temporal and spectral structure of the sounds used as stimuli. *Schouten* [31.24, 32, 33] considered the *strike note* of bells as a possible case of residue pitch. The *strike note* is a virtual pitch brought about by the combined effect of several spectral components that come close to harmonic frequency ratios. Since bell spectra are often quite complex with respect to the number and spacing of components, more than one strike note might become audible as a residue pitch (for empirical data [31.284]) or, in Terhardt's model, as *virtual* pitches [31.228, 230]. Also, a strike note can

be perceived in addition to a low spectral pitch. Quite often, the strike note differs slightly in pitch from the prime, the second strong partial in a typical minor-third bell [31.284, 285]. Another spectral pitch usually well audible is the minor third which, as a strong component in most bells [31.252], *stands out* against the quasiharmonic series formed by those components that give rise to the strike note. As a result of several spectral pitch cues perceived separately from the strike note as a virtual pitch, subjects find that most bell sounds are ambiguous as to pitch when heard individually. However, in a sequence of such sounds, musically trained subjects can judge the interval between a pair of bells, for instance, by singing or humming the interval they believe to be represented by the *main* pitch in each bell [31.251, 286]. One may view pitches that the auditory system provides in particular from complex inharmonic sounds as *emergent* in the sense that such percepts are the outcome of a probabilistic estimate aiming for a *best fit* of actual spectral content against a template (as reflected in auditory models involving statistics [31.187, 189, 228, 230]). There are sounds produced from musical instruments such as Javanese and Balinese gong chimes (e.g., *bonang*, *trompong*) that, due to spectral inharmonicity and spectral density, are quite ambiguous with respect to the number and quality of pitches they tend to evoke [31.93, 248, 287]. However, even for such sounds, pitch perception rarely leads to completely unexpected or novel results but is still causally related to the input. It is rather that, in a sample of subjects listening to such sounds, individuals may pick up different cues (more or less prominent spectral components or combinations of such components) on which they base their pitch estimate.

As mentioned above, the *acoustic bass* known from pipe organs has been explained as a difference tone f_2–f_1. The phenomenon of two physically real tones resulting in the sensation of at least one more tone was described, around 1750, by Georg Andreas Sorge (a German organist, music theorist, and composer), by Jean-Baptiste Romieu (a French lawyer and scientist), by Jean Adam Serre (a painter from Génève who also published on music theory), and by Guiseppe Tartini, the well-known Italian composer and music theorist. An inspection of the relevant text paragraphs (which are assembled by Rubeli as part of a commentary in his translation of *Tartini*'s treatise [31.288, pp. 58 ff.]) shows Romieu and Serre refer to a *low third tone* (troisième son grave) that is the fundamental of two other tones forming an interval of a fifth, a fourth, etc. This *third tone* could be a virtual pitch of the residue type as well as a difference tone. *Tartini* [31.288, pp. 70 ff.] also has this *third tone* (terzo suono); after presenting his examples for two-tone intervals leading to sensation of a third tone that is a difference tone, Tartini claims that *all intervals of the harmonic series 1, 1/2, 1/3, 1/4, etc. produce the same 'third tone', namely the tone at 1/2*. Though one would expect the third tone at 1 and not at $1/2$ (the series $1, 1/2, 1/3, \ldots$ refers to intervals expressed as string lengths), what is implied is that the additional pitch evoked by intervals each consisting of two consecutive harmonics is the same and is, as Tartini emphasizes, *the physical fundamental of the harmonic system*. In this respect, it would be a virtual fundamental pitch f_0 resulting from two primaries f_m, f_n at harmonic ratios rather than a difference tone. Though both the periodicity pitch f_0 corresponding to the missing fundamental and the difference tone $f_d = f_2$–f_1 are virtual in the sense that the sound signal outside the ear does not contain energy at f_0 or f_d, they have different origins. While f_0 is a product of *central* processing ([31.186]; Sect. 31.4), difference tones, of which $D_1 = f_2$–f_1 and $D_2 = 2f_1$–f_2 are most relevant, are generated in the cochlear BM-IHC-OHC circuit due to nonlinear input–output functions of OHC; combination tones (sum and difference tones) produce traveling waves in the cochlea that have a maximum at a CF corresponding to the sum or difference of the primaries [31.98, pp. 147–150]. Hence, the quadratic difference tone D_1 and the cubic difference tone D_2 are distortion products (quadratic and cubic refer to terms in the nonlinear transfer function [31.31, 289] and [31.159, Chap. 22]). Typically, the quadratic difference tone D_1 becomes audible at medium and higher sound levels (> 50 dB). In contrast, difference tones of the type $f_1 - k(f_2 - f_1)$, $k = 1, 2, 3, \ldots$ (for $k = 1, D_2 = 2f_1 - f_2$ results) become audible already at low sound levels, but only within a limited frequency region below f_1 [31.290]. Moreover, audibility and prominence of D_2 appears markedly dependent on level and on the frequency ratio f_2/f_1 while D_1 is hardly affected by level above threshold as well as the f_2/f_1 ratio (test sounds are included in *Houtsma* et al. [31.83], demonstration 34). Combination tones that are sums of primaries are difficult to hear. If the primaries are harmonic complex tones, sums can be easily masked by partials of the primaries.

Combination tones (sum and difference tones) and the role they may have in regard to pitch and consonance were an issue in many publications of the 19th and early 20th century (for a historical review see [31.291]). Exploring perception of combination tones, *Stumpf* [31.292] concluded that, for pairs of (almost) pure tones one of which is fixed in frequency while the other goes slowly up over the range of two octaves, only a few combinations are of relevance, namely f_2–f_1 (D_1), $f_2 + f_1$, $2f_1$–f_2 (D_2), $2f_2$–f_1 and, already to a lesser degree, $3f_1$-$2f_2$ and $3f_2$–$2f_1$. He condensed his observations into a scheme [31.292, p. 135, Fig. 13];

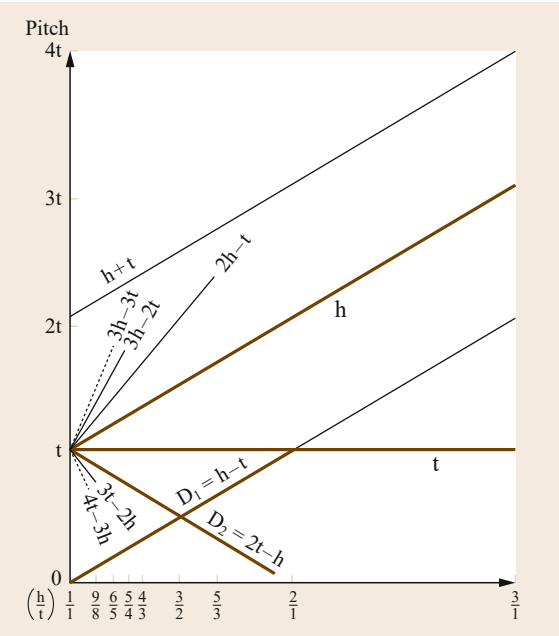


Fig. 31.14 Combination tones as explored by *Stumpf* [31.292]. Inside the coordinate system, h and t refer to high and low tone (of a pair of pure tones). On the ordinate, 0, t, 2t, 3t, 4t indicate the frequency relative to the low tone, t (constant in frequency). The abscissa gives the ratio of the two tones in musical intervals (from the unison 1/1 to the octave 2/1 and the 12th 3/1) ◀

that shows that the main effects of combination tones occur within the first octave, and rarely beyond the range of a twelfth (Fig. 31.14).

A scheme similar to Stumpf's can be found in more recent publications on psychophysics [31.293]. Hearing D_2 and possibly also D_1 is a condition that Tartini and other violinists, playing their instruments quite close to their ears, could have realized. Though combination tones are known to result from cochlear nonlinearities [31.98, 289], [31.31, Chap. 2], musicians and even Helmholtz have pondered the possibility of combination tones arising also from the acoustics of certain instruments. There are some indications that, when a harmonic interval such as a fifth or fourth is played as a double stop on a cello, relatively weak components can be detected in the spectrum corresponding to a frequency that is the difference of the two fundamentals; if these components ($f_2 - f_1$ and its multiples) are amplified and added to the original cello sound, they can provide a *third voice* to a two-part cello piece performed on stage (for details [31.294]).

31.6.4 Fusion, *Verschmelzung*, Consonance

One of the fundamental experiences of musical sound is that of several pure and/or complex tones presented simultaneously. A multiplicity of sounds emanating from one source or from several sources in regard to vibrational patterns and resulting sound structure either blend well or rather dissociate; the degree to which sounds *fit together* depends primarily on acoustic parameters such as frequency ratios and phase relationships [31.93, 159]. In regard to listeners, mixtures of sounds can give rise to sensations of roughness or beats on the one hand, which have been identified as characteristic of dissonance (Sect. 31.6.2); on the other, sounds blending well lead to a stable temporal (periodicity) and spectral (harmonicity) patterns in the auditory system, which are fundamental for a perceptual quality known as *consonance*. According to a definition given by Boethius (around the year 500) in his *De institutione musica* (Lib. I, 8, ed. Friedlein), consonance is the mixture of a high and a low tone that arrive at the ears pleasantly and uniformly. This experience was known already to Greek theorists (for source material see [31.295, 296]) experimenting with concords on string instruments. Ptolemaios, writing in the second century, grouped tones into three classes labeled homophonic, symphonic, and emmelic. Homophonic, he says (harm. Lib. I, 4, 7, ed. Düring [31.297]), are the octave and the double octave since two tones at these intervals give the impression of one tone, which implies hearing the intervals as simultaneities. Symphonic are the fifth and fourth as well as their extensions into higher octaves. Emmelic are those tones which, presented side by side (perhaps in succession, as melodic steps) are found pleasing to the ear (pairs that don't are considered ekmelic). Several Greek theorists use the term *krasis* (κρασις [31.295]) which can be translated as mixture, to denote the specific quality of consonance.

Stumpf, in his theory of consonance, adopted from Greek music theorists the aspect that two tones presented simultaneously at distinct intervals can blend to the extent that listeners take them for one. However, Stumpf's theory of *Verschmelzung* includes a range of additional considerations in regard to possible causes and psychological effects of consonance. Translating the German term *Verschmelzung* as used by *Stumpf* [31.40, 298, 299] into *fusion* tends to narrow the concept, which was intended to implement a psychological theory of consonance expanding the psychoacoustic approach taken by Helmholtz. According to *Helmholtz* [31.28, Sect. 10], a sensation of consonance results from such intervals of complex harmonic tones that have partials in common. For intervals in just intonation (JI), partial frequencies coincide due to small integer frequency ratios as are realized, in a perfect

match, for the octave (2 : 1), double octave (4 : 1), and the twelfth (3 : 1) as well as nearly so for the fifth (3 : 2). For Helmholtz, consonance requires hearing two concurrent harmonic complexes free from beats. Since the relative amount of beats and, correspondingly, sensation of roughness increases with harmonic ratio (being somewhat greater in 6 : 5 than in 5 : 4, and markedly so in 9 : 8 as compared to 5 : 4) as well as with tuning away from just intervals (as in ET), scaling of intervals on a dimension consonance–dissonance can be executed, by listeners, based on their sensation of beats and roughness either being absent or present to varying degrees. *Helmholtz* [31.28, p. 291] gave the following order due to coincidences of partials:

1. Octave 1 : 2
2. Twelfth 1 : 3
3. Fifth 2 : 3
4. Fourth 3 : 4
5. Major Sixth 3 : 5
6. Major Third 4 : 5
7. Minor Third 5 : 6.

According to Helmholtz's classification, octave, twelfth and double octave constitute *absolute consonances*, fifth and fourth are *perfect consonances* (as in Greek and medieval music theory), major sixth and major third are *medium consonances*, and minor third and minor sixth (5 : 8) are regarded *imperfect consonances*. Helmholtz acknowledged that the natural minor seventh 4 : 7 also bears to a consonant sensation, often more so than the minor sixth.

Stumpf [31.40, 298, 299] challenged the explanation of consonance offered by Helmholtz as being both incomplete and not free from errors. First, he argued that consonance or dissonance is experienced also for pairs of pure tones (as well as for certain chords made up of pure tones) where coinciding or beating partials do not come into play. Hence, the criterion of coinciding or beating higher partials does not hold for such cases, and cannot be the decisive basis for either consonance or dissonance perception. Second, Stumpf insisted that consonance and dissonance are also registered if two tones forming an interval are presented dichotically (one tone to each ear) so that there is no interaction of partials on the BM level, leaving consonance and dissonance for a *central* analysis stage. Third, with the decisive processing moved away from the inner ear to the brain, Stumpf stated consonance and dissonance are psychological experiences that have a quality of their own. Even if sensation of consonance and dissonance relate to acoustic features (such as patterns of spectral partials), these are merely foundations for the perceptual experience that, in regard to consonance, is peculiar with respect to *Verschmelzung*, a concept that, most of all, refers to the relations several tones can have in acts of perceiving.

Stumpf saw sensation, perception, and cognition combined into a hierarchical processing that leads from sensory input to apperception, the latter focusing on relational structures between individual elements as well as complexes of elements [31.300]. In his *Tonpsychologie* [31.39, 40] as well as in other works, Stumpf elaborated on parameters and criteria governing perception, which almost always includes judgments with respect to the structural composition of sound and music stimuli. One particular judgment aims at the appearance of a multiplicity of elements in pairs of sounds forming musical intervals. If such pairs give rise to a percept that implies structural unity and coherence, the elements tend to be integrated in perception into a Gestalt-like configuration. The corresponding experience is what *Stumpf* [31.40, pp. 65 ff., 128 ff.], [31.298] labeled *Verschmelzung* (for a detailed account of the concept see [31.301]). Though Verschmelzung is a scalable perceptual quality [31.40, 298, 302], it is not simply that of *fusion* since the elements making up an interval or chord remain to be noticed individually, in analytic listening, and they may be perceived as interpenetrating each other (*Stumpf* [31.303, p. 261] himself regarded *Verschmelzung* as *Durchdringung* = interpenetration of several constituents). Correspondingly, the *whole* resulting from a combination of (two or more) pure or complex tones is not unstructured as a sum yet is perceived as a coherent configuration of elements (such as in a geometrical figure or in a crystal lattice). Stumpf considered such structures that can be apprehended by musically skilled listeners as a *unity of relations* (German: *Verhältnisganzes*). Hence, even though *Verschmelzung* is experienced instantly and immediately (in holistic listening), subjects become fully aware of the elements blending into *Verschmelzung* through analytical listening and cognitive assessment (apperception, apprehension) of what has been perceived. To experience consonance of pairs of tones, chords, or more complex sonorities, subjects may be switching back and forth between analytic and holistic modes of listening. Holistic listening thereby should provide a salient percept of parts matching perfectly while the analytic mode leads to recognition of the individual parts and their relational structure within a coherent configuration. *Stumpf* [31.299, p. 394] offered a striking example of an *ideal concord* (*Zusammenklang*) consisting of five perfect major triads (4 : 5 : 6) played simultaneously at root frequencies of 100, 200, 400, 800 and 1600 Hz respectively. If synthesized from 15 pure tones, the waveshape plotted in Fig. 31.15 results.

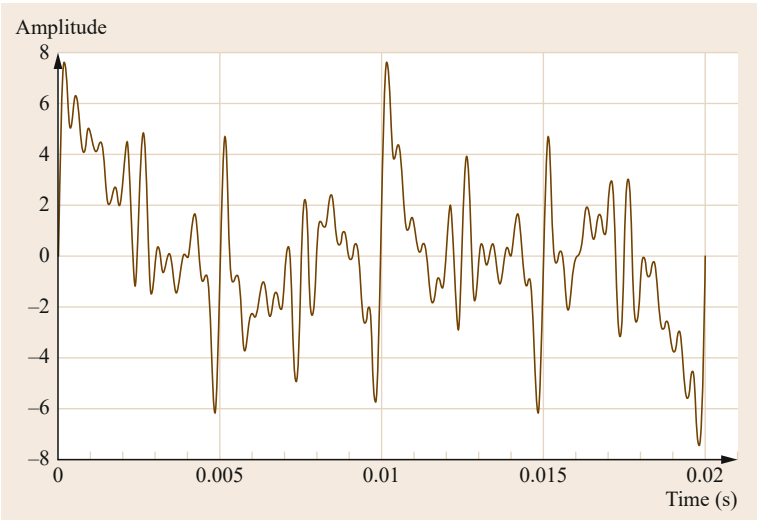


Fig. 31.15 Stumpf's *ideal concord* (5×3 harmonic components = five just triads)

As an inspection of the waveshape reveals, the main period is at 10 ms, and a second period at 5 ms. In an ACF analysis, in addition to these two candidates for periodicity pitches (reinforcing spectral pitches), a number of subharmonics below 100 Hz are detected. Listening to Stumpf's concord, one can try to identify its constituents as well as attend to the mixture or *fusion* of the 15 components which, to be sure, are not partials within a single harmonic series. For maximum fusion of partials, one can generate symmetric pulse-type sounds [31.301] that comprise very many harmonics covering the frequency range from 100–200 Hz up to 22–25 kHz. Such sounds lead to a maximum sensation of consonance since dozens of harmonics *fuse* or blend into another (in holistic listening). It is of musical interest that, in Baroque organs in particular of northern Germany, organ builders created various reed stops where the reed mechanism produces pulse sequences that excite up to, and even above, a hundred harmonics per tone played (the absolute number depending on the type of reed generator, the geometry of the resonator, and on register [31.304]). Reed stops as well as mixture stops in which also very many harmonics are audible for each tone played (in northern Germany, a stop known as *Terzzimbel* offering pure major thirds besides fifths and octaves was much in use) convey the experience of *Verschmelzung* to everyone listening.

As these examples demonstrate, consonance, experienced as *Verschmelzung*, does have an objective basis in the sound structure as well as in auditory processing. Sounds containing many harmonic components can be combined into chords, which in JI yield structures that again account for maximum periodicity and harmonicity in the signal as well as in peripheral auditory analysis [31.250, 301, 305, 306]; for audio demonstrations see [31.307]. In this regard, perceiving consonance is a natural phenomenon [31.26] and not something learned in the course of life (as has been suggested from a behaviorist perspective). Considering objective bases of consonance, various theories of consonance focusing on different aspects can be integrated into one. The criterion emphasized by *Helmholtz* [31.28], namely coincidence of harmonic partials and presence or absence of beats, is certainly of relevance since subjects according to some studies [31.308] prefer just intervals and, as explored in other investigations [31.309, 310], tend to accept tempered intervals within a range where the beating is not annoying. Taking into account that, in music, complex harmonic tones are predominant, it can be said that consonance of two or several such tones is maximum if these are strictly periodic in their waveshape, implying perfect harmonic spectra. If several such tones with their f_1 at small integer ratios and spectral envelopes composed like $A_n = 1/n$ (n = harmonic number) are combined phase-locked (onsets are identical) into a major chord, one will observe coincidence of partials, the criterion emphasized by *Helmholtz* [31.28] as well as by *Stumpf* [31.299]. Underpinning the same criterion, *Husmann* [31.311] reported that perception of consonance remains when pairs of harmonic complex tones are presented dichotically, presuming a *central spectrum* representation to process coinciding partials. When two or more harmonic complexes (with partials phase-locked) tuned to JI are presented diotically, strictly periodic excitation patterns are likely to occur on the level of the BM/CP and the AN. It is hardly surprising to find periodic input yields periodic excitation and neural responses [31.225, 312]. The criterion of converging pulse patterns to account for consonance had been introduced as a hypothesis in 19th century works

on music theory and psychology [31.313, 314]. In addition, harmonic complexes presented at intervals with small integer frequency ratios would be the condition suited for combination tones to occur even if these might be of minor importance. Moreover, chord structures in JI are suited to generate virtual pitches, one of which is clearly marked as the *root* of the chord (in the sense of Rameau's *basse fondamentale* [31.305]). In sum, there are several concurrent factors relevant for sensation and perception of consonance suited to demonstrate that consonance basically is a natural phenomenon. It relates closely to periodicity and harmonicity, two elementary principles found governing structures and processes in almost all areas of nature [31.315]. In regard to signal and systems theory as well as acoustics, these principles are at the core of the Wiener–Khintchine theorem (Sect. 30.2), which in turn can be used to model pitch perception as well as sensation of consonance. As a matter of fact, three harmonic complexes (each comprising the first seven harmonics) combined so as to realize a major chord 4 : 5 : 6 in JI are enough to give rise to maximum *fusion* and the most salient pitch configuration possible [31.250]; for similar considerations see [31.316]. To be sure, the principle of harmonicity as expressed by small integer ratios that govern spectral partial structures as well as formation of chords has nothing to do with *number mysticism* (the well-known phrase put forward by behaviorists to denounce *natural causes* relevant for music perception). The two principles fundamental to science, namely counting and measuring [31.317], can help to understand also sound and music (for applications see [31.159, 318, 319]). Stating a nativist theory of consonance does not imply that *harmony* in music can be, or should be, completely naturalized (rather, it needs to be given natural foundations). However, taking the interdependence of acoustics, psychoacoustics and musical structures into account, it makes sense to realize chords in music in appropriate tuning and spectral structure so that, for instance, chords intended to be perceived as consonant can be transformed into musical sound with maximum spectral harmonicity while minimizing roughness and beats ([31.318] with audio examples on CD). Conversely, dissonant sonorities thus can be realized with appropriate fine tuning of pitches and spectra. In this respect, a close correspondence between musical structure, actual sound profile and perceptual qualities will be established.

31.7 Categorical Pitch Perception, Relative and Absolute Pitch

In the course of Chap. 30, the difference between discriminable and identifiable stimuli (or stimulus features) has been stressed (Sect. 30.1.4). It is a distinction that was of prime importance in debates concerning *continuous* versus *categorical* perception that became an issue, most of all, in linguistics and phonetics where it was discussed in regard to phoneme boundaries (for background information [31.320] and [31.111, Chap. 9]). In a strong formulation, it has been stated that (emphasis theirs):

> *categorical perception refers to a mode by which stimuli are responded to, and can only be responded to, in absolute terms. Successive stimuli drawn from a physical continuum are not perceived as forming a continuum, but as members of discrete categories. They are identified absolutely, that is, independently of the context in which they occur.* [31.321, p. 234]

These authors further argued that subjects can discriminate between pairs of stimuli drawn from different categories *but not between stimuli drawn from the same category*.

The doctrine of categorical perception has spun debates in many areas of research (for an overview see [31.320] and a range of articles in [31.322]) including music psychology where it was discussed in conjunction with other concepts relating to tone systems, scales, and in particular intonation. An analysis of relevant publications reveals there are some obvious misconceptions in regard to theoretical and factual issues, which need some clarification. This concerns, first of all, the terms *category* and *categorization*.

The term *category* is often used synonymously with *class*, meaning that a number of n objects are sorted into m *categories* or classes ($n \gg m$) according to certain features or properties. This, however is just a description of a process that can be quite mechanical (like sorting many things into a few drawers) while *category* has a more profound meaning in areas like epistemology and ontology as well as in mathematics and linguistics. In regard to subjects building up coherent experience, *categorization* and *classification* involve judgments [31.323] even though these can be executed in a more or less *automated* manner [31.324, Chap. 2]; also [31.325, pp. 46 ff.]. Such judgments may pertain to, for example, segmentation of quasicontinuous sound streams we perceive (as in speech or music) into *units* that have a syntactic function and/or semantic meaning [31.111]. Segmentation processes involve sensation

and perception as one hears a speech or musical sound that is then analyzed as an acoustical signal, mainly so in bottom-up processing; however, segmentation of sound streams is more than finding transitory and quasistationary sections since subjects make use of their previous experience and recall both meaningful *units* (elements that are of relevance with respect to phonology, prosody, grammar, syntax, semantics) as well as rules of how such units should be connected sequentially or (regarding music) simultaneously. Cognitive processing of both speech and music involves memory and is achieved, in the main, in a top-down approach. The sensory-driven part and the cognitive part are combined in a complex processing model described in detail as *auditory scene analysis* [31.326].

Categorical perception in this respect involves sensory-driven analysis of acoustical signals combined with judgments pertaining to identification of *units* as well as an assessment of how they relate to other elements that occur within a given structure and context. Such assessments can be facilitated by referring to prototypes [31.324, 327] or by activating certain schemata that may be innate or may have been acquired in processes of learning [31.328–330]. However, in a quasi real-time process such as listening to music the rate of information available to listeners per time unit can be very high, to the effect that perceptual and cognitive processing is very demanding even for subjects with musical training. It has been calculated that, for example in Ligeti's *Atmosphères*, perceiving all of the very short notes distributed over many instruments in the score would come close to a flow of information of 719 bit/s [31.331, p. 72], which is several times that of the capacity of the auditory channel (estimated to allow for processing of about 100 bit/s [31.332, p. 56 ff.]). Even though the concept employed by Ligeti, in *Atmosphères* as well as in other so-called micropolyphonic works was not to have listeners perceive all the individual voices individually or even the very many notes these comprise (but, rather, the resultant textures and sound fields [31.333]), perception under the constraint of real-time processing, to be achieved successfully, requires the *information load* not to exceed the channel capacity, and possibly to be reduced in suited ways. There are several concepts of how information processing in human subjects can be optimized, for example reduction of the perceptual dimensionality of stimuli, a filtering process directed at finding distinctive features in the sensory input, and selective attention and suppression of *irrelevant* information [31.111, 326, 334, 335], [31.69, Chap. 21]. Also *categorical perception*, meaning that a continuous dimension is segmented into a rather small set of discrete categories into which a broad range of stimuli might be sorted with respect to certain features, would be an effective means to reduce the perceptual load.

Both speech and music are communication modes operating in the time domain so that, to facilitate successful transmission of information, a limited set of categories both on the side of the *sender* as well as that of the *receiver* seems useful. In order to make acoustic communication like speech or similar utterances possible, meaningful units must be distinguishable and, at least in principle, identifiable. This is why phonemes are different in spectral content (and often also in modulation parameters [31.336]). Differentiating such units by their acoustic cues can be achieved not only by humans but also by other mammals [31.337]. However, boundaries between phonetic categories are to some extent flexible [31.338]. Categorization in speech apparently is very much sensitive to context and adaptive in nature (for an overview see [31.339]). A serious problem that has been tackled in computer models of speech processing is recognition of allophones, that is, of sounds that are used to express a certain phoneme in a certain language but which differ slightly in temporal and spectral fine structure when spoken by many individuals (men, women, children). Though allophones can be distinguished to some extent, they must be close enough not to introduce phonemic boundary effects and, thereby, differences in meaning. The concept of *allophone* has been taken as an analogy for intonation variants in music [31.340, 341] in regard to categorical perception, implying that the scale steps in music would be equivalent to phonemes in speech. Comparing speech to music perception, things are different in many respects notwithstanding some convergent features. In the following, some of the points raised in articles on *categorical perception* of pitch [31.340–347] will be discussed. For this reason, a few more considerations relating to sensation and its measurement (Sect. 30.1.4) need to be included.

Sensation and perception are often viewed as processing of input information whereby the input is distributed along one or several dimensions. By definition, dimensions require effective parameters to be variable continuously between a lower and an upper limit value. In regard to pitch, the single parameter taken into account in some concepts is frequency. Since frequency, as a physical quantity, can be varied continuously, stimulation of the hearing system can be assumed to be continuous (at least in principle). However, even if excitation on the BM/CP level would match this characteristic, sensation measured psychophysically as a reaction to the stimulus [31.348] is discrete because of JND and DL thresholds (Sect. 31.6.1). In this respect, the discrimination process leads to segmenting the stimulus continuum into small (yet not infinitesimally small)

steps. Stimuli that can be discriminated still may appear to be quite similar as are, for instance, two pure tones of 1000 and 1005 Hz in tone height if presented at the same SPL. *Stumpf* [31.39, 40] discussed *similarity* as a general psychological concept that subjects make use of when comparing tones as to their *height*, their perceived loudness, or when evaluating some other attribute. *Stumpf* [31.39, pp. 111 ff.] defined equality of phenomena as sensed by subjects as *extreme similarity*, whereby, for a number of complex objects to be compared, similarity can arise from (a) equal ratios among parts or elements, or from (b) equal or identical parts found in several objects. Similarities found on the level of (a) and of (b) can add up to the *similarity of the whole*. The point that is decisive here is that similarity is scalable and can be translated into geometrical distances; put the other way round, a comparison between two stimuli with respect to the sensations they bring about involves comparisons of qualities, intensities, or of temporal and spatial parameters so that similarities are conceived as distances (a high degree of similarity means a small distance [31.93, pp. 404 ff.]). Obviously, these considerations are very close to those found later in concepts of perceptual scaling of similarity, and of multidimensional scaling (MDS) in particular [31.349, Chap. 11] and [31.348, Chap. 6]. Judgments pertaining to similarity seem to be necessary for perceptual and cognitive generalizations ([31.350] and the discussion in [31.351]). Categorization based on sensory input and perceptual processing can be viewed as a necessary part of cognitive generalization and, thus, part of the ongoing process each individual undertakes to acquire coherent and systematic knowledge (for epistemological implications see [31.352–354]).

Because of cognitive generalization, subjects tend to group stimuli that are discriminable yet which appear to be similar in view of certain features or parameter settings, as belonging to the same *category*. In regard to sensing tone height and pitch perception, small changes in frequency as the (sole) physical stimulus parameter bring about sensations of tone height (as well as brightness) that are still quasicontinuous since the increments are usually close to the size of JND/DL. With such stimuli, subjects can make judgments as to the difference of two tones (e.g., the second is higher or lower than the first) without being able to estimate (let alone precisely judge) the size of the difference. Pairs of pure tones that are just discriminable in tone height and brightness will hardly be perceived as representatives of different categories but rather as variants of the same basic stimulus. Hence, differences subjects sense between two tones in *height* or in another attribute in general must exceed a certain margin large enough to be considered as introducing a new *category*, which may be taken as a difference in perceptual quality. In this respect, a growing difference in stimulus parameter values and, correspondingly, sensation can be accompanied by perceiving increasing phenomenal dissimilarity between pairs or series of stimuli. In the case where the degree of phenomenal dissimilarity perceived from comparison of pairs or series of stimuli exceeds a certain limit, a transition from one perceptual category into the next is likely to occur. Categorization, from a perceptual and cognitive perspective, aims at recognizing stimuli as *members* or as *representatives* of a certain category. For classifications in general it is assumed that the members of each class are equivalent in regard to certain features and properties notwithstanding slight differences they may have in actual parameter values. Concerning music perception, the categories relevant for pitch typically are manifest as steps on a scale (see below). In a strict view pursued in speech studies, *categorical perception* means identification of certain units (e.g., phonemes) by both recognition and labeling (imagined as mentally putting a *tag* on each item perceived).

The three levels referred to above can be related to each other thus:

1. Physical stimulus: Continuously variable (e.g., frequency in Hz)
2. Sensation: Quasicontinuous (small increments, measurable in JND/DL thresholds)
3. Perception: Discrete *categories* (each comprising a range of stimuli that differ by small increments in parameter values but are perceived as being more or less *similar* with respect to certain features or properties).

The problematic point now is that, according to the strong formulation of categorical perception [31.321, p. 234], subjects are said to be unable to discriminate between stimuli drawn from the same category. Taking, for example, the steps of a chromatic musical scale (in ET) as nonoverlapping *categories*, this would mean that, since fulfilling the classificatory criterion of *intraclass equivalence*, every tone played in a range of about ± 40 cent around the fundamental frequency of a certain scale tone (say, $A_4 = 440\,\text{Hz}$) would be a suitable representative of that scale step. In fact, such views have been issued in a number of publications, and have been mixed up, in some cases, with another concept leading to very broad interpretations of the musical scale step viewed as a *category*. The concept in question is the well-known *law of comparative judgment* proposed by *Thurstone* [31.355, 356] as an expansion of the theorems issued by Weber and by Fechner. This law states, in short, that an observer, judging the JND in intensity (or a qualitative difference) between two stimuli

in what Thurstone calls the discriminal process is not consistent on successive trials. Rather, the sensations of the observer and their judgments are viewed as intrinsically variable so as to make up a distribution of data with a mean and a standard deviation for each stimulus. Comparing two stimuli, S_1 and S_2, a number of times, the judgments one observer obtains based on his or her sensations can be considered as two random variables forming two dispersions [31.349, Chap. 8 and 9], [31.357, pp. 13 ff.], [31.358].

Aspects of categorical perception and of Thurstone's scaling model have been instrumental, often implicitly, in explaining variances observed in pitch perception and intonation. As to the former, it has been argued, for example, that pitch perception must be categorical because *pitch* would be unidimensional as a stimulus property (a function of frequency), and identification for unidimensional stimuli would be restricted to very few *categories* on a *stimulus continuum* (consisting of arbitrary frequencies). As a *proof* for such assertions, Miller's well-known *magical number* 7 ± 2 [31.359] has been quoted. As a matter of fact, such assertions are invalid on several counts. First, Miller's rule concerns unidimensional stimuli. Even if one would be willing to accept the rule, which Miller derived from a few experimental reports from the early 1950s, it could not easily be applied to pitch since pitch is neither unidimensional in regard to musically relevant stimuli nor takes perceptual attributes of pitch into account. Even with pure tones, there are several sensory attributes (tone height, brightness, also sharpness, vocality, density, volume) varying with frequency [31.40, 59, 299]. In fact, there are no unidimensional stimuli in musical contexts. As had been found in early experiments using tones varied on several dimensions (frequency, intensity, duration, etc.), the average information transmission was about six bits per stimulus [31.360], which yields $2^6 = 64$. Though the elementary stimuli and the conditions employed in these experiments fall short of musical material (like complex harmonic tones presented in a melodic phrase or chord), it is obvious that even for stimuli offering a *modicum* of variation on several dimensions the information is sufficient to identify more alternatives than implied by the 7 ± 2 rule (which itself calls for a more critical assessment [31.361]). Second, pitch outside such constructs as the *mel scale* (Sect. 30.1.4, *Loudness Scaling* ...) is not a continuous random variable. Even if two pure tones are presented so that one is fixed in frequency and the other continuously glides up or down, there are perceptual discontinuities marked by small integer frequency ratios, the most obvious of which is the octave (Sect. 31.1.2). Subjects with some musical training register not only octaves but also frequency ratios corresponding to other basic intervals (like the fifth, fourth, major third [31.40, 298, 362]. These intervals have a Gestalt-like quality and are identified perceptually if the frequency ratio is realized within certain limits. Third, in identifying such intervals, acoustic cues (consonance and *fusion* versus roughness and beating) come into play [31.308, 310]; though these cues are predominantly effective in tone pairs presented simultaneously, judgments on successive intervals can draw on perceptual information available in STM [31.218] when comparing tonal relations as well as on schemata or templates stored in LTM [31.328–330].

It should be clear, from what has been discussed so far, that perception of pitch in fact is categorical in that the most fundamental musical intervals function as categories; however, these must be viewed taking acoustical properties as well as perceptual categorization and cognitive structuring of tonal relations into account and not as mere *baskets* of classification. The concept of *pitch class*, well known from theoretical works on 12-tone and serial music, in this respect is problematic since it openly or implicitly equates musical pitches with the keys of a piano tuned to ET12 so that there are but 88 pitches falling into 12 classes, octave registers and enharmonic spellings being ignored. This, to be sure, is a rather crude simplification that does not reflect interval recognition capabilities and apprehension of chord progressions of subjects experienced in performing and listening to tonal music.

Assuming categorical perception of pitch would be restricted to sorting tones as one hears or produces in, for example, singing or playing a violin, into 12 *pitch classes*, one would in fact expect 12 *frequency baskets* of equal extent, namely 100 cent each, little (if any) overlap at the boundaries and a more or less flat (platykurtic) distribution of data within each ET12 *pitch class*. In a simple graph, the distribution of pitches produced and/or perceived, for three consecutive classes, would look like Fig. 31.16.

Dispersions in this model have the same shape (reflecting the 12-tone gospel that all chromatic tones are of *equal importance*), and there is an equal likelihood for pitches falling into the frequency range under the flat part of each class while there are also equal boundary transitions between adjacent classes (implying the psychometric function would be equal for all pitch classes and boundaries also).

However, data from both intonation analyses of tonal music and experiments on perception and production of musical intervals reveal that this model does not apply. First of all, the *pitch categories* that are actually used in sequential listening or melodic execution of intervals are not of equal size but vary with the structural weight of tones in a scale. In experiments, categori-

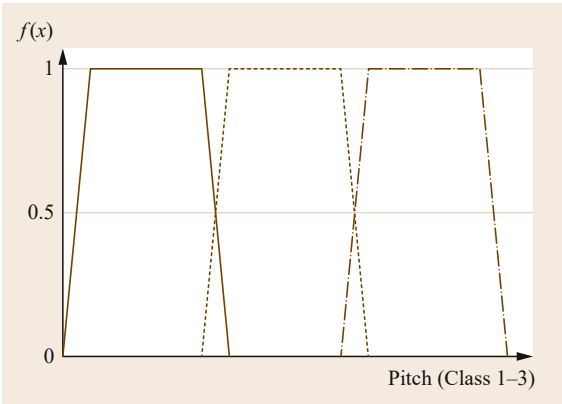


Fig. 31.16 *Pitch class* model of pitch perception and production

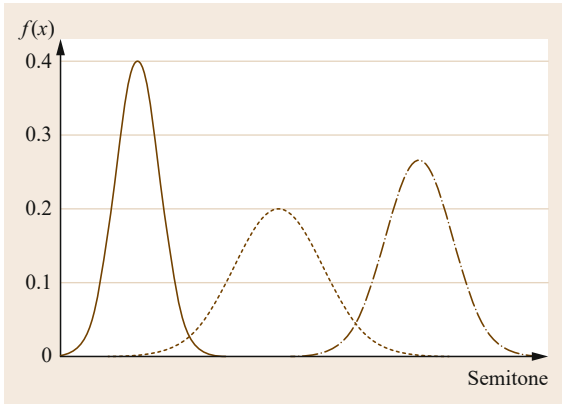


Fig. 31.17 Scale steps and *pitch categories* for three semitones viewed as PDF

cal identification of intervals yields curves of different width and smoothness, reflecting that some scale steps and melodic intervals are structurally more stable than others [31.340, p. 64, Fig. 4]. The same picture emerges in measurement of intonation where the range for structural important intervals such as the fourth, fifth, and octave in general is significantly smaller than for notes falling in between. For example, in a detailed study of professional violinists playing the Theme and Variation X from Paganini's Cappricio No. 24 in a-minor, *Janina Fyk* [31.92, Chap. 8, p. 185] found relatively small dispersions of about 25–30 cent for the perfect fifth, the minor sixth, major sixth and the octave for variation X (played at a slower tempo than the theme) while the dispersion for the major third and the perfect fourth were 39 and 40 cent, and the minor second and major second had 53 and 60 cent respectively. It is also of interest to note that, in various experiments, subjects distinguished between the augmented fourth and the diminished fifth [31.363, p. 133], [31.90, 91, 341]. The pitch difference between these intervals is ≈ 22 cent in both JI and Pythagorean systems (though in opposite direction!).

Taking differences of scale steps and intervals in regard to structural importance and actual dispersion data from perception and intonation measurements into account, an adequate model of *pitch categories* could be devised in terms of probability density functions (PDF) where stable categories (i. e., scale steps serving as perceptual and cognitive anchor points like prime, fourth, fifth) gain narrow (leptokurtic) dispersions, and those in between wider dispersions. In this way, Fig. 31.17 may be viewed as representing, for instance, the first three semitones of a scale (e.g., c–c#–d).

By the way, this model conforms to theory and practice of music in cultures where scales and modes are fundamental structures (as in Greek, Persian, Arab, Indian, and Turkish traditions). Typically, there are structurally important tones in scales framing intervals (octave, fourth, fifth) that can be filled with additional tones in a range of modal structures that in turn request certain intonation patterns (for theoretical background, examples, and empirical data from measurements see [31.52, 364–372]). The concept of *fixed* and *movable* tones in scales and modes is common to ancient Greek, Near Eastern, and Indian music traditions. One has to remember that stable tones corresponding to fundamental intervals within scales and modes of course have a melodic function but must be viewed also regarding types of drone accompaniment widely in use (the drone must not always be audible as sound but is implied in the performance of singers and solo melodic instruments as a referent). In *Western* theory and practice, tones besides melodic steps constitute harmonic relations even for tones played in succession; sensation and perception of consonance inevitably come into play since the STM *buffer* for pitches of tonal sequences is at least several seconds wide [31.218, 373].

Consonance is an important aspect that explains why different intonation patterns are not equivalent in realizing the same interval category. The quality of simultaneous intervals executed in music for musicians and listeners with musical training apparently depends on two factors, one being related to sensation of beating and roughness versus *fusion* and smoothness of partials in two or more complex tones, the other with an estimate of the size of the distance between tones [31.308–310]. The range within which intervals are judged as acceptable in most cases is ≈ 15–20 cent above and below the just ratio (like 3 : 2 or 5 : 4) while deviations greater than this range are perceived as mistuning [31.374, pp. 100 ff.]. We found in experiments where cadences were played in different tunings (Vallotti and Young, ET, JI, Meantone, Pythagorean) to

samples of students significant differences in ratings of intonation quality and of consonance [31.375, 376]. In particular, Vallotti and Young tuning received much higher ratings than tuning chords to the Pythagorean scale which, of course, was conceived for *melodic* execution as in the Dorian mode of ancient Greek music. Greek Dorian contains two whole tones $9/8 \approx 204$ cent and a *leimma* (Greek: *remainder*) of $256/243 \approx 90$ cent in each tetrachord; these are joined by another whole tone $9/8$ [31.52, pp. 183 ff.]. A melodic C-major scale derived from a chain of fifths [31.72], like $f-c-g-d-a-e-b \rightarrow c-d-e-f-g-a-b$, contains the same intervals as the Dorian mode and can, therefore, be played in a Pythagorean tuning (as in fact was observed by [31.89]; the violinists he investigated playing C-major scales came fairly close to Pythagorean tuning). However, for simultaneous intervals and chords, Pythagorean pitches are not recommendable since the major third of $81/64 \approx 408$ cent gives rise to beating while the major third in Vallotti and Young is close to 392 cent (about 6 cent wider than the just major third $5/4$) and quite *smooth* in regard to consonance. Since the difference between the major thirds in Vallotti and Young versus Pythagorean tuning is ≈ 15.5 cent, this gives a measure of the sensitivity subjects have for the consonance of major thirds in chords. For skilled performers, adjusting pitches within long held chords is common practice as has been demonstrated for barbershop [31.377] and madrigal ensembles [31.378], [31.305, pp. 66–67]; in an intonation test performed with two barbershop ensembles singing cadences, in the dominant seventh chord they produced almost the *natural seventh* $7/4$ (with, on average, 977 cent). In madrigal singing, one often finds frequency ratios in long held chords coming close to JI values [31.378]. Subtle adjustments and distinctions can be observed also in solo violin playing. Following observations according to which expert listeners were able to distinguish an augmented fourth and a diminished fifth with respect to correct intonation [31.341], *Fyk* [31.92, pp. 96 ff.] had two professional violinists play two four-tone melodic phrases, (1) $f_4-b_4-c_5-e_4$ and (2) $f\#_4-c_5-b_4-g_4$ with and without vibrato. The first melody can be conceived as a sequence IV–VII–VIII–III in C while the second is viewed as VII–IV–III–I in G. The intervals to be played then are (1) an augmented fourth and a minor second upwards, followed by a minor sixth downwards and (2) a diminished fifth upwards followed by a minor second and a major third downward. The intervals in cents averaged over many trials for both conditions show distinctive deviations from ET and also from JI and Pythagorean tuning as well as fair to close approximations to pitches (expressed as fundamental frequencies of complex harmonic tones as were recorded from the violins) defined by one or several of these tuning systems. It seems that violinists relate to several criteria guiding their intonation of intervals in melodic sequences, among them being tonal function of the tones within a scale and/or tonality, direction of interval (upward, downward), harmonic tension and leading note, expressive (vibrato) or straight intonation. Some of the pitches expressing basic intervals are more stable than others when measured over several trials; the range in such cases is small, and typically within $\approx$ 10–20 cent [31.92, pp. 110 ff.], [31.89]. Recent observations from professional violinists showed their pitch adjustment to tones modified in fundamental frequency is within a very small range ($2 < \mathrm{df} < 7$ cent), and includes adjustments performed even when the change of the stimulus frequency was below the perceptual threshold, suggesting an automated motor response probably independent from conscious perception [31.379].

The Russian musicologist *Nikolai A. Garbuzow* studied relative and absolute pitch and conducted many pitch and intonation measurements in the 1940s and 1950s (published in articles and monographs collected in [31.344]). From his observations and measurements, which covered not just pitch and intonation but also loudness, tempo, and rhythm, he came to propose that human hearing in general is of a *zonal nature* [31.344, 380]. In regard to intonation (in particular, of violinists) he found a considerable range of deviations from tuning systems (ET, JI, Pythagorean), however, many of the actual pitches played relate to the melodic and interval structure of the works he had selected as well as to phrasing and expression chosen by the musicians studied. Garbuzow also found that subjects were able to distinguish between several *shadings* or small steps within a *tonal zone*. In this regard, pitch is not so much a random variable (in the sense of a Thurstonian or similar scaling model) than it is a choice among reasonable alternatives. In a musical context, professional musicians *can* distinguish sharp from flat within pitch categories. Hence, a tonal zone should not be mistaken for a *pitch class* within which all selections would be *equally appropriate* for performers and listeners due to an assumption of intraclass equivalence of pitch realizations. Evidently, the difference between an augmented fourth and a diminished fifth matters to musically trained subjects [31.363, pp. 129 ff.] even though on a piano tuned to ET12 this would be the same *pitch class* (6HT). Optimal sizes and admissible ranges of intervals in intonation for musicians who control their performance by ear and are not specifically guided to realize intervals relative to a certain tone system and tuning depends, as empirical data indicate, much on musical context and expressive execution. In this respect, intonation is variable, yet not arbitrary, within

pitch categories. In terms of phonology, one might view different intonations within pitch categories as *allophones* [31.341] or, in regard to tonal languages, as *allotones*. However, the concept of categorical perception as developed in speech and language studies at least in the strong version [31.321] cannot be readily applied to music and, moreover, has been challenged by findings that showed discrimination as predicted by this theory might hold for stopped consonants but does not apply to vowels [31.381]. Sure enough, pitch as a perceptual quality and parameter of musical intonation relates to periodic sounds (such as vowels).

From empirical evidence available, it can be concluded that the *pitch categories* relevant in music reflect scale as well as melodic and harmonic interval structures. These categories and their interrelations are based on several factors, among them being:

a) Acoustic properties and neural processing characteristics as are evident for octaves and seem to be relevant also for some other intervals, in particular, the fifth and twelfth. There is striking evidence for the use of the pure fifth and pure fourth in many music cultures, and the pure fifth has been addressed as an interval of distinctive quality forming the most basic consonance next to the octave since antiquity (for a detailed discussion with reference to historical source material see [31.72]).
b) The experience of sensory consonance, on the one hand, and of dissonance as connected with beating and roughness, on the other, which are relevant not only in the perception of the octave, pure fifth and pure fourth but also other intervals (major and minor third, major and minor sixth, and to a lesser degree the tritone and the minor seventh).
c) The factor of proximity that is relevant for judging the distance between tones, in particular of tones representing adjacent scale steps [31.55]. Proximity of tones in a chromatic scale often has been viewed in regard to *similarity* (see above), however, there are two kinds of similarity effective in the perception of scales, one based on a convergence of components (as in the octave, twelfths, fifths; see above), the other addressing the nearness of components with respect to JNDs and distance estimates (this issue, the double nature of *similarity* in regard to tones and scales, and to pitch relations in general, has been discussed by *Stumpf* [31.39, 40, 59] and in many publications since). The factor of proximity becomes evident, in regard to the *nearness* and *adhesion* of tones, when one hears a scale in a system with more than 12 pitches per octave, for example the 31-tone system calculated by Christiaan Huygens in the 17th century and realized, on various keyboard instruments, by the Dutch physicist *Adriaan D. Fokker* and some modern composers [31.382, 383].

Going from one step to the next (in this system, the difference in fundamental frequencies is $1200/31 = 38.71$ cent, a so-called diësis), the transition can be still perceived as a small step, and also various intervals remain discriminable and even identifiable both in succession and when played simultaneously. Discrimination can be achieved since the step size of nearly 40 cent is large enough to distinguish melodic steps and intervals by their size, and multiples of the basic step in fact are good to very good approximations to common musical intervals (e.g., two diëses give a chromatic semitone, three a diatonic semitone, five result in a whole tone (meantone of 193.55 cent), ten in a perfect major third, etc.). However, it would require much training to be able to identify all the intervals that can be produced in this system as such, in particular in a real-time listening and/or performing situation. Therefore, listeners not familiar with this system may take some of the 31 pitches as detuned variants of the limited number of *pitch categories* they know from their musical practice (very often, this is playing a piano or other keyboard tuned to ET12). Obviously, the number of *pitch categories* available to listeners and performers as well as the conceptual framework of how they relate to each other depends on a process of learning and training notwithstanding the fact that at least some of these relations have a *natural* acoustic and perceptual basis. Perhaps the most decisive factor for identification of tones and intervals is previous storage of categories along with their distinctive features in LTM, and fast access to this stored information in a listening situation or when performing music.

The ability to identify tones according to established *pitch categories* in a fast and seemingly effortless operation has often been addressed in reports on so-called absolute pitch (AP; for overviews concerning relevant phenomena and publications see [31.384–388]. AP capabilities, which include also identification of musical keys [31.385], are often divided into *active* and *passive* [31.389], the former pointing to a subject's ability to produce a tone at a certain pitch indicated by a tone name (for example, A_4 or F_5) by singing or whistling while the latter means a subject can assign a tone name to a given stimulus. For reasons of convenience and familiarity of subjects with piano tones, these were the stimuli used most frequently in AP experiments [31.384, 385, 390, 391]. In order to avoid complex sounds that would perhaps offer more than one cue for pitch identification, sine tone generators have been employed as well. In general, all frequencies

are measured relative to ET12 and a *concert pitch* of $A_4 = 440\,\text{Hz}$, and deviations from these values subjects realize in singing or in adjusting the frequency of a tone generator are given in cents. Quite often, deviations within a certain range (expressed in ± cents around one of the ET12 pitch frequencies) are considered insignificant since they do not violate the boundaries of that *pitch class* or *category* (see above for *intraclass equivalence*). Relevant boundaries may be taken at ±30 cent [31.344, 380] or even at ±40 cent, with only short *transit zones* between categories of 20 cent. Moreover, octave errors are not always considered as relevant (hitting the correct *pitch class* is regarded as showing *active* AP ability). A *passive* AP ability means a subject is able to identify a (pure or complex harmonic) tone that she or he hears by correctly labeling it (e.g., *this is* G_5). The tone name signifies its pitch relative to, typically, 12 categories that are representing a chromatic scale. Very often, the tones used in AP experiments were those from a piano tuned to ET12 so that enharmonic differences pertaining to flats and sharps are ignored both in the stimuli and in the *categorical* judgments prevalent in subjects with this type of AP. Since piano tones can be fairly well distinguished both in regard to register (octaves) and within octaves due to tone height, timbre and brightness as well as remembering different frequencies with which these tones are used in works of music and in musical practice, *categorical* pitch identification of piano tones for subjects having received years of piano training is not too difficult, and often results in scores above chance level. However, such findings may indicate *absolute piano*, as [31.384, p. 436 f.] and [31.391, p. 271] remarked in a review that is quite critical of AP concepts in general, rather than AP. In view of common practice, AP means subjects are able to either categorize tones they hear with respect to 12 *pitch classes* as well as several registers, or that they can produce a certain pitch, within the limits of a chromatic ET12 scale, according to a tone name that represents the *label* of a certain category. The adjective *absolute* in AP refers to a form of immediate categorical judgment that should be executed without (external and/or internal) reference and without making comparisons. However, though absolute judgments relate to objects in isolation (no context provided), this is a condition that might be unrealistic, in certain respects. As *Stumpf* [31.39, p. 139] has pointed out, judgments on pitch involving *absolute* identification of tones by labeling them (e.g., F_3 or $C\#_5$) can be done so that the actual stimulus is compared to a reference stored in memory; in principle, a few reference pitches or just one like the most familiar A_4 (plus the octaves one can imagine as further anchor points) should suffice. It has been argued that,

> *even in so-called* absolute *identification, the listener is trying to transform the task into one of* relative *pitch, seeking to anchor the presented tone to an available context.* [31.392, p. 192]

While in fact some basic relational frames stored in LTM might be governing much of AP judgments, any comparison of a stimulus with stored anchors or even an adjustment of a certain pitch a subject produces in *active* AP tasks to a tone name (*give me a* $F\#_4$) takes time. Therefore, reaction times have often been measured in AP experiments [31.393]; it can be hypothesized that, with a complete set of pitch categories stored as a template in LTM, reaction time for performing either *active* or *passive* AP tasks is very short while it increases when only a partial reference frame or even a single pitch reference is available for internal comparisons providing the basis for *absolute* pitch judgments as well as for production of certain pitches. Also, it can be hypothesized that the error score for subjects (see below) reflects accessibility of internally stored references as well as the precision with which such references can be recalled from memory. In this respect, AP tasks address cognitive abilities (as the term *absolute tone consciousness* used by *Abraham* [31.389] and other researchers on AP indicates), namely (a) the capacity to draw, in a fast and reliable way, on a set or template of *pitch categories* stored in LTM, and (b) a capacity of correctly labeling tones heard or imagined in correspondence to these categories. The term *category*, as explained above, in regard to pitch perception indicates a certain frequency range or bandwidth while, in a strict sense, one could also argue that perceiving *absolute* pitch should confine this range to that of a single JND. However, it has been observed in experiments comprising larger samples of subjects with apparent traits of AP [31.385, 390] that such subjects have access to rather broad pitch categories (usually spanning a semitone) but often do not perform so well in tasks requiring precise discrimination of pitch changes or tuning differences [31.394, 395]. In this respect, many (but not all [31.90, 91]) subjects with AP indeed seem to come close to *categorical perception* as predicted by its strong formulation [31.321].

There are several hypotheses regarding the genesis and/or acquisition of AP [31.388]. One takes AP as an innate and possibly hereditary property that is rare (at least among subjects of Caucasian origin), and is often affiliated with other traits indicative of superior *musicality* as well as with general intellectual alertness observed in a child such as reading music, memorizing complete musical works, playing one or even several musical instruments, and creativity in composing and

improvising music at an early age (for detailed reports on musical child prodigy, see [31.396, 397]). A reason to assume innateness of AP was that some subjects had or have AP abilities (both *active* and *passive*) at a very early age where training can hardly account for actual performance. Also, some subjects display astounding AP abilities in that their categorization of tones in regard to pitch is achieved in a very fast, precise (i. e., very low error scores [31.393]), and seemingly effortless way, whereby categorization may extend to various musical and even to environmental sounds (apparently, there are also other animals besides humans capable of *absolute* pitch identification). Such observations do not deny effects of learning and practice since early onset of musical training seems to be another decisive factor. However, training alone cannot explain actual musical performance of subjects exhibiting clear AP abilities at a very early age (often, between four and ten) who may be viewed as *AP possessors* or as belonging to the AP phenotype [31.398, 399] to be distinguished from subjects who have acquired AP abilities in years. Some data show an aggregation of AP possessors within certain families [31.390, 399, 400], a factor that has been interpreted as indicating a genetic predisposition for the development of AP.

A second hypothesis also assumes that there is some innate disposition for developing AP at an early age (the time window mentioned in many publications is from about three to six years of age); acquisition of AP abilities within this *sensitive phase* (or *critical period*) is viewed as depending on appropriate training aiming at identification of tones according to fixed pitches and pitch categories, a task that involves of course labeling or mapping tones heard. Even more demanding seems production of isolated tones on hearing or reading a tone name (e.g., B^{b}_{5}). Sometimes conventional music education, which has a focus on relative pitch and pitch relations rather than on memorizing fixed pitches, is regarded as a factor deteriorating the initial AP disposition. Though there is, in certain countries, a practice of using solfeggio with *fixed do*, that is, do is always representing the note C, children usually learn to sing or play simple tunes (from which they memorize melodic contours) as well as intervals and chords with their inversions that are fundamental in tonal music. The issue of acquiring a template of pitch relations effectively replacing an initial AP disposition seems controversial for several reasons. Viewed in evolutionary terms, there are some indications that, for instance, certain bird species are responding to both relative and absolute pitch cues concurrently [31.401]. Though there are some hints that young children may respond to absolute features of a melody such as the pitch of the first note or the pitch range, there are also indications that they respond to pitch relationships [31.402, pp. 159 f.]. This seems unsurprising given that basic harmonic intervals like the octave and the fifth have a natural acoustic and perceptual basis (Sects. 31.1–31.4) so that such intervals should be learned by children with ease. In fact, children take advantage of the pure fifth and the major triad when processing melodic interval structures [31.403]. The decisive factor for both the development of AP abilities [31.388] and of good relative pitch perception indeed seems to be an early onset of musical training and practice. To be sure, there are more perceptual and cognitive skills that benefit from early onset of training (like learning foreign languages) so that this condition is by no means exclusive to AP. It has been suggested, for language acquisition, that infants are able to *map* critical aspects of a language they hear into neural substrates so that being exposed to ambient language sounds could be sufficient to induce such mappings [31.404]. One might probably see a parallel for developing AP in children who are exposed to music and will encode the pitches and intervals they practice, often on a piano or other keyboard [31.66, 405]. The rivalry between AP and relative pitch (RP) might then be understood so that a mapping of absolute pitches in infancy could be an obstacle for the mapping of scales and intervals in terms of relative *fluid* pitch templates later.

Another fact that should be considered when assuming an original AP *disposition* would exist in many or all of us is the late introduction of *absolute* pitch in acoustics and musical performance practice. Though calculation of absolute frequencies and definition of reference pitches (by means of the length of organ flue pipes) became accessible, in round figures, in the 17th century [31.18], there was no precisely defined *concert pitch* before the end of the 19th century (for a historical review, see the collection of articles in [31.406] and [31.407]). By about 1885, several European countries agreed on $A_4 = 435\,\mathrm{Hz}$, and only in 1939 was an international norm (ISO 16) for $A_4 = 440\,\mathrm{Hz}$ settled. In retrospect, *absolute pitches* defined by frequencies (measured in cycles per second (cps) (Hz)) was neither available to, nor needed for, musical performance practice in Europe (and elsewhere) before the 19th century. Two factors that played a significant role in the process of standardization were the spread of ET12 as the predominantly used keyboard tuning (roughly, between 1780 and 1860), and the development of the modern piano (fortepiano) in about the same era. It is of interest to note that ET12, the modern piano, and AP concepts, which became of interest in music education and music psychology by the end of the 19th century [31.389], relate closely to each other. It is this conglomerate that led to many experiments and their interpretations in the realm of *absolute piano* [31.384, 391].

The third hypothesis plainly states, in a behaviorist stance, that AP is nothing but a learned ability, and that almost everybody can achieve AP with adequate training conducted over a period long enough to make 12 pitch categories accessible from LTM and, if needed, STM (including sensomotoric coordination in *active* AP tasks). A classical case study often cited to prove the hypothesis is that of *Paul Brady* [31.408] who managed, at the age of 32, to acquire AP abilities within three months, in a tough effort where he spent many hours on tone and pitch identification. Though his achievement seems remarkable, it can hardly be compared to the performance of AP phenotype subjects who can do with moderate effort in childhood what Brady could as an adult after heroic training.

In what seems a typical *nature–nurture* controversy, there is factual evidence for both sides; there are observations indicating some differences between AP *possessors* and *nonpossessors* in regard to cortical activation patterns and the role of working memory and associative memory [31.387, 388, 409] and there are data from surveys of AP possessors [31.399, 400] relevant for determining possible familial and/or ethnic aggregation of AP traits. On the one hand, there is evidence for a distinguishable AP phenotype; on the other, there are many reports on positive effects of musical training on pitch identification tasks in particular during the *critical period* [31.410]. Also, it seems evident that members of ethnic groups using tonal languages such as Mandarin Chinese can acquire AP more easily (since they have to learn and encode certain *pitch categories* fundamental to the use of their native tongue anyhow). However, in many respects, the differences between AP possessors and musically trained subjects with good or even excellent *relative pitch* (RP) seem gradual rather than *absolute*. In experiments where musically trained RP subjects are employed as a control group, their scores in pitch and key identification tasks often are not much different from those of AP subjects who, to be sure, also show typical errors and limitations in pitch identification and pitch production tasks. Those typical errors include [31.385, 390, 392, 411]:

1. Missing the correct identification of the pitch/tone name of a given stimulus (piano tone, sine tone, or other) by a semitone up or down, or even by a whole step up or down.
2. Mistaking the tone/pitch presented for a fifth or a fourth up or down the target.
3. Identifying the correct category, but not the correct octave.

In regard to limitations, there are several considerations:

1. Typical findings are that many AP subjects have low error scores only for a gamut of one octave or, at best, two octaves (performance of Caucasian subjects usually is most reliable in the range C_4–C_5, and often also good in the range C_3–C_4, but markedly deteriorates below C_3 and, increasingly with pitch and register, above C_5).
2. Error scores are low for the diatonic tones of a scale, and markedly so for a diatonic C scale (corresponding to the white keys on a piano) while error scores are clearly higher for chromatic tones marked by accidentals.
3. According to observations, performance is to some extent dependent on the timbre of the tones presented for pitch identification; error rates increase for many subjects when tones from instruments other than the piano (with which most of the AP *possessors* are familiar) are used.
4. A final source of errors that is of interest in regard to *categorical perception* is a change in the tuning in which tones or musical phrases are played. Using tones recorded from instruments, piano cadences and excerpts from musical works in major and minor presented in three tunings (tuned to $A_4 = 440\,\text{Hz}$ or detuned by either $+40$ or -40 cent), *Heyde* [31.385] found that the tuning of -40 cent was correctly identified in nearly 62% of the judgments when the musical excerpt was in a minor key while identification of a tuning $+40$ cent of tones and musical excerpts in major keys was much poorer (with an average score of only 28.2% correct).

A factor one might think of for explanation is that tuning of instruments and *concert pitch* in orchestras nowadays often is higher than $A_4 = 440\,\text{Hz}$, and rather close to 445 Hz; AP subjects might therefore be inclined to take the higher tuning for the *concert pitch* they experience with modern orchestras. However, raising the pitch by 40 cent means $A_4 = 450.28\,\text{Hz}$, a difference that should be recognizable. Apparently detuning to the flat side was more noticeable for AP subjects (the score for correct identification averaged over all conditions reached 48.8%). There are more reports documenting that AP subjects have difficulties in correctly identifying intervals in the case where the tuning is changed by 50 cent or a transposition of the tonal context is introduced, while subjects with RP seem to adapt to such conditions more easily [31.394, 395].

As the data from *Heyde* [31.385] show, identification of keys (with the standard tuning at $A_4 = 440\,\text{Hz}$) was better than that for tones (77.3% of all judgments in her sample of AP subjects ($n = 47$) as compared to 65.6%), and identification of keys that are frequently

used in music (such with no or few accidentals) was better than that for keys with rare occurrence. In regard to the error scores obtained in such experiments, missing the target by one semitone (up or down the chromatic scale) seems perhaps the most frequent error with AP subjects, followed by errors by a whole step and octave errors. In comparison, errors mistaking the target for its perfect fifth or fourth are much less frequent (in Heyde's data, about 2% on average for tones, and 4.2% for keys). Mistaking the target category by a perfect fourth or by a perfect fifth notwithstanding the perceptual difference in tone height indicates that a relational principle of tone qualities as are expressed in musical intervals, and in the circle of fifths as a perceptual and cognitive framework, plays a significant role also for AP possessors (as was elaborated by [31.390]). Octave errors may arise from perceptual factors such as misjudging the tone height of a stimulus due to spectral convergence and phenomenal similarity between two complex tones one octave apart (Sect. 31.1.2); this type of error does not affect correct identification of the pitch category.

It is not quite clear what the cues are for complex tones presented in isolation to let subjects accomplish the task of identifying the appropriate pitch category. As has been elaborated above (Sect. 31.1.2), the concept of tone qualities taken as saturated elements within each octave was viewed in parallel to vision of colors. *Bachem* [31.48] consequently spoke of *tone chroma* to denote this peculiar quality assigned to 12 *chromatic* tones for a range of several octaves. However, a tone quality usually is defined referentially, by the relations one tone has with others in regard to a scale or the intervals they form [31.52, 72]. Without such a context, it is still a matter of debate what constitutes a particular *tone quality* or *tone chroma*. Assertions that such tonal qualities would *exist*, at least for AP possessors, as much as chromatic colors may *exist* for subjects capable of unimpeded vision, are built upon an analogy that does not explain the specifics of encoding and retrieval relevant for AP abilities. It has been conceded for a long time even by proponents of a special kind of *tone chroma* perception experienced by AP possessors [31.412, pp. 32 ff.] that perceiving *absolute* pitches is neither completely independent of relational ties nor free from limitations imposed by register, timbre, and other factors. In sum, except for very few cases (of the *child prodigy phenotype* like that reported in [31.396]) there is hardly an *absolute AP* that would cover the musical gamut in total as well as hold for many different timbres. Rather, there are various types and grades of AP as empirical data (including responses to questionnaires, self-assessments, error scores from experiments, etc.) demonstrate. Further, there are many subjects who have good or excellent RP abilities and have also encoded a number of fixed pitches because they have a long experience in tuning their instruments (such as bowed or plucked strings). Usually, such subjects can both identify and sing or whistle the relevant pitches. Though they might not qualify as *AP possessors*, in a narrow sense, they have encoded a range of pitches as anchors in LTM that they can use also in judgments on pitches falling in between anchor points [31.392]. A general ability to memorize pitch levels may exist in connection with the memory for certain songs. Observations on a fairly large sample of European subjects [31.413] recently showed that a certain percentage could reproduce a song of their choice relative to the original within a limit of ± 2 semitones.

Given that AP abilities are often observed in subjects who learned to play the piano at an early age [31.66, 405] and taking the role of the piano in music education into account, it seems feasible that many of these AP *possessors* encode absolute frequencies and brightness values for a number of complex tones of a diatonic or even a chromatic scale they experience intensively during the *critical period*. Piano tones seem suited to this task since the brightness (measured by the spectral centroid of complex tones) changes markedly with pitch. For example, for a chromatic scale C_3–C_5 played with the sound of a sampled grand piano (MIDI velocity for all notes was set to 80), the spectral centroid rises from about 1.2 kHz to ≈ 3.2 kHz while f_1 of the sounds rises from 130.8 to 523.3 Hz. In effect, brightness can be used as an additional cue to pitch for each single scale tone even if brightness is not independent of pitch but, rather, one component of the pitch of a complex as well as of a pure tone.

At this point, it should be emphasized that fundamental components of pitch as a product of sensation and perception, namely brightness, tone height, and *tone quality*, should not be confused with *dimensions*. In regard to models and methods of multivariate statistics, dimensions require to be (a) continuous (representing variables that, in principle, can express arbitrary values within lower and upper boundaries) and (b) linearly independent (so as constituting an n-dimensional vector space, see [31.414] and textbooks on multivariate analysis). While the condition of quasicontinuous representation and measurability obtains (within JND/DL limits) for brightness and *tone height*, whose magnitude increases with frequency, *tone quality* resists quasicontinuous grading inasmuch as it is conceived in terms of discrete, nonoverlapping and fairly wide pitch categories (like the 12 chromatic semitones on a keyboard). In regard to scaling [31.415, 416], there would be thus interval scales for brightness and *height*, and

an ordinal scale of *qualities*, the latter with the special feature of *octave equivalence*. Among the many objections brought forward against the two-componential theory of pitch was that, for complex harmonic tones such as produced by strings, woodwinds and brass, the percept of musical pitch is unitary and coherent, and not an aggregate of components [31.54, 55]. Dissociation of components would, therefore, be possible in abstract analysis (as also [31.37, 56] admits) but highly irrelevant for music perception where listeners (and musicians alike) would conceive of the pitch of a certain tone as a single point in a two-dimensional pitch–time-space (ordinate: pitch, abscissa: time). The position of a tone in this space relevant for music performance and pitch perception would be marked by its fundamental frequency (as in conventional Western staff notation where the ordinate corresponds, roughly, to log frequency [31.417]). In fact, reading either a melographic or a sonagraphic (spectrogram) representation of musical phrases in a 2-D (two-dimensional) representation (ordinate: frequency (lin or log) or cents; abscissa: time) reveals the structure of pitches, scale steps and intervals quite clearly [31.93, 137, 306, 418–420].

31.8 Scales, Tone Systems, Aspects of Intonation

In an evolutionary perspective, pitch perception was established firmly with vertebrate animals [31.421]. Birds (aves) express themselves by producing elaborate sequences of sounds that rightly are described as *birdsong* [31.422]. Many mammals likewise produce sounds at different pitches, and with different timbre; such utterances often serve the purpose of communication in a group, or to attract a female. In certain respects, human vocalizations are similar to those of other mammals, and music as used among humans can be viewed as a means of communication, if not survival that had to be invented in the course of hominid evolution [31.423]. Scale types and melodic patterns have been a topic often addressed in publications on comparative musicology since they were regarded as indicative of stages of musical development [31.424–427]. Well into the 20th century (before the advent of modern media and communication technology), most of the non-Western music cultures as well as many folk music traditions in Europe offered a broad range of scales and modal structures as well as a variety of tunings for instruments. Early investigations of *primitive* music (a term not to be read as derogative but in the sense of medieval Latin *primitivus* = the first of its kind) recorded in the field showed that in some cultures music in fact was very elementary as melodic formulae oscillate between no more than two pitch levels about a whole tone apart (but not always stable and to be viewed as relative pitch levels). Such forms were recorded in New Guinea and other remote places (e.g., Andaman Islands). However, melodic patterns built on a bichord or trichord were observed also among Finno-Ugrian peoples [31.428]. A lullaby recorded in New Mexico from an old Zuñi grandmother, in 1950 [31.429, Expl. 2], basically is a recitation on one stable pitch (G_4) with, repeatedly, a drop to a fourth below (D_4) held as a second stable pitch. In other cultures, prototypes of scales with four or five pitches were in use (some still are). A host of pentatonic scales has been observed in Old and New World cultures.

While those elementary forms of melody oscillating between two or three relative pitch levels can be viewed as exploring tone height differences rather than using distinct scale steps, rudiments of a scale become apparent for melodic phrases where two whole steps are combined with a semitone in between or where a whole tone is combined with an interval of a minor third (the gamut in both cases is a fourth; for examples from Australian Aborigines see [31.430]). A fourth often serves as a frame interval and is filled with two pitches in between. One of the widespread scales found in traditional music styles was and still is anhemitonic pentatony such as comprising the notes C–D–E–G–A. It offers whole tone steps close enough to each other to be perceived as *neighbors* in a scale so as to allow continuation of a melodic phrase upward or downward and intervals wider than a whole step to realize more expressive melodic gestures (in particular if the scale is extended to the octave). Anhemitonic and hemitonic pentatonic scales usually consist of stable pitch structures; in certain regions, the use of neutral thirds (the size is between a major and minor third) or other inflections of pitches may occur. In some music cultures the series of harmonic partials as available on aerophones and chordophones serves as melodic material [31.431]. Overtone singing (as in Mongolian *Chöömij* [31.432]), excitation of harmonic eigenmodes on duct flutes such as the large Slovak *Fujara* [31.433] and the use of the musical bow as in southern Africa are but a few examples of music based on harmonic partials. Evidently, there are several principles of scale formation [31.93, pp. 313 ff.]; a number of music cultures apparently realize an *additive* concept in that a nucleus at the size of a whole tone occupies the center of melodic

movement from which a few pitches up and down may be reached [31.428]. Others seem to pursue a *divisive* approach in that frame intervals (octave, fifth, fourth) are divided into a number of steps. Further, employing a basically pentatonic scale in a multipart context can necessitate octave transpositions of scale tones as one finds in East and Central African music genres [31.434, Chap. III]. In addition, the use of pentatonic modes in parallel voices easily introduces the fourth as a simultaneous consonance, and men and women (or men and boys), when singing the same melodic contour, often proceed in parallel octaves or fifths (as observed in non-Western and in Western musical practice [31.424, pp. 42 ff.]). Discovery of the octave, the perfect fifth and apparently also the perfect fourth as fundamental consonant intervals could be achieved in musical practice, and perhaps without much conceptualization. Moreover the octave, because of its natural foundations (Sects. 31.1.2 and 31.4), governs pitch perception in a way that seems difficult if not impossible to ignore in musical practice. Also the perfect fifth and the perfect fourth, for their distinct interval character and consonance (this applies especially for the perfect fifth used as a simultaneity) could serve as fundamentals in many musical structures; in fact, given their use in various musical cultures and genres they can be regarded musical *near-universals*. Further, the *sweetness* of major thirds and other consonant intervals apparently was discovered in indigenous music cultures of Polynesia and parts of Africa before the advent of Europeans who couldn't believe their ears when hearing multipart music that sounded quite similar to their *major tonality*, and musicologists and ethnographers were likewise surprised to hear melodic patterns based on harmonics (usually, from the second up to the 11th or 12th) in indigenous and folk music genres. Also, the tritone (suspect as *diabolus in musica* in medieval music theory) appears as melodic interval and even in simultaneities, in folk music contexts (e.g., vocal and instrumental music of Slovakia).

The variety of scales and modes (as to this concept [31.435]) utilized in vocal and instrumental music of many ethnic groups and cultures of course is the result of developments that are complicated as historical processes (and outside of the scope of the present study). In regard to the genesis of the tone systems that became basic to European art music, from about medieval times into the 20th century, one should, however, take notice of ancient Greek music theory (as well as its survival and partial reformulation in Persian, Arab, and Indian sources).

As has been emphasized in works on the history of mathematics, speculations on ratios of small integers can hardly be separated from empirical investigations in which string lengths corresponding to basic musical intervals were measured [31.436, Part. 2]. It was fairly easy to derive basic musical intervals from sections of a string on a kanon (the antique predecessor to the medieval monochord) as is evident from the *sectio canonis* of Euclides (third century B.C.E.; see [31.437, pp. 133 ff.] for the Greek text and a German translation; also [31.296, Chap. 14]). Euclides explains the relations of tones and intervals by line sections, which can be viewed as representing string sections. When he argues that consonant (in the Greek text: symphonous) tones give rise to but one sound in which they *fuse*, it is clear that he must have experienced by ear the intervals he elaborates on. Similarly, when Aristoxenos (fourth century B.C.E.), in his *Harmonics* [31.438, Chap. 7, pp. 158 ff.] explains the whole step as the difference between a (perfect) fifth and fourth, and the differences between concords and discords within the range of an octave, he refers to the actual experience of music, in particular to melody. Though mathematical speculation was a genuine part of Greek music theory (and vice versa [31.296, 436]) it must have included empirical observations early on. For example, basic intervals of the *Pythagorean* tone system can be derived by both a mathematical and a geometrical procedure (the latter applicable to line or string sections). In a context where the notion of *harmonia* is viewed as the principle uniting things that are unlike among each other, Philolaos (fifth century B.C.E. [31.439, p. 689], [31.438, pp. 37 f.]) determines the magnitude of *harmonia* (the octave) by the fourth and the fifth. The difference between the fifth $(2/3)$ and the fourth $(3/4)$ is a whole step $(8/9)$. He then states the tones between which intervals of a fourth or a fifth exist and concludes that the octave comprises five (whole) tones and two dieses. From the text, the tonal organization of the octave can be derived. A whole step can be subtracted two times from the fourth; one whole step subtracted from the fourth leaves the minor third $(27/32)$ while two whole tones subtracted from the fourth leave a *remainder* (the *leimma* $243/256$). The process of subtracting two whole steps from the fourth can be carried out also arithmetically; in a modern syntax this would read like

$$(4/3)/((9\times 9)/(8\times 8)) = 256/243\,.$$

In addition, the leimma can be found by a progression in fifths. One has to take notice that, in Greek music theory, scales are usually conceived in a downward direction. Five pure fifth downward from a given tone, the leimma is reached

$$f \leftarrow c \leftarrow g \leftarrow d \leftarrow a \leftarrow e\,.$$

Corrected for the octaves traveled, the ratio of the interval is calculated as

$$((2/3)^5) \times ((2/1)^3) = 256/243\,.$$

Without going into details pertaining to the construction and history of tone systems in antiquity and their survival in the Middle Ages in Europe and the Near East [31.52, 71, 73, 296, 438, 440, 441], the point of interest here is that the Greek *Dorian* scale, which is regarded as reflecting Pythagorean approaches to scale construction in progressions of fifths most clearly, has a parallel in the common European melodic major scale (Fig. 31.18)

Evidently, there are two identical tetrachords (Greek *tetrachord* means *four strings*) in both scales joined by a whole tone. The scale in question is suited for melodic lines, and intonation patterns of violinists come close to it [31.89]. However, the major third 81/64 (a *ditonus* since adding two whole tones 9/8) is too wide to *fuse* in a simultaneous interval, and this deficiency was already known in the Middle Ages when the use of major thirds in compositions of art music became an issue (early examples can be found, for example, in the so-called Robertsbridge Codex, dating circa 1350 [31.305]). Moreover, there was music composed (or adapted from folk music patterns already existing at the time) where parallel major and minor thirds are a constitutive feature like in the famous *Nobilis, humilis* from the Orkney Islands, a hymn from the 12th century written in praise of St. Magnus, Earl of Orkney [31.442]; parallels to *Nobilis, humilis* are found in Norwegian folk music.

The Greek theorist Archytas of Tarent (fourth–third century B.C.E.), according to Ptolemy (*Harmonikon*, p. 30 ed. Düring [31.297, p. 47]) elaborated on three different tetrachords, the diatonic, the chromatic, and the enharmonic. The intervals in these three tetrachords spanning a perfect fourth are

$$(9/8) \times (8/7) \times (28/27) = 4/3, \text{ which equals}$$
$$1512 - 1701 - 1944 - 2016$$
$$(32/27) \times (243/224) \times (28/27) = 4/3$$
$$1512 - 1792 - 1944 - 2016$$
$$(5/4) \times (36/35) \times (28/27) = 4/3$$
$$1512 - 1890 - 1944 - 2016\,.$$

It is easy to see that the three tetrachords have one interval in common (the diesis $28/27 = 2016/1944$) and that the whole numbers Ptolemy introduced for demonstration are just expansions of the fractions. The diatonic tetrachord offers a septimal whole tone 8/7 besides the common whole step 9/8, the chromatic

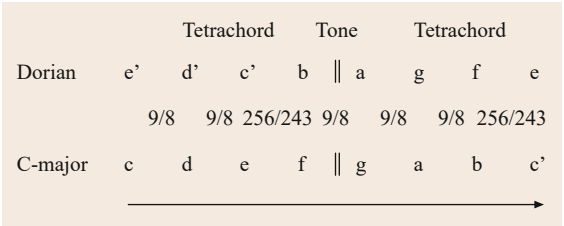


Fig. 31.18 Dorian scale and melodic major scale

tetrachord has a minor third 32/27, and the enharmonic a just major third 5/4. Though it seems that Archytas did try to describe tonal relations that were in use at his time [31.438, pp. 43 ff.], [31.443, pp. 413 f.], Ptolemy criticized some of the ratios as not fitting appropriate melodic intervals he was familiar with when played on a kanon. He, therefore, offered alternative solutions [31.297, p. 33 ff.], which include a tetrachord labeled

Diátonon sýntonon

$$(10/9) \times (9/8) \times (16/15) = 4/3,$$

in whole numbers

$$504 - 560 - 630 - 672\,.$$

This tetrachord comprises two different whole tones which, in combination, result in the just major third: $(10/9) \times (9/8) = 5/4$. The difference between the two whole tones is a *syntonic comma* (expressed in cents, $9/8 = 203.9$ and $10/9 = 182.4$ cent; the syntonic comma thus is 21.5 cent). The difference between a major third 5/4 and the perfect fourth 4/3 is the diatonic half tone of $16/15 = 111.7$ cent. Ptolemy claimed that all the intervals he calculated can be tested by attuning an eight-stringed kanon in a precise manner. The diatonon sýntonon had a prerunner ascribed to Didymos (first century), whose diatonon was of the form $(9/8) \times (10/9) \times (16/15)$; both became the blueprint for the modern harmonic scale, which can be written as noted in Fig. 31.19.

The ramifications of Greek and Hellenistic music theory were that several principles of scale construction were explored, some with a specific mathematical background such as defining intervals as superparticular ratios, $(m+1)/m$, a process that gives relevant musical

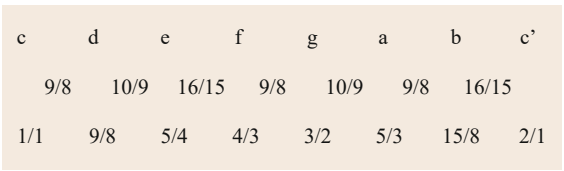


Fig. 31.19 Harmonic major scale

intervals in a straightforward way; for example,

Evaluate [Table [$(m+1)/m$, $\{m, 15\}$]]

turns out the list

$$\left\{2, \frac{3}{2}, \frac{4}{3}, \frac{5}{4}, \frac{6}{5}, \frac{7}{6}, \frac{8}{7}, \frac{9}{8}, \frac{10}{9}, \frac{11}{10}, \frac{12}{11}, \frac{13}{12}, \frac{14}{13}, \frac{15}{14}, \frac{16}{15}\right\}$$

Though such operations may seem removed from musical practice, this is not the case since intervals of the size 7/6 (the septimal minor third of 266.9 cent) had a musical function. In the Greek tone systems, the interval 7/6 is that between the tone called mese (the middle one) and the trite (a tone in the tetrachord above the mese [31.52, 71, 438]). Adding a diesis (paramese–trite) to a whole tone (mese–paramese) gives $(9/8) \times (28/27) = 7/6$. Such intervals were apparently part of musical practice at the time of Archytas [31.443, pp. 413 ff.] and are employed in genres of Near Eastern music to the present where also very small steps corresponding to the chromatic and, in particular, enharmonic genus as described by Archytas and other Greek theorists survived [31.370]. One can measure intervals such as 7/6 and even 36/35 in Turkish Ney music [31.418]. The whole tone 10/9 besides that of 9/8 can be found on fretted string instruments such as the Azerbaijan lute *tār* [31.93, p. 318, Tab. 18].

Perhaps the most basic operation relevant for the construction of tone systems and a variety of scales is a progression in fifths and fourths from whatever starting point is chosen. A sequence of five fifths (in downward direction) gives the leimma 256/243 (90.2 cent), a sequence of seven fifths upwards the so-called apotomē (2187/2048 = 113.7 cent, another semitone). Evidently, leimma and apotomē add up to the whole step

$$(256/243) \times (2187/2048) = 9/8\,.$$

Going eight fifths down from a given tone yields a complicated fraction, 8192/6561, which is, however, a close approximation to 5/4. The difference is less than 2 cent and thus below the threshold for pitch. For example, the tone F^b (eighth fifths down from C), when put into the same octave, is almost identical with the just major third above C (F^b = 384.4 cent, $\bar{E}$ = 386.3 cent). Thus, it is possible to derive an almost just major third by a progression in fifths, a process described by Şafī al-Dîn Urmavi, one of the leading scholars of Persian-Iraqi music theory, in the 13th century [31.73, Chap. 3]. He developed a scale comprising 17 tones per octave, which includes not only a number of almost just thirds but also the minor whole tone in addition to the major whole step and a septimal whole tone 7/8 (as expressed in string length; expressed in frequencies, it would be 8/7).

With the reception of Greek and Hellenic music theory in Europe, beginning in late Roman times and pursued up to the 15th century [31.440, 441], many of the intricacies of the Greek tone systems, their division into tetrachords and their internal organization according to the three genera (diatonic, chromatic, enharmonic) were still known as is evident from Boethius (around 500 AD) whose treatise *De institutione musica* became most influential for medieval theorists. Guido of Arezzo, in his *Micrologus* (circa 1026 AD) refers to dissonances arising from false intonation of singers as well as from sharpening or lowering the size of a certain interval too much [31.444, pp. 134–135]; several manuscript copies of the *Micrologus* include passages where Guido is criticizing singers to substitute a diesis for a semitone. It can be shown [31.440, p. 162 f.] that the diesis Guido had in mind is the diesis 28/27 of Archytas; added to the whole tone 9/8 it gives a septimal minor third: $(9/8) \times (28/27) = 7/6$. Echoes of antiquity were not just a matter of numerological speculation which, to be sure, must always be viewed in conjunction with geometrical diagrams and measurements on the monochord; mensuration of organ pipes played a role as well [31.441, 445]. Apparently, even musical practice at certain places included systematic microtonal inflections of pitches in basically diatonic melodic textures as can be traced in the Tonary of St. Bénigne de Dijon (Antiphonarium Codex Montpellier, late 10th century [31.446, p. 565 ff.]).

The fundamental problem experienced, in music theory and instrument building, in the course of the Middle Ages and ever since, is that of incommensurability: intervals and their multiples derived from different prime numbers do not converge. It was known to Greek theorists that a chain of 12 fifths was not equal to seven octaves: $(3/2)^{12} \neq (2/1)^7$. The 12 fifths overshoot the seven octaves like the ratio 531 441 : 524 288; the difference is small but significant and amounts to 23.46 cent (known as the *Pythagorean comma*). To make 12 fifths equal seven octaves (this is the range spanned by the progression in fifths), the interval of the fifth must shrink by a small amount to 2.996615/2 instead of 3/2. However, this is a modern calculation involving decimals while medieval music theory still was based on small integer ratios; besides the superparticular ratio $(m+1)/m$, there was the superpartiens ratio $(m+1+n)/m$ like the major sixth 5/3, and the multiplex ratio as in the octave 2/1; further, certain combinations of these proportions were allowed [31.440, p. 64 ff.]. In Pythagorean tuning, a perfect fifth of course comprises a perfect fourth and a whole tone. If the fifth is conceived to span from the diatonic note B

to the accidental F#, to be perfect, the fifth would need the following intervals (leimma, tonus, tonus, leimma, apotomē):

256/243 * 9/8 * 9/8 * 256/243 * 2187/2048 = 3/2

B c d e f f#

Tuning a chain of 12 fifths up and down from a center (g) would result in such a scheme

a^b e^b b^b f c g d a e b f# c# .

There are 11 perfect fifths and, to force the chain into a *circle*, one interval c# - a^b that falls short of a perfect fifth at 678.5 cent (sometimes labeled *wolf fifth* for its howling sound quality). Of the major thirds, eight are of the Pythagorean ratio 81/64 but four are almost just due to the schismatic equation known from Şafī al-Dîn Urmavi (see above). Pythagorean tuning can be assumed for the 14th century (as reflected in organ dispositions and surviving music of the time [31.447]). Tunings and temperaments became an issue when, around 1350–1400 AD, works appeared that made use of the major third as a simultaneity. A long controversy was started among theorists before, in the 16th century, both the major third (5/4) and the minor third (6/5) as well as their complementary intervals (the minor sixth 8/5 and the major sixth 5/3) were acknowledged as consonances, along with their extensions into the next octave (minor and major tenth, etc.). *Gioseffe Zarlino* [31.448, Lib. I, Cap. 36–40], [31.305, p. 72 f.] applied both the arithmetic and the harmonic mean to the division of intervals and demonstrated that the fifth contains the just major and the just minor third.

In regard to tuning, in collections of early organ music like the *Buxheimer Orgelbuch* (≈ 1460/70) there are pieces that still feature fifths and fourths as fundamental consonances; however, there are also pieces where in particular the long held final chord includes a major third (e.g., no. 216, where the final chord is $C_3 - G_4 - E_5$), indicating that the tuning of organs had changed, or was about to change at that time, from a basically Pythagorean pattern to one that could adequately account for major thirds. For example, *Ramos de Pareja* [31.449, Cap. 5] proposed a tuning that combines a number of perfect fifths and just major thirds. By about 1600, the major and minor third and their complementary intervals were widely in use. Works like the *Paduana lachrimae* (written by John Dowland and adapted to keyboard by Sweelinck and other composers of that era) are full of simultaneous thirds that bring the emotional message of the piece to the fore.

The just major third 5/4, which had been calculated in antiquity (see above), introduced another prime number into music and thus intervals incommensurable to their Pythagorean counterparts. It was not possible to tune a conventional keyboard so that it offered perfect fifths and just major thirds for all keys. Among the host of solutions that were tried, beginning in the 15th century, we find various temperaments (aiming at a compromise between tuning in fifths and tuning in major thirds) on the one hand, and expansion of the number of keys per octave on the other (for detailed accounts see [31.450–453]; for mathematical aspects see [31.53, 319, 454]). Quite often, both approaches were combined to result in keyboard instruments that had from 14–31 or even 53 keys per octave and offered either a regular temperament or a selection of pitches from just intonation. Around 1600–1700, the most common approach to tuning was that of so-called meantone temperament, which actually comprises a variety of tunings. Meantone temperament was invented to feature just major thirds, of which up to eight can be realized on a keyboard tuned to so-called quarter-comma meantone temperament (Fig. 31.20; the signs $0, -1, -2, +1$ denote comma differences of tones/pitches relative to tones in the 0-series of fifths).

The temperament that was advocated by *Michael Praetorius* [31.455], among others, is implemented by tuning four pairs of just major thirds ($b^b - d - f\#$, $f - a - c\#$, $c - e - g\#$, $e^b - g - b$), just thirds being written in the vertical and connected by the sign | in this scheme. With only 12 keys per octave available, the price to be paid for no less than eight just major thirds is a detuning of the remaining four major thirds (which are 428 cent wide). In addition, most of the fifths (written in horizontal direction and connected by ... in this scheme) are narrowed by 5.5 cent while its complementary interval, the fourth, is widened by the same margin in order to sum to the octave (696.5 + 503.5 = 1200 cent). Adding two such narrowed fifths and subtracting an octave, the

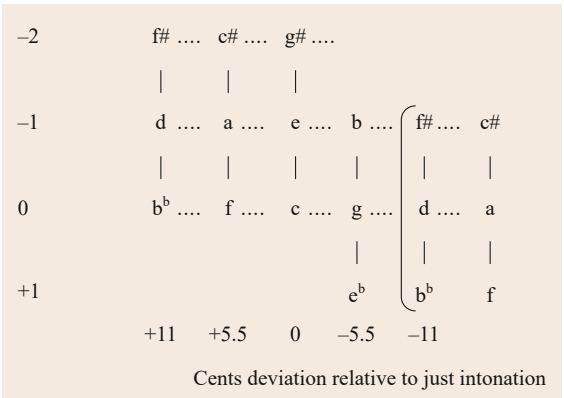


Fig. 31.20 Scheme of quarter-comma meantone temperament for 12 keys

c	c#	d	e^b	e	f	f#	g	g#	a	b^b	b	c'
0	75.5	193	310.5	386	503.5	579	696.5	772	889.5	1007	1082.5	1200

Fig. 31.21 Quarter-comma meantone scale

meantone of 193 cent is found (which is 11 cent flat relative to the major second $9/8 \approx 204$ cent). A problematic interval in 1/4-comma meantone temperament is the fifth $g^{\#} - e^b$ which, at 738.5 cent, is far too wide. The total error of this tuning system relative to just intonation pitches can be reduced significantly if the system would be expanded beyond 12 pitches and tones per octave (expansion would be possible to the right of the bracket in Fig. 31.20 as well as to the left side). With only 12 keys/pitches per octave, the scale tuned to 1/4-meantone temperament is as displayed in Fig. 31.21.

The intervals are given in cents, which of course is a modern measure based on logarithms (for historical approaches to meantone temperaments [31.450–452]). Evidently, the pitch of the tone d, calculated as the geometric mean, is in the middle of the major third c – e (numerically, the meantone can be calculated from $\sqrt{(5/4)} = 1.11803 = 193.2$ cent), which makes d the *meantone* to c and e; by the same relation, f# is the meantone between e and g#. The scale comprises two different semitones, a chromatic (75.5 cent) and a diatonic (117.5 cent), which are suited for the highly chromatic music of the 17th century. The scale offers c#, f# and g# as well as e^b and b^b but lacks d^b, g^b and a^b as well as d# and a#; thus, the major and minor chords that can be actually played on an organ in 1/4-comma meantone tuning, with little inharmonicity of the resulting sound patterns [31.456] are C, F, G, B^b, D, A, E, E^b and c, g, a, d, e, b, f#, c# (uppercases = major, lowercases = minor). Among the intervals available in 1/4-comma meantone tuning are two augmented sixths ($e^b - c\#$ and $b^b - g\#$) that are almost 966 cent wide, which is very close to the natural seventh $7/4 \approx 969$ cent. It seems that Frescobaldi and Scheidt recognized this option for enhanced harmonic expressivity in some of their keyboard works [31.452, pp. 341–343].

It should be noted that *Fogliano* [31.457, Lib. III, Cap. 2/3] had proposed a harmonic division of the monochord, which yields a scale analogous to the harmonic scale; the fifth d – a in this scale ($10/9 \times 16/15 \times 9/8 \times 10/9 = 40/27 \approx 680.5$ cent) is one syntonic comma flat of a perfect fifth. Also, d – f is a minor third ($32/27 \approx 294$ cent) one comma flat against the just minor third ($6/5 = 316$ cent). To provide for correct intervals (the fifth d – a and the perfect fourth f – b^b), Fogliano inserted two more tones (the d above c as 9/8 and the b^b as 16/9 [31.458, pp. 70 f., pp. 106 ff.]). The 14 tones/pitches included

$$C\,\tfrac{25}{24}\,B\,\tfrac{16}{15}\,D\,\tfrac{81}{80}\,D\,\tfrac{16}{15}\,B\,\tfrac{25}{24}\,E\,\tfrac{16}{15}\,F\,\tfrac{25}{24}\,B\,\tfrac{27}{25}\,G\,\tfrac{25}{24}\,B\,\tfrac{16}{15}\,A\,\tfrac{16}{15}\,B\,\tfrac{81}{80}\,B\,\tfrac{25}{24}\,H\,\tfrac{16}{15}\,C$$

Fig. 31.22 Fogliano's monochord division

in *Fogliano*'s scale [31.457] (plate fronting fol. XXXV, tone designations are his) can be seen in Fig. 31.22.

This scale can be written (including cents [31.459, p. 94 f.]) like in Fig. 31.23.

If ordered in a tone net of pure fifth and just major thirds [31.52, p. 242], these 14 tones/pitches yield a structure similar to that shown in Fig. 31.20 except that the fifths are now perfect (Fig. 31.24).

With these 14 tones/pitches, eight just major chords (C, F, G, D, A, E, B^b, E^b) and seven just minor chords (c, g, a, d, e, f#, c#) can be played. Thus, if this system would be implemented on a keyboard instrument, with two additional keys a great improvement of the tuning and the range of just intervals and chords could be achieved. In fact, in particular during the second half of the 16th century and the first half of the 17th century, there were many attempts at developing keyboards with more than 12 keys per octave [31.450, 452, 460]. One of the reasons was to accommodate the chromatic and enharmonic styles of music aiming at a renaissance of musical genera known from Greek antiquity; another was to solve the problems left with meantone tuning. A practical solution implemented on a number of organs and also on harpsichords was to expand the keyboard by one or two subsemitonia, namely split keys that were needed for d# and e^b as well as for g# and a^b to get rid of the so-called *wolf fifth* ($g\# - e^b$ has 738.5 cent in 1/4-comma meantone temperament). The difference between the g# available in this tuning (Fig. 31.20) and an a^b needed for certain chords (a^b major, f-minor) but not available with only 12 keys is the diesis $128/125 \approx 41$ cent, which is the difference between three just major thirds and the octave (e.g., $c - e^{-1} - g\#^{-2} = 1158.9$ cent; the exponent −1, −2 indicates the pitch is one syntonic comma or two flat compared to the note of the same label in a chain of perfect fifth; see below). Consequently, to make meantone temperament workable in most or even all keys, one needs more than 12 pitches by adding from one or two subsemitonia to split keys for the five accidentals (some instruments had additional keys also for $e\# \approx f^b$ and for $b\# \approx c^b$). Otherwise, one has to restrict the gamut to a central range of keys. The gamut of keys that can be played in 1/4-comma meantone temperament with

c'	$c^{\#-2}$	d^{-1}	d	e^{b+1}	e^{-1}	f	$f^{\#-2}$	g	$g^{\#-2}$	a^{-1}	b^{b}	b^{b+1}	b	c'
1/1	25/24	10/9	9/8	6/5	5/4	4/3	25/18	3/2	25/16	5/3	16/9	9/5	15/8	2/1
0	70	182	204	316	386	498	568	702	772	884	996	1018	1088	1200

Fig. 31.23 14-Note-scale (derived from Fogliano)

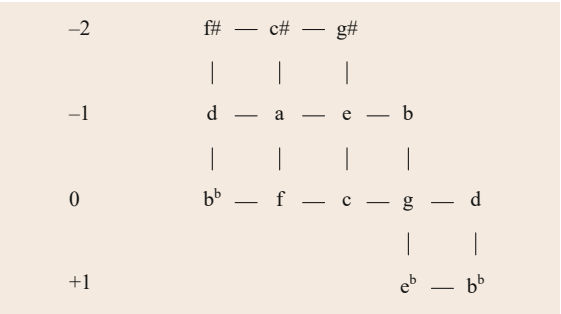


Fig. 31.24 Pitch structure derived from Fogliano's 14-tone scale

a standard keyboard (12 keys) usually is from E^{b} to A (that is, from three flats to three sharps). Most works of the era conform to this range. Some require notes that are not available, or only in some approximation that might, however, be acceptable when viewed from expressing affects such as distress and grieve by means of extreme chromaticism [31.453, p. 179 ff.], [31.451, p. 237 ff.].

Around 1680–1750, a range of *well-tempered* tunings for keyboards with just 12 keys per octave had been developed that form a compromise between tuning as many just thirds as possible (implying imperfect fifths and fourths as in quarter-comma meantone temperament) and tuning perfect fifths (implying *Pythagorean* intervals, see above). The problem is to determine pitches suited to representing intervals based on three prime numbers {2, 3, 5} in such an approximation to just ratios that all tones are usable both as melodic and harmonic intervals with the constraint that only 12 keys and pitches are available on a conventional keyboard. The number of tones/pitches needed for all intervals to be realized in major and minor keys is fairly large; the regular 53-tone temperament that was favored by *Bosanquet* [31.51] and Helmholtz offers a good approximation for intervals based on prime numbers {2, 3, 5}. If also intervals based on the prime number seven are considered, such as the natural seventh (4/7) available as the seventh harmonic on wind and brass instruments, and actually used in musical practice (from folk music of Norway, Slovakia and Switzerland to Benjamin Britten's *Serenade for tenor, horn, and strings*, op. 31), the number of pitches per octave goes up significantly [31.52]. In any event, the basic condition is $n \gg m$, where n is the number of pitches/tones required per octave for just intonation or good approximations to just pitches, and $m = 12$ is the number of keys actually available on a conventional keyboard to be tuned accordingly. Hence, *well-tempering* (from Latin temperari = to adjust, to balance) comes down to an optimization process where one must try to find *best fits* according to several criteria, among them usability of all tones in melodic and harmonic functions in as many keys as possible, along with a high degree of consonance and a low degree of roughness. Nowadays, these goals can be achieved on an algorithmic level and by means of signal processing [31.318]; by about 1680, theorists could not but calculate manually (though some did already with the help of logarithms that had been introduced, in the 17th century, into calculation of musical intervals and tunings [31.450, 451]). In practice, musicians in general had to find proper pitch assignment to the 12 keys from their professional experience in tuning organs, harpsichords, and clavichords; the temperaments implemented on keyboards had to stand an empirical test, namely that of expert listeners such as J.S. Bach taking part in examinations of newly built organs. To conclude from historical reports, sensitivity for judging by ear different tunings (and their relative deficiencies) must have been very high. The task of *well-tempering* gained urgency when compositions ventured into sharps and flats outside the gamut well covered by quarter-comma meantone (e.g., D. Buxtehude, *Praeludium in E*, BuxW 141). Proposals for well-tempering such as offered by *Werckmeister* [31.461, 462] should not be confused with *equal temperament* (ET12); a *well-tempered* tuning typically offers intervals (fifths, thirds, semitones) of different size and so also maintains differences between keys with respect to degrees of consonance versus beats or roughness. A *well-tempered clavier*, therefore, has to be distinguished from a keyboard tuned to equal temperament (ET12). There have been many hypotheses as to Bach's *recipe* for tuning a harpsichord or organ [31.463–466]. Though differing in details, these proposals all converge on a nonequal temperament for which indications can be found in the keys as well as in harmonic and melodic textures in Bach's works for keyboards (in particular, organ works [31.467, 468]). Recently, a new evaluation of Book I of Bach's Well-Tempered Clavier has been carried out [31.469] by employing dissonance calculations based on the approach of *Sethares* [31.318]. Also, organ temperaments relevant for an adequate

performance of Bach's works have been evaluated empirically [31.470].

While quarter-comma meantone temperament and concepts of *well-tempering* were in use, also equal temperament was pursued both empirically and by means of calculation. Among the empirical approaches there was Vincenzo Galilei's proposal of fretting a lute to the ratio 18/17, which leads to semitones of 99 cent [31.459, p. 57]. On a theoretical basis, there were attempts to find an equal temperament by geometrical divisions of the monochord whereby the *Pythagorean comma* (23.5 cent, the excess of a sequence of 12 perfect fifths against seven octaves, see above) was split in small proportions to each of the 12 fifths [31.451, 459]. A strictly arithmetic equation for an equal temperament (ET12) derived from a subtle compression of the fifth away from the perfect ratio can be written thus (in a modern syntax)

$$\text{Solve}[(x/2)^{12} - (2/1)^7 == 0, x]//N\,.$$

The set of solutions includes $x \to 2.996615$, the value mentioned above, and also $x \to 1.49831$, which results as a decimal from the fraction 2.996615/2. Expressed in logarithms, $1200\log_2(2.996615/2)$ is exactly 700 cent, the size of each fifth in ET12. Calculation of equal temperament was performed, in Europe, by Simon Stevin (probably 1585, or around 1605), who returned to the issue known from Euclid's critical account of Aristoxenus that the octave comprises six equal tones. After some initial attempts [31.17, Chap. 2.3], [31.451, p. 179 f.], Stevin calculated the tempered minor third from

$$\sqrt[4]{\tfrac{1}{2}}$$

and the major third from

$$\sqrt[3]{\tfrac{1}{2}}\,.$$

He came very close to exact values for all 12 semitones. His treatise though was not published before 1884. Shortly before Stevin, in China the prince Chu Tsai-yu is said to have calculated ET12 by about 1584, however, the approach he took apparently is based on proportions and divisions of string lengths while a more abstract calculation of

$$\sqrt[12]{\tfrac{1}{2}}$$

for the semitone as it seems is not stated explicitly as a mathematic formula in his works [31.471, 472].

There were several concepts of tempering a chain of fifths so that a circle could be formed [31.459, p. 156 ff.]. *Johann Neidhardt* [31.473] offered a number of temperaments, still using a monochord for divisions of strings; a criterion featured in his concept for tuning fifths and major thirds in different *circles* are beat frequencies. His final proposal [31.473, p. 50 ff.] for an equal tempered scale was still derived from proportions, to result in fifths that are all 1/12 of a Pythagorean comma narrowed, major thirds that are all 7/12 comma wide, and minor thirds all narrowed by 8/12 comma. This is in fact ET12. However, at the time ET12 had been calculated correctly, it was hardly put to practice, for several reasons. One had to do with the laborious process of retuning in particular organs, another with the results of retuning that were found musically unsatisfying. As *Mattheson* [31.474, p. 144 f.] criticized, semitones tuned to ET12 *all sound out of tune, namely, if one does not hold them as such but against the singers and instruments, in particular the trumpet* [...]. Mattheson correctly recognized that the beat frequencies in ET12 change wildly with register and interval (the term *gleichschwebende Temperatur* that was used in many German publications as a synonym for ET12 is incorrect and misleading).

The quest for ET12 around 1700–1750 has to be viewed in regard to developments in music and music theory. Concepts of modes and clausulae derived from the eight Gregorian modes (and their expansion to 12 by *Glarean*, in his *Dodecachordon* [31.475]), had been complemented with new ideas concerning major and minor. A decisive factor was the acceptance of the just major and minor third as consonances and a discussion of how to use them in triads and harmonic progressions (as elaborated by [31.448]). Though traditional concepts of *mode* lasted well into the 18th century, even with modifications (as is evident from many works of J.S. Bach), the distinction between *hard* and *soft* keys (according to major and minor thirds) inherent in a diatonic scale d – e – f – g – a – b – c was found relative because every note of the scale could be changed by alterations up and down as needed [31.476, pp. 14–15]. *Heinichen* [31.477, 478] presented a circle (*Musikalischer Circul*) as a guide to understand relationships between common major and minor keys as well as a guide for modulation between such keys in a situation where an organist needs to improvise in a preludium. He [31.477, p. 262, Sect. 6] explained that, in the genus chromaticum covering the sharps, the musically relevant extreme would be reached with H-Dur (B-major), and in the genus enharmonicum covering the flats, at B^b minor (which Heinichen saw as a *tone* (tonus, key) hardly in use anyway). Further expansion around the circle, Heinichen warned, would be of no avail; this remark does not support the idea that modulation through all 24 keys was intended or that the *Circul* of 1711 reflects ET12. The compass regarded as musically relevant by *Heinichen* [31.477] is from five sharps

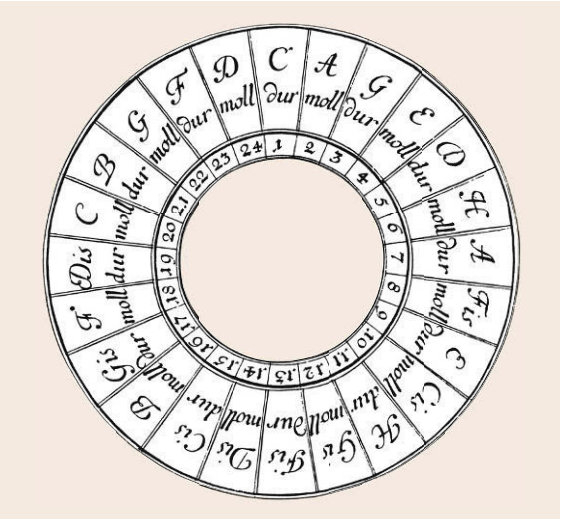


Fig. 31.25 *Musicalischer Circul* (after [31.478])

to four flats at the utmost. The circle in 1728 was closed by equating d# with e^b, and dismissing, besides e^b, keys that would have called for even more flats (Fig. 31.25 from [31.478, p. 837]).

Even though the 1728 *Circul* did not necessarily presume ET12, at least some *well-tempered* tuning close to ET12 seems implied in the circular arrangement of keys. In musical mathematics it was known that a chain of pure fifths could not be transformed into a circle and would result in a spiral instead if continued over many octaves [31.319, p. 156 f.]. However, there are concepts comparable to a *circle of fifths* outside of equal temperament like the cycle of Ṭhaṭ in north Indian music [31.366]; also Chinese historical sources from the Zhou period onward [31.479, pp. 65–67] elaborate on chains of pure fifths and fourths as a backbone of musical scale construction.

In regard to the normative interval of the perfect fifth, *Drobisch* [31.480, p. 32] described ET12 as *but one special case of the system of fifths* besides other variants (of which he reported and calculated several [31.480, 481]). However, ET12 became the standard tuning for *Western* music as played on keyboards since the fifths and fourths in this tuning system are nearly perfect and the major and minor thirds as well as the corresponding sixths are tolerable; the major third in ET12 at least is less prone to creating roughness than the Pythagorean 81/64 interval. With the spread of ET12 the *rivalry* between several prime numbers as a basis for musical intervals as encountered by Fogliano and Zarlino had been solved. But also ET12 has its cost: the *chromatic scale* in ET12 consists of 12 equal steps of 100 cent each and neither offers a chromatic (25/24 = 70.67 cent) nor a diatonic (16/15 = 111.73 cent) semitone; the very term *chromatic* thus is misleading inasmuch the distinct qualities of scales and modes based on different semitones (as are used in truly chromatic works, e.g., Sweelinck's *Fantasia chromatica*) have been eliminated. Also eliminated with ET12 on keyboards are differences in the relative consonance or dissonance of major and minor chords played in different keys that are clearly perceived in quarter-comma meantone temperament, prompting composers to express certain affective and emotional states by using different keys. Such differences were still available, to some extent, with *well-tempered* tunings, some of which can be regarded as outperforming ET12 since a gradation of keys is maintained while none of the intervals and chords sounds harsh (e.g., sixth-comma meantone, Vallotti/Young; [31.482, Chap. 18–5]). A problem introduced with ET12 on organs is the discrepancy between two tunings; while the pitches of the keyboard are now in ET12, pipes in the mixture stops are still in just intonation. A combination of both gives rise to audible beats and roughness in particular if major thirds are included in the pipe ranks (as in a historical *Terzzimbel*, which is very present in the sound of an organ).

The process of rationalization and simplification that led to ET12 has consequences for music theory, too. The development of modern major and minor tonalities in the 16th and 17th centuries as well as the chromaticism that became a formative factor demanded more than 12 (pitches/tones) per octave both in conceptualization and in musical performance. If one opted for just thirds in triads, and seeing that the just major and minor third add up to the perfect fifth (as Zarlino did), the implication is a web of pure thirds and fifths as embedded in a tone net (Fig 31.26).

–2-series	h	fis	cis	gis	dis	ais	eis	his		
–1-series	c	g	d	a	e	h	fis	cis	gis	dis
–0-series	as	es	b	f	c	g	d	a	e	h
+1-series	fes	ces	ges	des	as	es	b	f	c	g
+2-series				bes	fes	ces	ges	des	as	es

Fig. 31.26 Tone net from pure fifths (3/2) in the horizontal and major thirds (5/4) in the vertical

For readability, the tones/pitches are designated here with their labels as are used in Germany and neighboring countries. The series are ordered vertically with respect to syntonic commas; for example, in just intonation the tone *as* (a^b) in the +2-series is two syntonic comma higher in pitch than the corresponding tone in the 0-series while the tone *h* (b) in the -2-series is two commas lower than the tone *h* in the 0-series. The tone net shown in Fig. 31.26 of course is a section of a structure than can be extended in the vertical and in the horizontal as far as is needed (if one incorporates also intervals based on the prime number 7, the structure becomes three-dimensional; [31.52] and below).

As is obvious from the tonal relations involved, a simple cadence C – a – F – d – G – C in just intonation would require the following tones and pitches

−1 – series	d	a	e	h	
0 – series		f	c	g	d

There are two different tones for *d* relative to the center, $c = 1/1$, which have frequency ratios $9/8$ and $10/9$ that would be used in the d-minor and in the G-major chord respectively. If we consider another simple cadence in minor $c - f - B^b - E^b - f - G - C$, it is clear that the f-minor chord needs an a^b from the +1-series. The quest for enharmonic instruments in the 16th and 17th centuries [31.450] must be viewed, besides the aspect of exploring the ancient genera, as a serious attempt to give melodic and harmonic tonal relations as conceived and elaborated in compositions an adequate musical and acoustical realization. As one can hear and apprehend from cadences played on instruments like the *Clavicymbalum universale* (19 tones/pitches per octave [31.455, Cap. XL]; a replica was built by Keith Hill in 1983, now in the Organeum at Weener/Frisia), differences between chords with flats and chords with sharps are distinct and identifiable (of course, with appropriate training). In sum, the long struggle with tunings and temperaments as was pursued from about 1400–1800 would have been abandoned much earlier if differences in the tuning of scales and chords had been musically and perceptually irrelevant.

With the implementation of ET12 as a standard, the history of scales, as *von Hornbostel* [31.483, p. 13 f.] once put it, *inevitably ends*. Though chromaticism is found embedded in 19th century works written for the modern piano (for contrasting examples from Renaissance, Baroque, and 19th century music see [31.484]), it lacks the tension generated from two (or more) different sizes of semitones; the concept *enharmonic* totally changes its meaning under the constraints of ET12 in that, instead of marking a distinctive difference between sharps and flats (e.g., g# vs. a^b), both can be now substituted for each other by means of so-called *enharmonic equivalence*. Hugo Riemann, eminent music theorist, insisted that the difference between keys far apart relative to the circle of fifth (like g^b and f#) can be apprehended and even perceived (in *inner perception* that does not require actual sensory input [31.485]) from *reading* scores. *Riemann* [31.486, p. 117 f.] maintained that *orthography* is a decisive factor for music cognition, and for cognitive processing of harmony in particular while the effect of different keys taken as enharmonically identical would vanish for the listener of music. Riemann thus advocated an approach where musical textures have to be conceived (and heard in *inner perception*) in the tone system that represents pitch relations in just intonation while execution and sound production would be done in some temperament that would more or less level those differences (one of his major studies was on Beethoven's piano sonatas). Riemann, who at the beginning of his career followed concepts that aimed at giving music theory acoustical and psychoacoustic foundations as were provided by *Helmholtz* [31.28] and pursued also by *Arthur von Oettingen*, another physicist and music theorist known as a proponent of *harmonic dualism* [31.487, 488], later retreated to a dichotomy of music as conceptualized in *Tonvorstellungen* [31.489], on the one hand, and music as played and heard, on the other [31.300, 305]. In contrast, both Helmholtz and Oettingen regarded just intonation as the acoustical, perceptual and conceptual fundament of harmony, and both also owned special keyboard instruments to explore harmonic sequences in just intonation. *Stumpf* [31.490], acknowledging that the major and the minor tonality call for a *dualistic* perspective in harmony, but seeing that part of the formulations given by Oettingen and Riemann were at odds with common musical experience, tried to reconcile the sensory and the cognitive approach by stating that consonance and dissonance is a matter of sensation while conceptualization of harmonic structures is achieved in terms of *concordance* and *discordance*, respectively.

There are good reasons to take just intonation (JI) as an acoustical and psychoacoustic basis as well as a valid reference system for music theory: evidently, most of the complex sounds used in music are composed of harmonic partials so that chords combining harmonic complexes again superimpose perfectly if their fundamentals are in small integer ratios (Sect. 31.6.4). A chord played in JI has a clear and unambiguous sound structure, which in turn allows clear determination of tonal functions and harmonic relations among chords [31.52, 488]. Moreover, virtual pitches are elicited under this condition in an unambiguous way, and the difference between a major and a minor chord based on the same tone and fundamental (say, C-major, c-minor, both in *root* position, f_1 at 244 Hz) can be demonstrated em-

pirically as one of virtual pitch and not one of spectral centroid or brightness [31.305]. Therefore, if one would want to investigate concepts such as Rameau's *basse fondamentale* in a precise manner, it should be done with chord sequences realized in just intonation. JI is not advocated for obscure *numerological* speculation or a tedious effort for *purity* of intonation but for the sake of giving music theory, and in particular harmony, a valid acoustic and psychoacoustic basis as well as a meaningful reference system. Different from the time of Helmholtz, Oettingen, and Riemann when JI was awkward (and costly) to realize on keyboards, calculation of relevant pitches (see Appendix in [31.52] for a closed system of 171 pitches based on prime number {2, 3, 5, 7} ratios) and implementation as a software–hardware platform nowadays is feasible [31.318].

If no particular precision is required or wanted, for both musical performance and conceptualization, ET12 will do. It is in fact kind of a 'mean tone system' in many regards. First, the tritone in this system is calculated as the geometric mean that exactly halves the octave. Second, the semitone in ET12 is about the average between a chromatic and a diatonic semitone, and the two thirds in this system are about the average taken from the just and the Pythagorean ratios (however, with a closer approximation of Pythagorean values). So, if one assumes that intonation of singers and instrumentalists, in many if not most areas of *Western* musical practice, lies somewhere in between the two systems (Pythagorean, JI), the probability that data from measurements will approach more or less ET12 pitch values is high in particular if such data are averaged per *pitch class* (as in many publications on intonation); averaging data this way unfortunately disregards melodic and harmonic functions of tones/pitches within musical structure. What is achieved thus, is interpreting averaged intonation data referenced to an averaged ET12 system of pitches. Further, to keep things plain in terms of *categorical perception*, one may suggest that a pitch in music should be viewed as representing a diatonic scale step and its low and high alterations as well as a broad range of intonation patterns admissible for that pitch [31.55]. Such a view implies perception of pitches always can be reduced to that of pitch *categories* in a diatonic scale. To simplify things even further, one might add some assertions like the difference between a natural seventh (969 cent) and that available on a piano (1000 cent) would be within the tolerances of the ear anyway so as to justify neglecting it. If empirical data indicate otherwise [31.374, 377], one can still ignore them to keep things simple for elementary music education or similar purposes outside musical science.

Intonation in musical performance should realize certain scale types, modes and harmonic structures. Intonation can be guided (and is also limited) by sensory, cognitive, sensomotoric, acoustic, technical and other factors. For a singer or instrumentalist, a note in a conventional notation indicates f_1 of a certain tone that he or she aims to produce. However, notation often is only relative as, for reasons of convenient reading, simplifications in regard to accidentals might be made (in particular, to avoid double accidentals). In effect, of the ≈ 35 pitches that can be expressed per octave [31.50] without introducing additional symbols for intonation, often a selection is used that serves readability rather than precise intonation. If *enharmonic equivalence* is applied to accidentals, notation cannot escape ambiguity with respect to pitch information. Hence, intonation of advanced tonal music (such as string quartets, saxophone quartets, or orchestral works) requires harmonic analysis that must be done before any performance takes place. As part of such an analysis, intonation signs have to be added to the conventional notation as needed [31.52]. Musicians familiar with certain genres and their harmonic and interval framework may not need such auxiliary information, however, one has to remember that musical performance often is done in public and with the constraints of real-time execution; therefore, to reduce stress as well as notational ambiguity, auxiliary intonation signs can help. If notation is ambiguous in regard to pitch information and intonation, musicians in general try to do as good as is possible, playing or singing a pitch that they are certain fits the relevant *pitch category*, with an option to do some fine-tuning in an ensemble in particular on long-held notes. One can observe subtle pitch adjustments in chords in small ensembles performing madrigals, even of the truly chromatic and enharmonic genus (for an instructive audio demonstration, listen to the CD accompanying [31.491] as well as to recordings of madrigals from Monteverdi and Gesualdo performed by The Consort of Musicke, London). The same mechanism of pitch adjustment toward JI is observed in barbershop singing [31.374, 377] and also in string and saxophone quartets unless extensive vibrato is used by one or several players. Vibrato, which was sparingly used in Baroque music (and despised still by Leopold Mozart as a violinist), has become a means not only of operatic expression (as in belcanto singing) but also a means to disguise intonation deficiencies or insecurity. Vibrato extent measured in lyric singing can exceed ± 100 cent, on single notes [31.492]. In Fig. 31.27, vibrato of a soprano singing the word *Engeln* in Richard Wagner's Lied *Die Engel* (*Wesendonck-Lieder*) is shown. Pitch oscillates widely around B_4; detailed analysis with a gammatone filter bank revealed that the pitch shifts between ≈ 505 Hz and ≈ 585 Hz (a range of ≈ 250 cent).

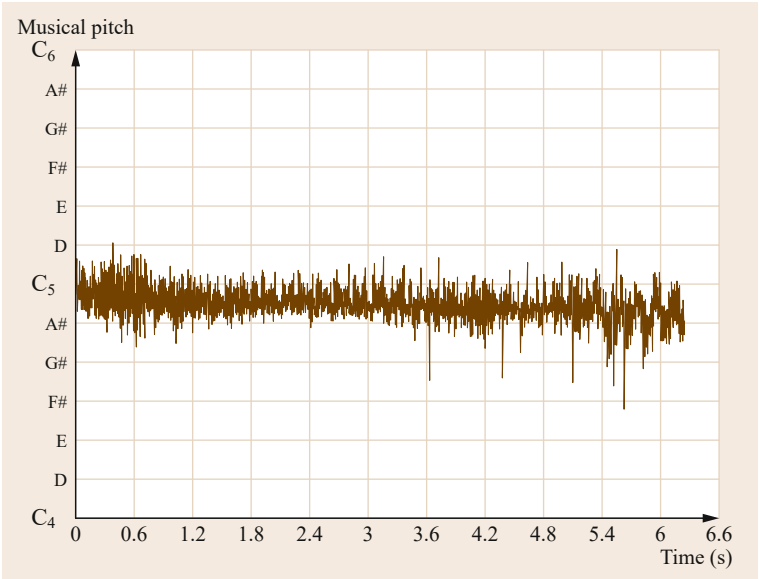


Fig. 31.27 Vibrato, soprano singing one word (*Engeln*), pitch shifts around B_4

Though listeners in general can derive a *mean pitch* from tones played with regular vibrato [31.493], vibrato to the extent shown in Fig. 31.27 blurs tonal structures. On the other hand, microtonal inflections as are produced by string bending on an electric guitar in blues and rock music often span a well-defined pitch range, and can be perceived accordingly. A musician producing precise string bendings as did the late Stevie Ray Vaughan in songs like *Texas Flood* [31.494] of course controls with his ears the sensomotoric process of fingering and applying force (increasing with pitch) to a string up to a target pitch level. Sensomotoric action and auditory pitch control are thus parts of a systemic feedback loop.

One last aspect to be included into this section is the assertion according to which a number of non-Western music cultures use scales in some equal temperament (ET). A popular view is that *gamelan sléndro* expresses ET5 and that the scale known from Thai gong chimes and xylophones (as in *piphat* ensembles) is in ET7. Hence, slendro is understood as 5×240 cent and the Thai scale as 7×171.4 cent. Also a number of African xylophones have been viewed as expressing either ET5 or ET7 (for background information see [31.93, 248, 495]. It seems that Alexander J. Ellis introduced the thesis of non-Western ET scales, unfortunately from measurements that can be called superficial at best. His argument, in short, was that there is not one *natural* scale (i. e., a scale based on small integer string or frequency ratios) but many different scales. To underpin his claim of diversity, Ellis strangely enough stressed that not only Europeans were using ET but also cultural groups as far away as South East Asia. Ellis' thesis has been reiterated time and again even though empirical data on idiophone tunings was scarce before empirical measurement could be done with suitable equipment and precision (i. e., from about 1970 to the present). Moreover, most studies on non-Western scales and tunings started from measurements of single spectral components (for example, by low-pass filtering complex inharmonic sounds) whereby frequency readings of the lowest spectral peak in each sound were taken as representing its *pitch* as well as a scale step in a certain tuning. Such an approach might hold for harmonic complex tones where one might regard f_1 as decisive for (place and periodicity) pitch, with the other harmonics reinforcing the overall periodicity (Sect. 31.4). In complex inharmonic sounds, however, there can be a number of strong spectral components in frequency ratios quite different from $f_n = nf_1$, $n = 1, 2, 3, \cdots$ (see analyses in [31.93]), which may be perceived as discrete spectral pitches. In addition, a number of strong inharmonic components in a sound can give rise to a virtual pitch at a frequency other than that corresponding to the lowest mode of vibration (as found in the spectrum). To illustrate the problem, one example shall be given. Taking a *saron ricik* in slendro tuning (a metallophone of six bronze plates corresponding to six *keys*; the instrument in question is part of the Gamelan *Kyahi Timbul*, manufactured in Yogyakarta), the f_1 components from six keys/sounds are displayed in Fig. 31.28.

Though one interval is considerably larger than the theoretical 240 cent, and another one smaller (as is typical for most slendro tunings), one could still take these differences as deviations from a *mean* scale step intended as realizing ET5. However, if the sounds from

Key	1	2	3	4	5	6
Hz	548.59	628.99	720.11	843.39	977.39	1114.04
Cents		236.8	234.2	273.6	255.3	226.6

Fig. 31.28 Javanese *saron ricik*, slendro tuning, spectral f_1 frequencies and cents

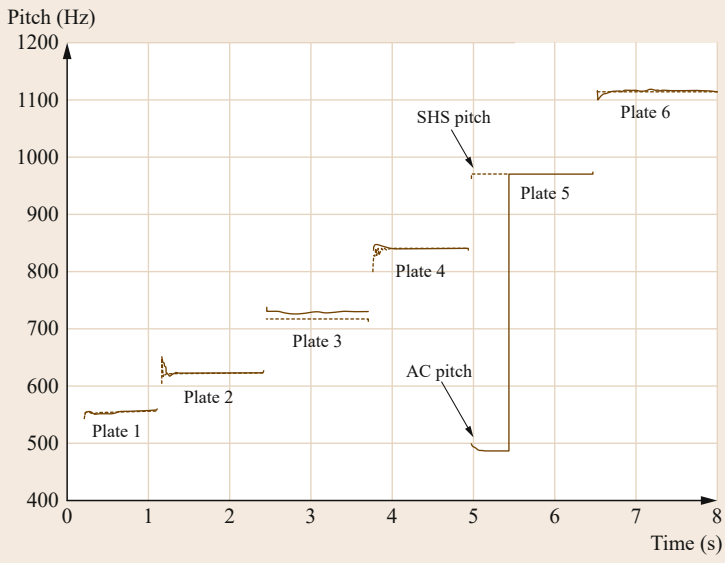


Fig. 31.29 Javanese *saron ricik*, pitch frequencies (AC and SHS). AC pitch for plate/key no. 5 at first is at 486 Hz (which is a subharmonic at 1/2 of the most likely pitch frequency). *Solid line*: AC pitches, *dashed line*: SHS pitches

Key	1	2	3	4	5	6
Pitch	554.1	621.69	717.13	841.39	970.87	1114.75
Cents		199.27	247.25	276.63	247.81	239.24

Fig. 31.30 Javanese *saron ricik*, slendro tuning, measured pitch frequencies (AC, SHS) and cents

the six plates and keys are subjected to a measurement of virtual pitches (by either AC or SHS algorithm, see Fig. 31.29), the deviations from a unit step assumed to be close to 240 cent do not disappear; instead, a scale of unequal steps becomes more likely as the five intervals between six keys are now as can be seen in Fig. 31.30.

Evidently, the octave is slightly enlarged at 1210 cent. Though the series of pitches measured does not lend to viewing *slendro* as ET5, it seems possible that Javanese musicians regard all intervals as *functionally equivalent* (the Javanese concept of *pathet* apparently combines features of *key*, *range*, and perhaps also *mode* [31.496]).

31.9 Geometric Pitch Models, Tonality

Concepts of scales and modes often involve geometrical structures (as does the very concept of a *scale*). Ordering of tones was done in a spatial arrangement at least since the time of Archytas. There are many technical terms in Greek music theory that are derived from geometry (or the other way round, geometrical concepts derived from measurement of string sections on the kanon [31.436, Part. 2], [31.443]) such as *diastema* (interval on a straight line or string) and *horoi* (endpoints of a line or string, also the two points that define an interval). That which is divided by two horoi is a *difference* (*hyperoché*), most of all, a difference in pitch. There are more terms and concepts with a geometrical background that have implications also for perception: Aristoxenus elaborates on the *topoi* of tones, which is the range within which so-called movable tones (*kinoumenoi*) can change their position while certain tones (*hestotes*, the tones framing tetrachords as well as whole steps at the bottom of the tone system and between disjunct tetrachords) are fixed in pitch [31.438].

The term topoi here refers to the positions a movable bridge or fret can take on a kanon; a change of topos means a change in the interval structure within a tetrachord, and a *difference* in pitch or pitches. According to Aristoxenus [31.438, p. 135 f.] there is a limit for the smallest difference that can be adjusted precisely on the kanon and also identified by ear (this of course is not a JND but a musically relevant interval, a small diesis that roughly is about 40 cent wide). The tones of the Greek tone systems (11, 15 or 18 tones either spanning an octave plus a fourth or two octaves [31.71], [31.438, p. 11 ff.]) can be conceived as ordered along a dimension low–high. One of the interesting issues discussed in Greek musical mathematics and philosophy is the problem of discrete entities and continuous magnitudes (as well as their distinctive differences and possible relations). The musical concepts of interval, boundaries and *difference* are part of this discourse.

One has to take the historical background into account when studying medieval sources on scales or even modern publications on pitch. In *Herbart*'s model of a one-dimensional *Tonlinie* [31.497, 498], tones are considered as points in a continuum; tones can be compared with respect to their linear distance. The idea behind this concept is that stimuli such as tones differ in certain parameter values, for instance frequency; differences are quantifiable and can be expressed as distance functions. In regard to sensation, it can be assumed that a growing difference in parameter values results in decreasing similarity or increasing dissimilarity of pairs of stimuli (Sects. 31.1, 31.7); hence, the effect on sensation can likewise be expressed as a distance function [31.93, p. 404 ff.]. *Stumpf* [31.39, p. 144] agreed to Herbart's concept, stating that the manifold of pure tones has but one dimension. The decisive argument for Stumpf was that, of any three tones A, B, C, only one can be in the medial position. Consequently, the three tones must fall on the same line. *Wundt* [31.499, p. 59 f.] added that it is *simple tone sensations* (sensations of sine tones of a given frequency) that make up the continuous manifold of but one dimension. Also *Mach* [31.500, p. 217] agreed on the one-dimensional construct, saying that the series of tones has to be viewed as an analogy to a space of one dimension, which can be conceived as a vertical line (since tones are ordered along this dimension from *low* to *high* in pitch) or as a line that runs in a median plane from front to back. Mach held that a series of tones (which are sensed as pitches) does not have symmetry; this implies there is no center or reference point. However, according to Mach sensation of a certain tone must be fixed to one distinct place in this one-dimensional tonal space.

The one-dimensional model of tones reflects aspects relevant for manifolds and continua as mathematical and physical structures; it might be recalled that Leibniz and Brentano had discussed foundations of continua while Riemann and Helmholtz contributed to differential geometry and higher-dimensional spaces [31.93, p. 404 ff.]. *Stumpf* [31.303, p. 165 ff.] underpinned the idea that one-dimensional continua of pitches (*Tonhöhen*), intensities, brightnesses, etc. are conceptual abstractions, while any tone we hear of course does have a pitch, an intensity, and a brightness as inherent properties. However, the construct of the *Tonlinie* could have been judged as unsuited for psychophysics since it disregards a basic perceptual experience, namely octave equivalence. A more appropriate model was proposed by *Opelt* who, in a preliminary booklet in [31.501] and in a more detailed publication of [31.313] offered a theory of pitch perception based on sequences of impulses and periodicity detection. He tabulated the intervals within one octave (see the Table in Fig. 31.31, no. 23) as logarithms corresponding to millioctaves, that is, $1\,\mu o = \sqrt[1000]{2} = 2^{1/1000} = 1.0006934$. The just major third 5/4 has (in rounded figures) 322 μo, the perfect fifth 3/2 has 585 μo; in ET12, the major third has 333.33 μo, the fifth has 583.33 μo, and the tritone of course has 500 μo. Opelt did not advocate ET12 except as a compromise needed for keyboard tuning. Instead, he [31.313, p. 40 ff., 66] correctly saw that ET31 would be needed as a tuning where all major chords 4 : 5 : 6 : 7 : 8 could be played with good approximation to just harmonic intonation. Since 31 tones and pitches per octave is a temperament that affords ingenuity and costs if implemented on a keyboard [31.382, 383], Opelt dismissed the natural seventh (his chord scheme thus became 4 : 5 : 6 : K : 8) and suggested 19-tone equal temperament (ET19) as a pitch and tone system relevant for conceptualizing music while, for practical reasons, ET12 would have to do on keyboards. To be sure, division of the octave in ET19 was propagated later also by *Yasser* [31.502]. The pitches and tones available in ET19 include a nearly just minor third 6/5 (315.8 cent) and an almost just major third (379 cent) while the fourth is about 7 cent wide (505.3 cent) and the fifth is about 8 cent narrowed (694.7 cent).

Opelt's approach included fundamentals of vibration and pitch, consonance and harmony, as well as aspects of rhythm perception, which are all treated from the fundamental principle of isochronous pulse sequences. It is of interest to note that he apparently was the first to offer a geometrical model of pitch in three dimensions. He first derived numerical calculations for two scales (ET19 for apprehension of music and ET12 for the limitations imposed by keyboards) along with a geometrical model that comprises the relevant sections on various (linear and logarithmic) lines (see Fig. 31.31, no. 19, included in [31.313] as plate

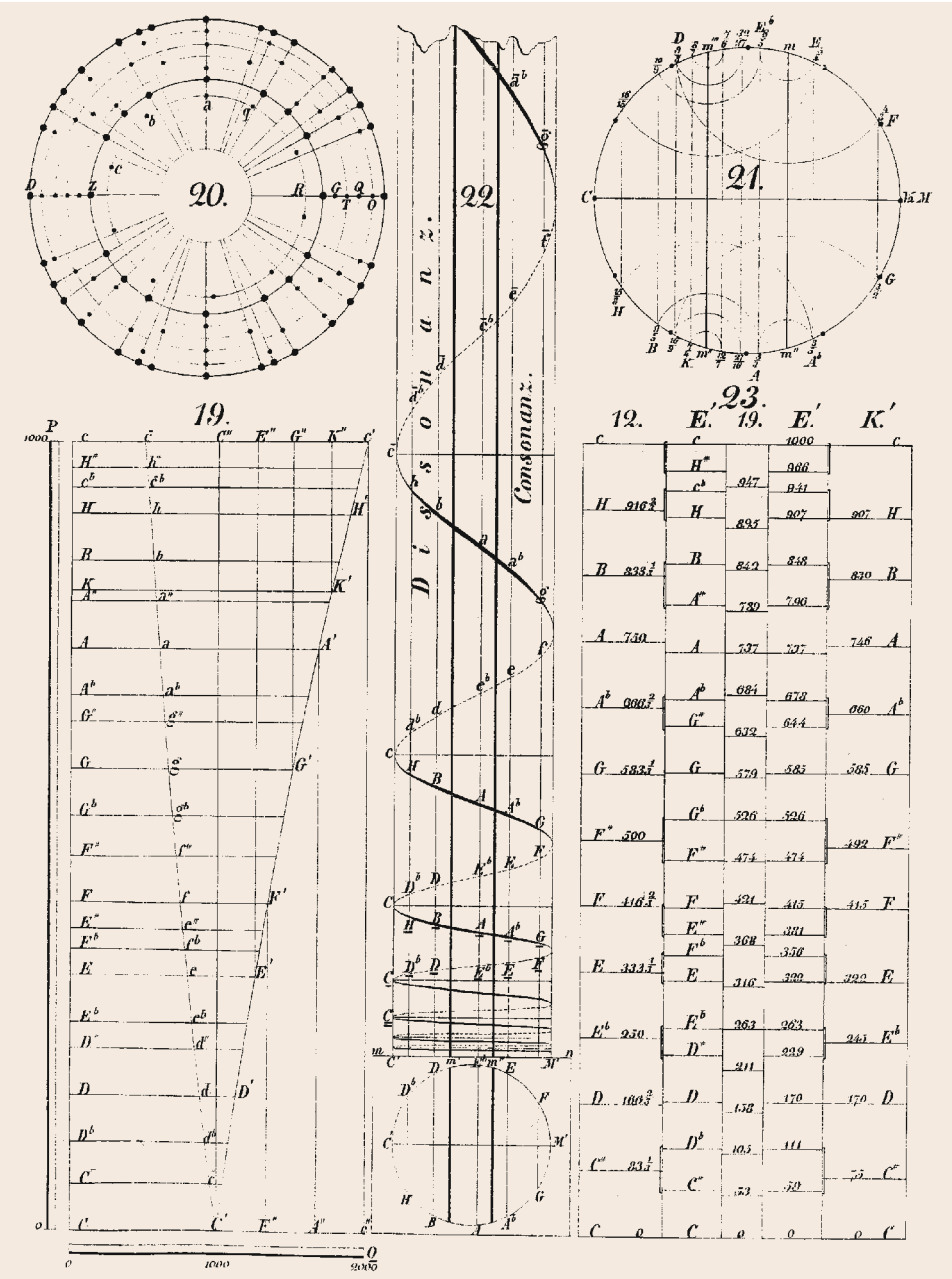


Fig. 31.31 Tone curves, tone circles and tone column designed by *F.W. Opelt* [31.313, 501]

VI). As a next step, his scales are transformed into circles (Fig. 31.31, nos. 20, 21) which also permit the finding of the pitches/tones for chords in each key. Different from later models, the cross-section of the circle (C–M in no. 21) in fact is *not* meant to yield the tritone in ET12 that Opelt considered musically irrelevant. The point M opposite the starting point C is only needed for symmetry within each octave. Finally, the logarithmic line (C′c′ in no. 19, labeled *Toncurve* = tonal curve by Opelt) representing the pitches is wound around a cylinder whose base is the circle of pitches. Complementary intervals adding up to the octave (e.g., fourth and fifth, major third and minor sixth) are found on the same ordinate. With the ascent of the logarithmic pitch scale wound around the cylinder (labeled *Tonsäule* = *tone column*; Fig. 31.31, no. 22), the scale becomes full circle after exactly one octave. The model Opelt designed nicely accounts for the rise of absolute frequency values

per octave; therefore, the helix angle of the tone curve increases in proportion to the frequency level reached in higher octaves.

Opelt's model was adapted by *Drobisch* [31.481, 503], who, however, changed the concept inasmuch as Opelt's conceptualization of just pitch relations (already simplified to ET19) is now further simplified to ET12. As a consequence, the circle of ET12 pitches is halved by f# opposite the starting tone/pitch of c. *Drobisch* [31.481, p. 35 ff.] refined Opelt's model by implementing radii from the center of the tone circle to the tones located on the tone curve wrapped around the cylinder [31.481, Fig. 3]. Thereby, he obtained a helix (like a staircase in a round tower [31.300, p. 24]).

Opelt's model of a spiral of tones/pitches wrapped around a cylinder and Drobisch's adaptation became fairly influential; *Lotze* [31.57, p. 212 f.] referred to the model and explained the musical scale phenomenon as a combination resulting from linear elevation (on the ordinate) and a declination sideways (with respect to the tone curve). From the perspective of his two-componential theory of pitch, *Révész* [31.37, p. 20] argued that a representation of the continuous element of pitch (height) together with the periodic element (tone quality) can only be achieved in a three-dimensional model like the spiral. He objected, though, on the grounds that the spiral would be ostensive yet not quite correct since its curvature does not account for the constant direction of tone height (since it rises linearly, a straight line would be more apt). *Révész* [31.56, p. 75] included a graph of the spiral wrapped around a cylinder into his account of two-componential pitch; it was, however, close to Opelt's original version and not the helix designed by Drobisch. To avoid misinterpretations, one has to observe that, in the spiral or helix model, two *components* of pitch are mapped on a geometrical construct of three *dimensions*. As has been underpinned above (Sects. 31.1, 31.7), these components cannot be regarded as independently variable (at least not in a musical situation where common instruments and vocalists generate tones).

A slightly modified construct of the *tone column* involving the spiral was presented by *Ruckmick* [31.504] who suggested a *tonal bell* as a geometrical model to incorporate another spatial attribute of sound: volume (Chap. 32). *Wellek* [31.505, 506] gave a detailed analysis of pitch and tonal attributes (height, brightness, volume, weight, density) that are variable as a function of frequency in regard to their dimensional (or, rather, nondimensional) structure. He [31.505] argued that, in musical perception and cognition, time would also be conceived spatially, and indeed is a dimension of tonal space since music and even a simple scale unfolds in time. The only other true dimension is that of low–high. In order to incorporate several components from the range of tonal attributes, a possible geometrical representation according to *Wellek* [31.505] would look like a slim cone or pyramid, however askew with the base close to the front and the peak more back in space (to account for *volume*, similar to the *tonal bell* of Ruckmick).

More recently, Opelt's spiral and Drobisch's helix have been expanded to a double helix and even more complex structures by *Shepard* [31.507, 508]. One aspect not covered explicitly in the tonal column of pitch is the relation of tones in a chain of fifth as well as its variant, the circle of (tempered) fifths (Sect. 31.8). Though it is obvious that the diatonic and the chromatic scale can be derived from a chain of perfect fifths ordered into one octave, the ascending tone curve (Fig. 31.31) does not reveal the progression in fifth as a separate *dimension*. Taking complex tones composed of sinusoidals spaced in octaves and garnered under a symmetrical spectral envelope [31.78], one can, to some extent, isolate *tone quality* by keeping the spectral centroid relevant for the sensation of both *tone height* and *brightness* of sounds constant (Sects. 30.2 and 31.1.2). With tone height and brightness kept on the same level, a scale of so-called *Shepard tones* thus is a movement in a plane circle. From the circle of tempered fifths then a double helix can be formed [31.507, Figs. 2, 3], which, together with the *chroma cycle* of ET12, results in a double helix wrapped around a torus in four dimensions. If the linear dimension of *height* is added, a five-dimensional structure incorporating the double helix wrapped around a helical cylinder results [31.507, 508].

The model is consistent in regard to certain requirements, namely, invariance under transposition, octave equivalence, and uniformity of scale steps. These requirements in general are fulfilled if one uses ET12 as is done by *Shepard* who [31.508, p. 316, Footnote. 4] also relies on categorical perception to anchor his model on the *chroma cycle* so as *to disregard the distinction between a sharp of one note and the flat of the note just above (e.g., $C^{\#}$ versus $D^{\flat}$)*. The five-dimensional model, impressive as a geometric structure, is linked to findings from probe-tone experiments [31.509], which also employed *Shepard tones*. The probe-tone method demands that subjects rate how well a probe fits to an element presented before. Elements can consist of diatonic or chromatic scales, single chords, or combinations of chords such as cadences. Perhaps the most common element is a diatonic scale in either major or minor followed by the probe tone. Also chords and cadences in major and minor have been employed in a host of experiments. The judgment done by subjects on scales of 1 (poor) to 7 (good) apparently is

one involving a goodness-of-fit estimate. For each trial one obtains a profile; for instance, 12 chromatic *Shepard tones* are presented as probes following a diatonic scale based on any of the these 12 chromatic pitches (also in ET12). Data from such experiments have been subjected to multidimensional scaling (MDS), and the findings have been interpreted as reflecting an internalized hierarchy of tonal relations [31.327, 509].

Most of the probe tone profiles obtained from such ratings exhibit a characteristic shape that is reminiscent of another profile, namely that of roughness and dissonance versus consonance as developed by *Helmholtz* [31.28] and repeated in a host of studies since (Fig. 31.12). For example, the following probe tone profile was obtained from a group of 45 students in an introductory course in music psychology (Hamburg, December 2001) for a C-major scale presented with synthesized piano-like tones (Fig. 31.32).

Since the ratings are on an ordinary scale, median values (*dashed line* in Fig. 31.32) were calculated along with the means. The highest ratings are found for C (median = 6.0, mean 6.0) and G (mean = 5.29, median = 6). The other diatonic scale tones also obtained higher mean and median ratings (E = 4.24 mean, median = 4) than the accidentals, of which C# scored lowest (mean = 2.44, median = 2). The tone B scored high at 5.58 (mean) and 6 (median), which is unsurprising since this probe tone is a repetition of the diatonic scale (C, D, . . ., B) offered before as *context*. The ratings for D, E, F, G, and A can also be explained from both the general familiarity musically trained subjects have with a C major scale and from the effect of STM in which the whole scale will be stored. Different from the short buffer available in language as a *phonological loop* [31.510], STM for tone sequences and thus scales, is much

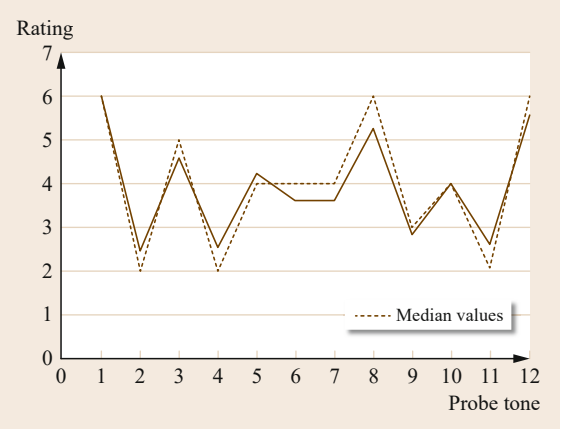


Fig. 31.32 Probe-tone profile, C-major scale, 12 probe tones C, C#, D, . . ., B based on means obtained from 45 subjects; *dashed line* = median values

greater [31.218] and can easily accommodate the *context* which, in this experiment, took about 2.7 s presentation time. The probe followed after a break of 250 ms and lasted for less than 500 ms, thereby the diatonic scale offered as context, the break and the probe add up to ≈ 3.5 s (this is a time window assigned to STM even from a conservative perspective [31.373, Chap. 4]). The probe can be compared to the complete diatonic scale to assess whether it fits from a musical point of view; probes representing diatonic scale steps thereby should receive significantly higher ratings than accidentals, what in fact they did (Fig. 31.32). In addition, there is a psychoacoustic effect if the probe is in marked dissonance with the scale and in particular with the last note of the *context*. The mean values for 12 scale tones in Fig. 31.32 correlate positively at $r_{xy} = 0.5$ with the means of consonance ratings for 12 intervals (from unison to major seventh) shown above in Fig. 31.12; since the data were obtained from two independent samples (2001, $n = 45$; 2009, $n = 51$), the correlation indicates that sensory consonance–dissonance plays a role also in probe tone ratings. *Leman* [31.511] demonstrated in simulation experiments involving a model of the auditory periphery as well as an echoic memory module that probe-tone profiles can be generated from the acoustical input (even from Shepard tones) without reference to hierarchic tonal schemes stored in LTM. Evidently, the *context* stored in STM as a sequence of pitches is sufficient for a perceptually driven comparison of the probe to the context in order to find whether or not the probe *fits*. This bottom-up processing approach is corroborated by findings from event-related potential (ERP) studies, which show that musicians are faster in encoding a regular sequence (*Neuhaus* [31.512] condition P^oD^o) than nonmusicians. Apparently, musicians need about three tones to make preliminary inferences as to the structure that is processed. Such inferences seem to involve expectancies about the direction in which a sequence is likely to continue as well as the shape it might adopt (there are indications that a frontal N300 component occurring in a time range of ≈ 300–600 ms after onset relates to expectancies with respect to sequential processing; see discussion in [31.513]). While processing of auditory input stored in STM seems feasible for rather short and simple tone sequences in regard to feature extraction and comparative tasks (e.g., an estimate of the goodness-of-fit of a probe tone), more complex structures call for the involvement of LTM. Consider, for instance, identification of modal structures in music where listeners have to analyze a melodic-diastematic formation in order to recognize which tones carry certain modal functions (like repercussa, finalis, confinalis [31.435, 514]). Guido (*Micrologus*, cap. xi) remarked that we can only

identify a mode beyond doubt after the very last note has been heard. One of the reasons is that a scale at the base of a mode can be different in upward and downward direction (as is often the case in Arab and Turkish Maqām/makam and in Indian Rāg [31.365, 366]). To provide even expert listeners with some *aide-mémoire* to recall complex modal structures interwoven with characteristic melodic patterns, Arab and Turkish classical music offer the Taqsim/taksim (as a fairly extended introduction to a maqām/makam) and North and South Indian classical music likewise offer an Alap in a similar function to introduce the Rāg succeeding it. Recognition of extended modal structures embedded in elaborate melodic-diastematic formations of course requires that listeners can reactivate their knowledge stored in LTM.

Finding the *tone center* or even the *key* of a piece of music has been achieved algorithmically, in a number of studies many of which used artificial neural networks (ANN [31.328, 329, 515–517] and [31.327, Chap. 4]). Experimental data suggest that the tone center, like pitch in general, is an emergent property that appears if a processing unit (an ANN or possibly a human) is exposed long enough to appropriate sound stimuli. Apparently, there are processes of self-organization in ANN (once they have been trained) capable of determining relationships between chords and keys organized in the ET12 circle and to store the results as schemata [31.518, 519]. Moreover, it has been shown in many experiments that ANN can accomplish a broad range of classificatory tasks from pitches and keys to meter, rhythm, timbres, and even genres and styles (see, for example, articles collected in [31.520]).

The causal approach to *tone semantics* [31.328–330] can draw on features inherent in tonal sequences and chordal patterns (in particular if realized with complex harmonic tones in JI). The normal cadence I–IV–I–V–I in Western harmony can be conceived as combining three triads related by perfect fifths and fourths; the consonance of fifths as well as identical tones in pairs of chords accounts for the perceptual affinity of chords within such a cadential schema. (In a *probe chord* experiment conducted with 45 students in 2001 in our institute, the G-major chord played with complex tones following a C-major chord as a reference did yield a mean of 4.47 on a scale of 1−7; SD = 1.29.) Human listeners familiar with *Western* music of the 18th and 19th century will make inferences as to a *tonal center* (or a sequence of several centers in a symphony or sonata) of a work they listen to from the occurrence of certain chord progressions, cadential schemata, as well as from the metrical accents and formal structure in which parts are arranged. For example, in a short and fairly simple piece like Chopin's Prélude in A-major the tonic occurs on the first quarter of the first bar, to be followed by the dominant seventh (bars no. 1, 2) and recursion back to the tonic in bars 3 and 4, etc. Thus, the center is clearly marked by the $T - D^7 - T$ progression that is a standard feature of indicating *tonality* and/or *key*. The only mildly *surprising* part of the work in question comes in bar 12 where a phrase ends on a $F\#^7$-chord. The note f# of this chord relates to the tonic by a minor third; $F\#^7$, as a seventh chord, can be conceived as the dominant of B (which, in turn, is the *secondary* dominant (in Riemann's terminology, *Wechseldominante, DD*) to A). The harmonic tension thus generated, however, is *released* in a plain progression of major chords $F\#^7 \rightarrow B \rightarrow E^7 \rightarrow A$ based on relations of fifths and fourths, respectively. Listeners familiar with Western harmony will have no difficulty in finding the tonal center and to apprehend harmonic functions of chords in musical textures like this. One could argue that the Prélude mentioned makes use of tonal schemata that demonstrate a *natural* as well as a *logical* organization; both aspects played a role in music theory from Zarlino to Riemann, and beyond [31.521, Chap. 16], [31.52, 300, 305]. *Riemann* [31.486, 522] developed an elaborate system of tonal relations and harmonic functions that reflects their *logical* character (in terms of *propositions* involving thesis, antithesis, and synthesis) as much as their *natural* foundations (the graded affinity of tones due to different degrees of consonance). This system can be condensed into a geometric structure (Fig. 31.33) where c is the center [31.523].

Riemann's system of harmony was devised to analyze and describe works composed between, roughly, 1700 and 1850. He devoted special studies to the piano sonatas of Beethoven, among other subjects. By the time of Beethoven, tonal harmony had evolved into advanced modulations while the fundamental cadential schemata familiar to listeners were still in use. A study of Western tonality of course cannot dispense with the historical background that is suited to reveal the very gradual transition from modal polyphony to functional harmony in music and music theory, between circa 1400 and 1750 [31.458, 524, 525]. For this process one can find a sample of *typical* and also a number of rather *untypical* cases both in theoretical writings and in musical compositions. Take, for example, one of Josquin des Prés' so-called *gospel motets* (*In principio erat verbum*) marking the transition from the 15th to the 16th century. In this work (recorded by, among others, René Clemencic and Musica Antiqua and Prague Madrigalists), one may observe many features known from modal techniques [31.526] while the cadences exhibit also strains already pertaining to major–minor tonality. In contrast, there are works written in the 20th century that fully ex-

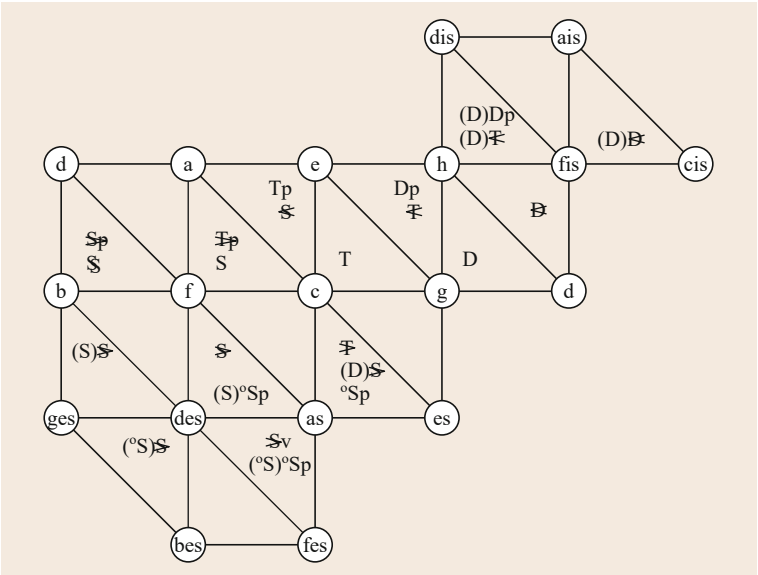


Fig. 31.33 Riemann's system of tonal relations and tonal functions (after [31.523, p. 81])

plore harmonic tonality and also incorporate elements known from modality (e.g., Vaughan Williams' *Fantasia on a Theme by Thomas Tallis*). To account for the harmonic structure of works such as Vaughan Williams' *Fantasia* or complex chordal textures as offered by Wagner (as in the *Tristan-Vorspiel*) or Strauss (in *Salomé* and other works), resorting to 12 *pitch classes* and a circle of 12 *chromatic* tones conceived in ET12 (by analogy to the 12 keys of a piano) is inappropriate. For example, in the *Fantasia* the so-called *five-chord motive* (based on a progression $G-F-g-A^{b}-G^{b}$) alone calls for 12 different tones/pitches. Even a composition clearly written for keyboard like the first *duetto* from part III of the *Clavier-Übung* by Bach (e-minor, BWV 802) according to the notation supervised by Bach has 17 pitches; this is a fact to be considered as evidence in an analysis of the melodic and vertical interval structure and, consequently, in regard to intonation. If one concedes this duetto can be played on a keyboard with but 12 keys and pitches, it is likely that by such a reduction some ambiguous intonation and a loss of the distinctness of intervals used in both melodic progression and simultaneity will be introduced.

The point that seems decisive for music theory and psychology is whether tonal relations embedded in interval and chord structures of musical works open to analysis shall be realized acoustically as pitch structures so that a correspondence between the two is achieved (within close limits). This is an issue that has led to debates and practical solutions as early as the Renaissance (for instance, Vicentino's 1555 treatise and his attempts at practical realizations of enharmonic music [31.450, 491]) and ever since. In the 19th century, *Helmholtz* [31.28], *Bosanquet* [31.51], *von Oettingen* [31.487, 488] and *Shohé Tanaka* [31.527] were among the leading scholars who advocated JI along with enharmonic keyboard instruments capable of rendering audible tonal structures far beyond diatonicism. Of course, such instruments were difficult to master and useful rather for demonstrating how specific chord progressions and modulations found in certain works of music were conceived, and should be apprehended, in analytical listening. One of the means of analysis that was recommended to be used in combination with enharmonic keyboards was the so-called tone net of perfect fifths and just major thirds proposed first by the mathematician *Leonhard Euler* [31.528] as a means of describing tonal relations in a clear and comprehensive way. The physicist and music theorist *Arthur von Oettingen* [31.487, 488] took up this concept [31.529]. He argued that, with respect to a tonal center defined by a certain tone and pitch (say, c), the tones and intervals shown in Fig. 31.34a are perceived as related.

The tones closely related to the center are the major third above and below this tone and the fifth above and below the center tone (the relations establish two symmetry axes). Also related to the center (but to a lesser degree) are the major third above and below the two fifths above and below the center. In sum, there are eight tones connected in this way to each center. To be sure, similar schemes adopted by Eitz and some other researchers reduced the number of tones related to the center to six (two fifths, two major thirds, two minor thirds).

Riemann [31.489, p. 19 ff.] adopted the reduced version of the tone net; the reason was that he saw

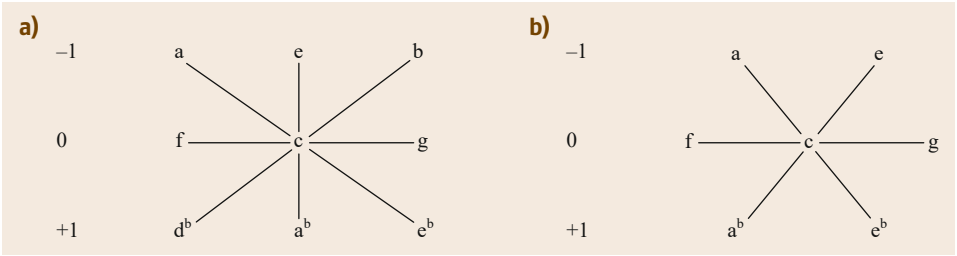
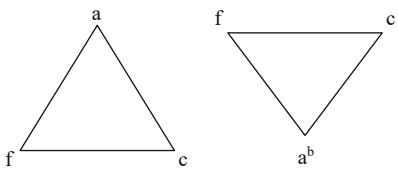
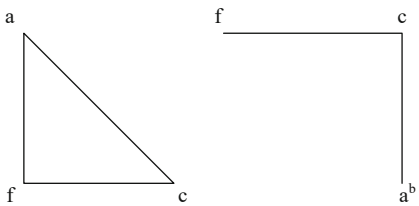


Fig. 31.34 **(a)** Tone center, related tones and symmetry axes (after [31.488, p. 386]), **(b)** Reduced tone net (adapted from [31.488, p. 384])

a constant shape of three tones forming either a major or a minor chord.



However, a constant shape also results from the more refined version of the tone net offered by *Euler* [31.528] and von Oettingen (one may either use triangles or open lines).



More advanced versions of the tone net were developed by *Fokker* [31.382, 530] who, following ideas already issued by Christiaan Huygens on incorporating intervals based on the prime number seven, pointed to the possibility of a three-dimensional structure. *Vogel* [31.52, 529] further elaborated this concept and developed a three-dimensional lattice structure (Fig. 31.35) that incorporates sequences of perfect fifths in the horizontal, sequences of just major thirds in the vertical, and the natural seventh (4/7 and 7/4) above and below each tone in the third dimension.

The main reason to represent tonal relations as geometrical structures in a tone net is clarity and comprehensibility. For example, the *distance* between certain chords and keys can be found by common geometrical operations (shift, translation), and such distances can be used for assessing perceptual and cognitive issues (the tone net allows quantification of chord progressions and of the *affinity* several chords or sonorities may have for musically trained listeners). In a historical perspective, the tone net is a late consequence of developments in music and music theory that started in about 1500 when (a) just major and minor thirds as well their complementary intervals were frequently used in musical compositions and when (b) the harmonic division of the octave, of the perfect fifth, of the just major third and even of the whole step yielded a range of musical intervals (3/2, 4/3, 5/4, 6/5, 9/8, 10/9, 16/15, 25/24) that could not be accounted for with a chain of perfect fifths (as in the Pythagorean tradition [31.451, 458]). Both aspects can be studied in works of music and in theoretical writings such as published by *Fogliano* [31.457] and *Zarlino* [31.448] where tonal relations include those of the prime numbers 2, 3, and 5. Further steps in the ongoing exploration of tonal relations include various cadential schemata (already in the 16th and more so in the 17th century), the dominant seventh chord as a means of *closure* based on tension and relaxation (in the 18th century and massively in the 19th century), harmonic chord progressions based on relationships of major and minor thirds, and finally harmonic textures involving septimal relationships between sonorities (as in Wagner's *Tristan* [31.52]). These developments included various types of chromaticism and even enharmonic tonal structures. In sum, musical composition diverged more and more from the diatonic scales and modes that were fundamental in church and secular music before 1500. By about 1850–1910 even functional harmony was expanded to the very limits (as in works of Wagner, Reger, Strauss and some of their contemporaries). In view of these developments, music theory could not adequately deal with complex harmony on the basis of diatonicism and its curricular correlate known as *Stufenlehre*. *Riemann*'s theory of harmonic functions [31.486, 522] was a serious attempt to deal with the shortcomings and failures of diatonic *Stufenlehre* and to supply more appropriate tools for music analysis and harmony. *Riemann* [31.50, 486] regarded JI as useful for conceptualizing tonal relations (intervals, chords) but as a pitch system unnecessary for musical practice. He [31.486, 489] instead opted for an approach where the music theorist should carefully *read* the score in order to apprehend the tonal relations unfolded in works of music as *tone images* (German: *Tonvorstellungen*). Correct *tone images* then

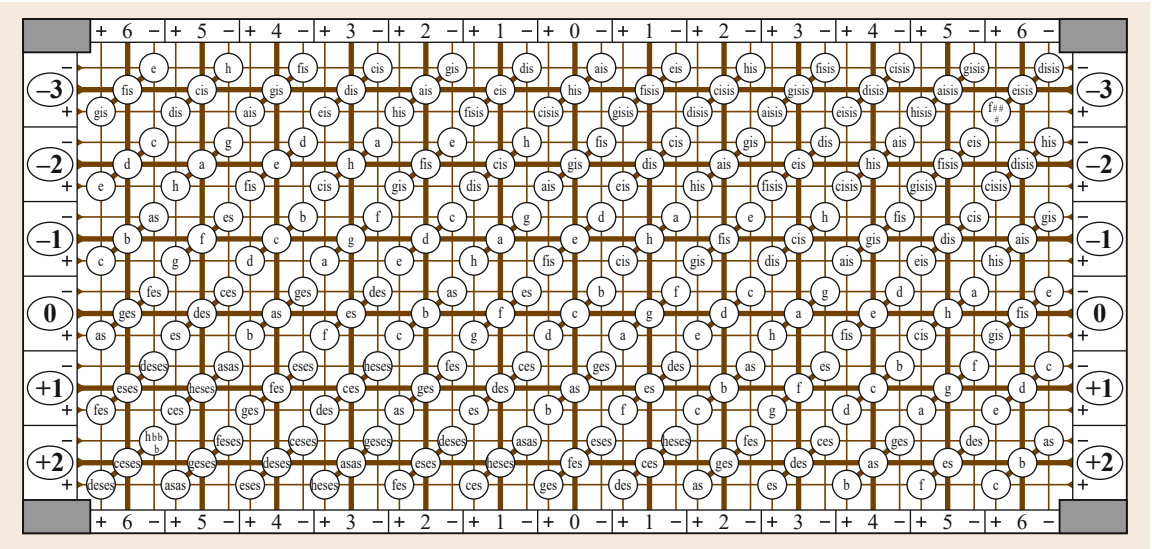


Fig. 31.35 3-D (three-dimensional) tone net (after [31.52, 529])

would gear our perception of actual music as played so that we, cognitively, would *hear* correct pitch relations even if intonation was deficient. Though such a process of *pitch correction* (within limits) governed by apperception of tonal relations plays a role for listeners, the alternative of matching correct intonation of tones/pitches to an appropriate melodic and harmonic analysis of musical works seems reasonable.

Among the scholars who, in the 20th century, worked on advanced systems of harmony as well as on the development of musical instruments suited to musical performance in tunings close to JI were the composers and musicologists *Siegfried Karg-Elert* [31.531] and *Eivind Groven* [31.532] (see articles in [31.533]), the physicist and music theorist *Adriaan Fokker* [31.382, 383] and the musicologist *Martin Vogel* [31.52].

References

31.1 H.P. Hesse: The judgment of musical intervals. In: *Music, Mind, and Brain*, ed. by M. Clynes (Plenum, London 1982) pp. 217–225

31.2 O. Külpe: *Grundriss der Psychologie auf experimenteller Grundlage* (Engelmann, Leipzig 1893)

31.3 A. Houtsma: Pitch salience of various complex sounds, Music Percept. **1**, 296–307 (1984)

31.4 H. Dai: Harmonic pitch: Dependence on resolved partials, spectral edges and combination tones, Hearing Res. **270**, 143–150 (2010)

31.5 N. Viemeister, C. Plack: Time analysis. In: *Human Psychophysics*, ed. by W. Yost, A. Popper, R. Fay (Springer, New York 1993) pp. 116–154

31.6 D. Eddins, D. Green: Temporal Integration and temporal resolution. In: *Hearing*, ed. by B. Moore (Academic, San Diego 1995) pp. 207–242

31.7 B. Moore: *Introduction to the Psychology of Hearing*, 5th edn. (Academic, Amsterdam 2007)

31.8 W. Bürck, P. Kotowski, H. Lichte: Der Aufbau des Tonhöhenbewußtseins, Elektr. Nachrichtentechn. **12**, 326–333 (1935)

31.9 A. de Cheveigné: Pitch perception models. In: *Pitch. Neural Coding and Perception*, ed. by C. Plack, J. Oxenham, R. Fay, A. Popper (Springer, New York 2005) pp. 169–233

31.10 L. White, C. Plack: Temporal processing of the pitch of complex tones, J. Acoust. Soc. Am. **103**, 2051–2063 (1998)

31.11 E. Terhardt: *Akustische Kommunikation* (Springer, Berlin 1998)

31.12 E. Zwicker, H. Fastl: *Psychoacoustics. Facts and Models*, 2nd edn. (Springer, Berlin 1999)

31.13 W. Shofner: Comparative aspects of pitch perception. In: *Pitch. Neural Coding and Perception*, ed. by C. Plack, A. Oxenham, R. Fay, A. Popper (Springer, New York 2005) pp. 56–98

31.14 A. Kohlrausch, A. Houtsma: Pitch related to spectral edges of broadband signals, Phil. Trans. R. Soc. London Ser. B **336**, 375–382 (1992)

31.15 A. Houtsma: Pitch perception. In: *Hearing*, ed. by B. Moore (Academic, San Diego 1995) pp. 267–295

31.16 E. de Boer: On the 'residue' and auditory pitch perception. In: *Handbook of Sensory Physiology*, Vol. V, 3, ed. by W.D. Keidel, W.D. Neff (Springer, Berlin 1976) pp. 479–583

31.17 H.F. Cohen: *Quantifying Music. The Science of Music at the First Stage of the Scientific Revolution 1580–1650* (Reidel, Dordrecht 1984)

31.18 S. Dostrovsky, R. Cannon: Entstehung der musikalischen Akustik (1600–1750). In: C. Dahlhaus, S. Dostrovsky, J.T. Cannon, M. Lindley, D. Walker: *Hören, Messen und Rechnen in der frühen Neuzeit*, Geschichte der Musiktheorie, Vol. 6, ed. by F. Zaminer (Wissenschaftliche Buchgesellschaft, Darmstadt 1987) pp. 7–79

31.19 A. Wood: *The Physics of Music*, 6th edn. (Methuen, London 1962)

31.20 R. Beyer: *Sound of our Time. Two Hundred Years of Acoustics* (Springer, New York 1999)

31.21 A. Seebeck: Über die Bedingungen der Entstehung von Tönen, Annalen der Physik und Chemie **53**, 417–436 (1841)

31.22 A. Seebeck: Ueber die Sirene, Annalen Phys. Chem. **60**, 449–481 (1843)

31.23 S. Ohm: Ueber die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen, Annalen Phys. Chem. **59**, 513–565 (1843)

31.24 J.F. Schouten: The residue, a new component in subjective sound analysis, Proc. R. Netherl. Acad. Arts Sci. **43**, 356–365 (1940)

31.25 J. Schouten: The residue revisited. In: *Frequency Analysis and Periodicity Detection in Hearing*, ed. by R. Plomp, G. Smoorenburg (Sijthoff, Leiden 1970) pp. 43–54

31.26 H.P. Hesse: *Die Wahrnehmung von Tonhöhe und Klangfarbe als Probleme der Hörtheorie* (Gerig, Köln 1972)

31.27 A. Seebeck: Über die Definition des Tones, Annalen Phys. Chem. **63**, 353–368 (1844)

31.28 H. von Helmholtz: *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (Vieweg, Braunschweig 1870), 3rd edn. 1870, 5th edn. 1896, 6th edn. 1913

31.29 G. von Békésy: *Experiments in Hearing* (Wiley, New York 1960)

31.30 R. Plomp: *Experiments on Tone Perception* (van Gorcum, Assen 1966)

31.31 R. Plomp: *Aspects of Tone Sensation* (Academic, London 1976)

31.32 J. Schouten: The perception of pitch, Philips Techn. Rev. **5**, 286–294 (1940)

31.33 J. Schouten: The residue and the mechanism of hearing, Proc. R. Netherl. Acad. Arts Sci. **43**, 991–999 (1940)

31.34 E. de Boer: *On the 'Residue' in Hearing*, Ph.D. Thesis (Univ. Amsterdam. s'Gravenhage, Uitgeverij Excelsior 1956)

31.35 R. Ritsma: Periodicity detection. In: *Frequency Analysis and Periodicity Detection in Hearing*, ed. by R. Plomp, G. Smoorenburg (Sijthoff, Leiden 1970) pp. 250–263

31.36 B. Delgutte: Physiological models for basic auditory percepts. In: *Auditory Computation*, ed. by H. Hawkins, T. McMullen, A. Popper, R. Fay (Springer, New York 1996) pp. 157–220

31.37 G. Révész: *Zur Grundlegung der Tonpsychologie* (Veit, Leipzig 1913)

31.38 J.-P. Rameau: Démonstration du Principe de l'Harmonie (Paris: Durand Pissot 1750)

31.39 C. Stumpf: *Tonpsychologie*, Vol. 1 (Barth, Leipzig 1883)

31.40 C. Stumpf: *Tonpsychologie*, Vol. 2 (Barth, Leipzig 1890)

31.41 J. Rose, J. Brugge, D. Anderson, J. Hind: Phase-locked response to low-frequency tones in single auditory-nerve fibers of the squirrel monkey, J. Neurophysiol. **30**, 769–793 (1967)

31.42 E. von Hornbostel: Psychologie der Gehörserscheinungen. In: *Handbuch der normalen und pathol. Physiol*, Vol. 11, ed. by A. Bethe (Springer, Berlin 1926) pp. 701–730

31.43 L. Demany, C. Semal: Dichotic fusion of two tones one octave apart: Evidence for internal octave templates, J. Acoust. Soc. Am. **83**, 687–695 (1988)

31.44 P. Cariani, B. Delgutte: Neural correlates of the pitch of complex tones. I. Pitch and pitch salience, J. Neurophysiol. **76**, 1698–1716 (1996)

31.45 P. Cariani, B. Delgutte: Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch, J. Neurophysiol. **76**, 1717–1734 (1996)

31.46 M. McKinney, M. Tramo, B. Delgutte: Neural correlates of the dissonance of musical intervals in the inferior colliculus. In: *Physiological and Psychophysical Bases of Auditory Function*, ed. by D.J. Breebart (Shaker, Maastricht 2001) pp. 83–89

31.47 K. Ohgushi, T. Hatoh: The musical pitch of high frequency tones, Adv. Biosci. **83**, 207–212 (1992)

31.48 A. Bachem: Tone height and tone chroma as two different pitch qualities, Acta Psychol. **7**, 80–88 (1950)

31.49 F. Brentano: Von der psychologischen Analyse der Tonqualitäten in ihre eigentlich ersten Elemente. In: *Atti del V congresso intern. Psichol. Roma 1905* (1906) pp. 157–165, repr. in: F. Brentano: *Untersuchungen zur Sinnespsychologie* (Duncker & Humblot, Leipzig 1907), pp. 101–125

31.50 H. Riemann: Das chromatische Tonsystem. In: *Präludien und Studien*, Vol. 1, ed. by H. Riemann (H. Seemann, Leipzig 1895) pp. 183–219

31.51 R. Bosanquet: *An Elementary Treatise on Musical Intervals and Temperament* (Macmillan, London 1876), repr. Diapason Pr., Utrecht 1987

31.52 M. Vogel: *On the Relations of Tone* (Verlag für Systematische Musikwissenschaft, Bonn 1993)

31.53 M. Lindley, R. Turner-Smith: *Mathematical Models of Musical Scales* (Verlag für Systematische Musikwissenschaft, Bonn 1993)

31.54 G. Albersheim: *Zur Psychologie der Ton- und Klangeigenschaften unter Berücksichtigung der Zweikomponententheorie und der Vokalsystematik* (Heitz, Straßburg 1939), repr. Körner, Baden-Baden 1975

31.55 G. Albersheim: *Zur Musikpsychologie*, 2nd edn. (Heinrichshofen, Wilhelmshaven 1979)

31.56 G. Révész: *Einführung in die Musikpsychologie* (Francke, Bern 1946)
31.57 H. Lotze: *Medicinische Psychologie oder Physiologie der Seele* (Weidmann, Leipzig 1852)
31.58 H. Lotze: *Geschichte der Aesthetik in Deutschland* (Cotta, München 1866)
31.59 C. Stumpf: Über neuere Untersuchungen zur Tonlehre, Beitr. Akust. Musikwiss. **8**, 305–344 (1914)
31.60 C. von Maltzew: Das Erkennen sukzessiv gegebener musikalischer Intervalle in der äußeren Tonregion, Beitr. Akust. Musikwiss. **7**, 37–133 (1913)
31.61 C. Plack, A. Oxenham: The psychophysics of pitch. In: *Pitch. Neural Coding and Perception*, ed. by C. Plack, A. Oxenham, R. Fay, A. Popper (Springer, New York 2005) pp. 7–55
31.62 W. Köhler: Akustische Untersuchungen (Ph.D. Thesis, Univ. Berlin), Beitr. Akust. Musikwiss. **4**, 134–182 (1909)
31.63 W. Köhler: Akustische Untersuchungen I–III, Z. Psychol. **54**, 241–289 (1910); **58**, 59–140 (1911); **64**, 92–105 (1913); **72**, 1–192 (1915)
31.64 G. Rich: A study of tonal attributes, Am. J. Psychol. **30**, 121–164 (1919)
31.65 W. Straub: Tonqualität und Tonhöhe, Arch. ges. Psychol. **69**, 289–395 (1929)
31.66 K. Miyazaki: Musical pitch identification by absolute pitch possessors, Percept. Psychophys. **44**, 501–512 (1988)
31.67 R. Boynton, C. Olson: Salience of chromatic basic color terms confirmed by three measures, Vision Res. **30**, 1311–1317 (1990)
31.68 M. Gazzaniga (Ed.): *The Cognitive Neurosciences*, 4th edn. (MIT Press, Cambridge 2009)
31.69 N. Birbaumer, R.F. Schmidt: *Biologische Psychologie*, 7th edn. (Springer, Heidelberg 2010)
31.70 L. Marks: On cross-modal similarity: The perceptual structure of pitch, loudness, and brightness, J. Exp. Psychol.: Hum. Percept. Perform. **15**, 586–602 (1989)
31.71 M. Vogel: *Die Enharmonik der Griechen. Bd 1: Tonsystem und Notation* (Gesellschaft zur Förderung der Systematischen Musikwissenschaft, Düsseldorf 1963)
31.72 J. Handschin: *Der Toncharakter. Eine Einführung in die Tonpsychologie* (Atlantis, Zürich 1948)
31.73 L. Manik: *Das Arabische Tonsystem im Mittelalter* (Brill, Leiden 1969)
31.74 W.J. Dowling, D.L. Harwood: *Music Cognition* (Academic, Orlando 1986)
31.75 A. Wright, J. Rivera, S. Hulse, M. Shyan, J. Neiworth: Music perception and octave generalization in Rhesus monkeys, J. Exp. Psych. **129**, 291–307 (2000)
31.76 F. Attneave, R. Olson: Pitch as a medium: A new approach to psychophysical scaling, Am. J. Psychol. **84**, 147–166 (1971)
31.77 P. Briley, C. Breakey, K. Krumbholz: Evidence for pitch chroma mapping in human auditory cortex, Cerebral Cortex **23**, 2601–2610 (2012)
31.78 R. Shepard: Circularity in judgements of relative pitch, J. Acoust. Soc. Am. **36**, 2346–2353 (1964)
31.79 R. Shepard: Demonstrations of circular components of pitch, J. Audio Eng. Soc. **31**, 641–649 (1983)
31.80 J.-C. Risset: Computer, synthesis, perception, paradoxes, Hamburger Jahrb. Musikwiss. **11**, 245–258 (1991)
31.81 A. Schneider: Was haben Ligetis Études pour piano mit Shepard-Skalen zu tun? Über "auditorische Illusionen", Vertige und Columna infinită. In: *Mikrotöne und mehr. Auf György Ligetis Hamburger Pfaden*, ed. by M. Stahnke (von Bockel, Hamburg 2005) pp. 81–104
31.82 H. Purwins, I. Normann, K. Obermayer: Unendlichkeit – Konstruktion musikalischer Paradoxien. In: *Mikrotöne und mehr. Auf György Ligetis Hamburger Pfaden*, ed. by M. Stahnke (von Bockel, Hamburg 2005) pp. 39–80
31.83 A. Houtsma, T. Rossing, W. Wagenaar: *Auditory Demonstrations* (Philips Acoust. Soc. Am. Eindhoven, New York 1987)
31.84 A. Schneider: Über ein Hörexperiment mit einer Stimmung Adriaan Fokkers. In: *Musikwissenschaft–Musikpraxis*, ed. by K. Bachmann, W. Thies (Mueller-Speiser, Anif/Salzburg 2000) pp. 142–153
31.85 H.P. Hesse: Sonanz- und Distanzurteil bei musikalischen Intervallen. Über die Verbindlichkeit mathematischer Intervalldefinitionen, Mikrotöne **2**, 147–158 (1988)
31.86 E. Terhardt: Die Tonhöhe harmonischer Klänge und das Oktavenintervall, Acustica **24**, 126–136 (1971)
31.87 J. Sundberg, J. Lindquist: Musical octaves and pitch, J. Acoust. Soc. Am. **54**, 920–929 (1973)
31.88 B. Rosner: Stretching and compression in the perception of musical intervals, Music Perception **17**, 101–114 (1999)
31.89 F. Loosen: Intonation of solo violin performance with reference to equally tempered, Pythagorean, and just intonations, J. Acoust. Soc. Am. **93**, 525–539 (1993)
31.90 A. Rakowski: Acoustics and psychoacoustics of musical performance. In: *Intern. Musikwiss. Kgr. Mozartjahr 1991 Baden–Wien*, ed. by I. Fuchs (H. Schneider, Tutzing 1993) pp. 303–312
31.91 A. Rakowski: Categorical perception in absolute pitch, Arch. Acoust. **18**, 515–523 (1993)
31.92 J. Fyk: *Melodic Intonation, Psychoacoustics, and the Violin* (Organon, Zielona Góra 1995)
31.93 A. Schneider: *Tonhöhe–Skala–Klang. Akustische, Tonometrische und Psychoakustische Studien auf Vergleichender Grundlage* (Orpheus, Bonn 1997)
31.94 D. Webster, A. Popper, R. Fay (Eds.): *The Mammalian Auditory Pathway: Neuroanatomy* (Springer, New York 1992)
31.95 A. Popper, R. Fay (Eds.): *The Mammalian Auditory Pathway: Neurophysiology* (Springer, New York 1992)
31.96 R. Fay, A. Popper (Eds.): *Comparative Hearing: Mammals* (Springer, New York 1994)
31.97 D. Oertel, R. Fay, A. Popper (Eds.): *Integrative Functions in the Mammalian Auditory Pathway*

(Springer, New York 2002)
31.98 J.O. Pickles: *Introduction to the Physiology of Hearing*, 3rd edn. (Emerald, Binkley 2008)
31.99 R. Altschuler, R. Bobbin, B. Clopton, D. Hoffman (Eds.): *Neurobiology of Hearing: The Central Auditory System* (Raven, New York 1991)
31.100 G. Ehret, R. Romand (Eds.): *The Central Auditory System* (Oxford Univ. Press, New York 1997)
31.101 S. Gelfand: *Hearing. An Introduction to Psychological and Physiological Acoustics*, 4th edn. (Dekker, New York 2004)
31.102 J. Eggermont: Between sound and perception: Reviewing the search for a neural code, Hearing Res. **157**, 1–42 (2001)
31.103 J. Rauschecker, B. Tian: Mechanisms and streams for processing of "what" and "where" in auditory cortex, Proc. Natl. Acad. Sci. **97**, 11800–11806 (2000)
31.104 J. Rauschecker: Processing streams in auditory cortex. In: *Neural Correlates of Auditory Cognition*, ed. by Y. Cohen, A. Popper, R. Fay (Springer, New York 2013) pp. 7–43
31.105 J. Casseday, T. Fremouw, E. Covey: The Inferior colliculus: A hub for the central auditory system. In: *Integrative Functions in the Mammalian Auditory Pathway*, ed. by D. Oertel, R. Fay, A. Popper (Springer, New York 2002) pp. 238–318
31.106 G. Ehret: The auditory midbrain, a "shunting yard" of acoustical information processing. In: *The Central Auditory System*, ed. by G. Ehret, R. Romand (Oxford Univ. Press, Oxford 1997) pp. 259–316
31.107 M. Banks, P. Smith: Thalamocortical relations. In: *The Auditory Cortex*, ed. by J. Winer, C. Schreiner (Springer, New York 2011) pp. 75–94
31.108 J. Schnupp, C. Honey, B. Willmore: Neural correlates of auditory object perception. In: *Neural Correlates of Auditory Cognition*, ed. by Y. Cohen, A. Popper, R. Fay (Springer, New York 2013) pp. 115–149
31.109 G. Langner: *The Neural Code of Pitch and Harmony* (Cambridge Univ. Press, Cambridge 2015)
31.110 G. Eska: *Schall und Klang. Wie und was wir Hören* (Birkhäuser, Basel 1997)
31.111 S. Handel: *Listening. An Introduction to the Perception of Auditory Events* (MIT Press, Cambridge 1989)
31.112 R. Aibara, J. Welch, S. Puria, R. Goode: Human middle-ear transfer function and cochlear input impedance, Hearing Res. **152**, 100–109 (2001)
31.113 E. de Boer: Auditory physics. Physical principles of hearing theory. II, Phys. Rep. **105**, 141–226 (1984)
31.114 E. de Boer: Mechanics of the cochlea: Modeling efforts. In: *The Cochlea*, ed. by P. Dallos, A. Popper, R. Fay (Springer, New York 1996) pp. 258–317
31.115 L. Robles, M. Ruggero: Mechanics of mammalian cochlea, Physiol. Rev. **81**, 1305–1352 (2001)
31.116 J. Bronzino (Ed.): *The Biomedical Engineering Handbook*, 2nd edn. (Springer, Heidelberg 2000)
31.117 J. Lighthill: Energy flow in the cochlea, J. Fluid Mech. **106**, 149–213 (1981)
31.118 E. de Boer: Auditory physics. Physical principles of hearing theory. I, Phys. Rep. **62**, 87–164 (1980)
31.119 E. de Boer: Auditory physics. Physical principles of hearing theory. III, Phys. Rep. **203**, 125–231 (1991)
31.120 S. Ramamoorthy, D.-J. Zha, A. Nuttall: The biophysical origin of the traveling-wave dispersion in the cochlea, Biophys. J. **99**, 1687–1695 (2010)
31.121 G. Emadi, C.P. Richter, P. Dallos: Stiffness of the Gerbil basilar membrane: Radial and longitudinal variations, J. Neurophysiol. **91**, 474–488 (2004)
31.122 B. Kimberley, D. Brown, J. Eggermont: Measuring human cochlear travelling wave delay using distortion product emission phase response, J. Acoust. Soc. Am. **94**, 1343–1350 (1993)
31.123 W. Rhode, A. Recio: Study of mechanical motions in the basal region of the chinchilla cochlea, J. Acoust. Soc. Am. **107**, 3317–3332 (2000)
31.124 I. Russell, K. Nielsen: The location of the cochlear amplifier: Spatial representation of a single tone on the guinea pig basilar membrane, Proc. Nat. Acad. Sci. USA **94**, 2660–2664 (1997)
31.125 R. Patuzzi: Cochlear micromechanics and macromechanics. In: *The Cochlea*, ed. by P. Dallos, A. Popper, R. Fay (Springer, New York 1996) pp. 186–257
31.126 M. Ruggero, N. Rich, A. Recio, S. Narayan, L. Robles: Basilar-membrane responses to tones at the base of the chinchilla cochlea, J. Acoust. Soc. Am. **101**, 2151–2163 (1997)
31.127 M. Ruggero, S. Narayan, A. Temchin, A. Recio: Mechanical bases of frequency tuning and neural excitation at the base of the cochlea: Comparison of basilar-membrane vibrations and auditory nerve-fiber-responses in chinchilla, Proc. Nat. Acad. Sci. **97**, 11744–11750 (2000)
31.128 D. Greenwood: Critical bandwidth and the frequency coordinates of the basilar membrane, J. Acoust. Soc. Am. **33**, 1344–1356 (1961)
31.129 D. Greenwood: A cochlear frequency-position function for several species – 29 years later, J. Acoust. Soc. Am. **87**, 2592–2605 (1990)
31.130 D. Greenwood: The mel scale's disqualifying bias and a consistency of pitch-difference equisections in 1956 with equal cochlear distances and equal frequency ratios, Hearing Res. **103**, 199–224 (1997)
31.131 E. Lopez-Poveda: Spectral processing in the peripheral auditory system: Facts and models. In: *Auditory Spectral Processing*, Intern. Rev. Neurobiol., Vol. 70, ed. by D. Irvine, M. Malmierca (Elsevier, Amsterdam 2005) pp. 7–48
31.132 S. Jia, D. He: Motility-associated hair-bundle motion in mammalian outer hair cells, Nature Neurosci. **8**, 1028–1034 (2005)
31.133 J. Ashmore: The remarkable cochlear amplifier, Hearing Res. **266**, 1–17 (2010)
31.134 E. Evans: Auditory processing of complex sounds: An overview, Phil. Trans. R. Soc. London Ser. B **336**, 295–306 (1992)
31.135 R. Nobili, F. Mammano: Biophysics of the cochlea II: Stationary nonlinear phenomenology, J. Acoust. Soc. Am. **99**, 2244–2255 (1996)

31.136 F. Mammano, R. Nobili: Biophysics of the cochlea: Linear approximation, J. Acoust. Soc. Am. **93**, 3320–3332 (1993)

31.137 A. Schneider, R. Mores: Fourier-time-transformation (FTT), analysis of sound and auditory perception. In: *Sound–Perception–Performance*, ed. by R. Bader (Springer, Cham 2013) pp. 299–329

31.138 M. Russo, N. Rožić, M. Stella: Biophysical cochlear model: Time-frequency analysis and signal reconstruction, Acta Acustica/Acustica **97**, 632–640 (2011)

31.139 R. Nobili, F. Mammano, J. Ashmore: How well do we understand the cochlea?, Trends Neurosci. **21**, 159–167 (1998)

31.140 A. Oxenham, J. Bernstein, H. Penagos: Correct tonotopic representation is necessary for complex pitch perception, Proc. Natl. Acad. Sci. USA **101**, 1421–1425 (2004)

31.141 M. Chatterjee, J. Zwislocki: Cochlear mechanics of frequency and intensity coding. I. The place code for pitch, Hearing Res. **111**, 65–75 (1997)

31.142 A. Hudspeth: The cellular basis of hearing: The biophysics of hair cells, Science **230**(4727), 745–752 (1985)

31.143 A. Palmer: Neural signal processing. In: *Hearing*, ed. by B. Moore (Academic, San Diego 1995) pp. 75–121

31.144 H.P. Zenner: *Hören. Physiologie, Biochemie, Zell- und Neurobiologie* (Thieme, Stuttgart, New York 1994)

31.145 L. van Noorden: Two channel pitch perception. In: *Music, Mind, and Brain*, ed. by M. Clynes (Plenum, London 1982) pp. 251–269

31.146 M. Escabi, H.L. Read: Neural mechanisms for spectral analysis in the auditory midbrain, thalamus, and cortex. In: *Auditory Spectral Processing*, International Review of Neurobiology, Vol. 70, ed. by M. Malmierca, D. Irvine (Elsevier, Amsterdam 2005) pp. 207–252

31.147 C. Schreiner, R. Froemke, C. Atencio: Spectral processing in auditory cortex. In: *The Auditory Cortex*, ed. by J. Winer (Springer, New York 2011) pp. 209–234

31.148 M. Sachs, E. Young: Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rates, J. Acoust. Soc. Am. **66**, 470–480 (1979)

31.149 R. Winslow, P. Barta, M. Sachs: Rate coding in the auditory nerve. In: *Auditory Processing of Complex Sounds*, ed. by W. Yost, C. Watson (Erlbaum, Hillsdale 1987) pp. 212–224

31.150 P. Cariani: Temporal coding of periodicity pitch in the auditory system: An overview, Neural Plasticity **6**, 147–172 (1999)

31.151 E. Wever, C. Bray: Action currents in the auditory nerve in response to acoustical stimulation, Proc. Natl. Acad. Sci. USA **16**, 344–350 (1930)

31.152 E. Wever: *Theory of Hearing*, 2nd edn. (Wiley, New York 1957)

31.153 D. Kim, W. Rhode, S. Greenberg: Responses of cochlear nucleus neurons to speech signals: Neural encoding of pitch, intensity and other parameters. In: *Auditory Frequency Selectivity*, ed. by B. Moore, R. Patterson (Plenum, London 1986) pp. 281–286

31.154 N. Wiener: Generalized harmonic analysis, Acta math **55**, 117–258 (1930)

31.155 A. Khintchine: Korrelationstheorie der stationären stochastischen Prozesse, Math. Annalen **109**, 604–615 (1934)

31.156 J. Licklider: Basic correlates of the auditory stimulus. In: *Handbook of Experimental Psychology*, ed. by S.S. Stevens (Wiley, New York 1951) pp. 985–1039

31.157 J. Licklider: A duplex theory of pitch perception, Experientia **7**, 128–134 (1951)

31.158 L. Jeffress: A place theory of sound localization, J. Comp. Physiol. Psychol. **41**, 35–39 (1948)

31.159 W. Hartmann: *Signals, Sound, and Sensation* (AIP/Springer, New York 1998)

31.160 J. Licklider: Auditory frequency analysis. In: *Information Theory*, ed. by C. Cherry (Butterworth, London 1956) pp. 253–268

31.161 J. Culling, Q. Summerfield, D. Marshall: Dichotic pitches as illusions of binaural unmasking. I. Huggins' pitch and the "binaural edge pitch", J. Acoust. Soc. Am. **103**, 3509–3526 (1998)

31.162 M. Slaney, R. Lyon: A perceptual pitch detector. In: *Intern. Conf. Acoust., Speech and Signal Process. (ICASSP-90), Albuquerque*, Vol. I (1990) pp. 357–360

31.163 R. Meddis, M. Hewitt: Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification, J. Acoust. Soc. Am. **89**, 2866–2882 (1991)

31.164 R. Meddis, M. Hewitt: Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity, J. Acoust. Soc. Am. **89**, 2883–2894 (1991)

31.165 R. Lyon, S. Shamma: Auditory representations of timbre and pitch. In: *Auditory Computation*, ed. by H. Hawkins, T. McMullen, A. Popper, R. Fay (Springer, New York 1996) pp. 221–270

31.166 M. Slaney: Connecting correlograms to neurophysiology and psychoacoustics. In: *Psychophysical and Physiological Advances in Hearing*, ed. by A. Palmer (Whurr, London 1998) pp. 563–569

31.167 D. Pressnitzer, R. Patterson, K. Krumbholz: The lower limit of melodic pitch, J. Acoust. Soc. Am. **109**, 2074–2084 (2001)

31.168 A. Schneider, V. Tsatsishvili: Perception of intervals at very low frequencies: Some experimental findings. In: *Systematic Musicology: Empirical and Theoretical Studies*, ed. by A. Schneider, A. von Ruschkowski (P. Lang, Frankfurt 2011) pp. 99–125

31.169 D. Junius: *Temporal and Spatial Aspects of Hearing as Revealed by Auditory Evoked Potentials*, Ph.D Thesis (Carl von Ossietzky Universität, Oldenburg 2005)

31.170 S. Greenberg, W. Rhode: Periodicity coding in cochlear nerve and ventral cochlear nucleus. In: *Auditory Processing of Complex Sounds*, ed. by W. Yost, C. Watson (Erlbaum, Hillsdale 1987)

pp. 225–236
31.171 E. Smith, M. Lewicki: Efficient auditory coding, Nature **439**, 978–982 (2006)
31.172 T. Tolonen, M. Karjalainen: A computationally efficient multipitch analysis model, IEEE Trans. Speech Audio Process. **8**, 708–716 (2000)
31.173 E. Evans: Pitch and cochlear nerve fibre temporal discharge patterns. In: *Hearing. Physiological Bases and Psychophysics*, ed. by R. Klinke, R. Hartmann (Springer, Berlin 1983) pp. 140–145
31.174 E. Javel, J. Horst, G. Farley: Coding of complex tones in temporal response patterns of auditory nerve fibers. In: *Auditory Processing of Complex Sounds*, ed. by W. Yost, C. Watson (Erlbaum, Hillsdale 1987) pp. 237–246
31.175 W. Horst, E. Javel, G. Farley: Coding of spectral fine structure in the auditory nerve. I. Fourier analysis of period and interspike interval histograms, J. Acoust. Soc. Am. **79**, 398–416 (1986)
31.176 L. Cedolin, B. Delgutte: Pitch of complex tones: Rate-place and interspike interval representations in the auditory nerve, J. Neurophysiol. **94**, 347–362 (2005)
31.177 L. Cedolin, B. Delgutte: Spatio-temporal representation of complex tones in the auditory nerve. In: *Hearing. From Sensory Processing to Perception*, ed. by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, J. Verhey (Springer, Berlin 2007) pp. 61–69
31.178 C. Schreiner, G. Langner: Laminar fine structure of frequency organization in auditory midbrain, Nature **388**, 383–386 (1997)
31.179 C. Schreiner, G. Langner: Coding of temporal patterns in the central auditory nervous system. In: *Auditory Function. Neurobiological Bases of Hearing*, ed. by G. Edelman, W. Gall, W. Cowan (Wiley, New York 1988) pp. 337–361
31.180 G. Langner, C. Schreiner, M. Albert: Tonotopy and periodotopy in the auditory midbrain of cat and Guinea fowl, Adv. Biosci. **83**, 241–248 (1992)
31.181 G. Langner, C. Schreiner, U. Biebel: Functional implications of frequency and periodicity coding in the auditory midbrain. In: *Psychophysical and Physiological Advances in Hearing*, ed. by A. Palmer (Whurr, London 1998) pp. 277–284
31.182 G. Langner: Periodicity coding in the auditory system, Hearing Res. **60**, 115–142 (1992)
31.183 A. Palmer, I. Winter: Cochlear nerve and cochlear nucleus responses to the fundamental frequency of voiced speech sounds and harmonic complex tones, Adv. Biosci. **83**, 231–237 (1992)
31.184 G. Langner, M. Albert, T. Briede: Temporal and spatial coding of periodicity information in the inferior colliculus of the awake chinchilla (Chinchilla laniger), Hearing Res. **168**, 110–130 (2002)
31.185 P. Joris, C. Schreiner, A. Rees: Neural processing of amplitude-modulated sounds, Physiol. Rev. **84**, 541–577 (2004)
31.186 A. Houtsma, J. Goldstein: The central origin of the pitch of complex tones: Evidence from musical interval recognition, J. Acoust. Soc. Am. **51**, 520–529 (1972)
31.187 J. Goldstein: An optimum processor theory for the central formation of the pitch of complex tones, J. Acoust. Soc. Am. **54**, 1496–1516 (1973)
31.188 A. Gerson, J. Goldstein: Evidence for a general template in central optimal processing for pitch of complex tones, J. Acoust. Soc. Am. **63**, 498–510 (1978)
31.189 J. Goldstein, A. Gerson, P. Srulovicz, M. Furst: Verification of the optimal probabilistic basis of aural processing in pitch of complex tones, J. Acoust. Soc. Am. **63**, 486–497 (1978)
31.190 P. Srulovicz, J. Goldstein: A central spectral model: A synthesis of auditory nerve-timing and place cues in monaural communication of frequency spectrum, J. Acoust. Soc. Am. **73**, 1266–1276 (1983)
31.191 F. de Ribaupierre: Acoustical information processing in the auditory thalamus and cerebral cortex. In: *The Central Auditory System*, ed. by G. Ehret, R. Romand (Oxford Univ. Press, Oxford 1997) pp. 317–397
31.192 J. Eggermont, X. Wang: Temporal coding in auditory cortex. In: *The Auditory Cortex*, ed. by J. Winer (Springer, New York 2011) pp. 309–328
31.193 Y. Fishman, I. Volkov, D. Noh, C. Garell, H. Bakken, J. Arezzo, M. Howard, M. Steinschneider: Consonance and dissonance of musical chords: Neural correlates in auditory cortex of monkeys and humans, J. Neurophysiol. **86**, 2761–2788 (2001)
31.194 R. Zatorre: Pitch perception of complex tones and human temporal lobe function, J. Acoust. Soc. Am. **84**, 566–572 (1988)
31.195 M. Tramo, P. Cariani, C. Koh, N. Makris, L. Braida: Neurophysiology and neuroanatomy of pitch perception: Auditory cortex, Annals N.Y. Acad. Sci. **1060**, 148–175 (2005)
31.196 R. Zatorre, J. Zarate: Cortical processing of music. In: *The Human Auditory Cortex*, ed. by D. Poeppel, T. Overath, A. Popper, R. Fay (Springer, New York 2012) pp. 261–294
31.197 T. Hackett, J. Kaas: Auditory cortex in primates: Functional subdivisions and processing streams. In: *The Cognitive Neurosciences*, 3rd edn., ed. by M. Gazzaniga (MIT Press, Cambridge 2004) pp. 215–232
31.198 C. Pantev, M. Hoke, K. Lehnertz, B. Lütkenhöner, G. Anogianakis, W. Wittkowski: Tonotopic organization of the human auditory cortex revealed by transient auditory evoked mangnetic fields, Electroencephalography Clin. Neurophysiol. **69**, 160–170 (1988)
31.199 M. Sutter, C. Schreiner: Topography of intensity tuning in cat primary auditory cortex: Single-neuron versus multiple-neuron recordings, J. Neurophysiol. **73**, 190–204 (1995)
31.200 C. Pantev, O. Bertrand, C. Eulitz, C. Verkindt, S. Hampson, G. Schuierer, T. Elbert: Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous electric and magnetic recordings, Electroencephalography Clin. Neurophysiol. **94**, 26–40 (1995)

31.201 D. Bendor, X. Wang: The neuronal representation of pitch in primate auditory cortex, Nature **436**, 1161–1165 (2005)

31.202 D. Schwarz, W. Tomlinson: Spectral response patterns of auditory cortex neurons to harmonic complex tones in alert monkey (Macaca mulata), J. Neurophysiol. **64**, 282–298 (1990)

31.203 D. Hall, C. Plack: Searching for a pitch center in human auditory cortex. In: *Hearing. From Sensory Processing to Perception*, ed. by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, J. Verhey (Springer, Berlin 2007) pp. 83–89

31.204 D. Hall, D. Barker: Coding of basic acoustical and perceptual components of sound in human auditory cortex. In: *The Human Auditory Cortex*, ed. by D. Poeppel, T. Overath, A. Popper, R. Fay (Springer, New York 2012) pp. 165–197

31.205 P. Van Dijk, D. Langers: Mapping tonotopy in human auditory cortex. In: *Basic Aspects of Hearing*, ed. by B.C. Moore, R.D. Patterson, I.M. Winter, R.P. Carlyon, H. Gockel (Springer, Berlin 2012) pp. 419–425

31.206 I. Nelken: Feature detection by the auditory cortex. In: *Integrative Functions in the Mammalian Auditory Pathway*, ed. by D. Oertel, R. Fay, A. Popper (Springer, New York 2002) pp. 358–416

31.207 J. Kanwal, G. Ehret: Communication sounds and their cortical representation. In: *The Auditory Cortex*, ed. by J. Winer, C. Schreiner (Springer, New York 2011) pp. 343–368

31.208 I. Nelken: Processing of complex sounds in the auditory system, Curr. Opin. Neurobiol. **18**, 413–417 (2008)

31.209 N. Staeren, H. Renvall, F. de Martino, E. Goebel, E. Formisano: Sound categories are represented as distributed patterns in the human auditory cortex, Curr. Biol. **19**, 498–502 (2009)

31.210 J. Bizley, K. Walker, B. Silverman, A. King, J. Schnupp: Interdependent Encoding of pitch, timbre, and spatial location in auditory cortex, J. Neurosci. **18**, 2064–2075 (2009)

31.211 I. Nelken, O. Bar-Yosef: Neurons and objects: The case of auditory cortex, Front. Neurosci. **2**, 107–113 (2008)

31.212 T. Griffith, C. Micheyl, T. Overath: Auditory object analysis. In: *The Human Auditory Cortex*, ed. by D. Poeppel, T. Overath, A. Popper, R. Fay (Springer, New York 2012) pp. 199–223

31.213 X. Wang: Neural coding strategies in auditory cortex, Hearing Res. **229**, 81–93 (2007)

31.214 R. Zatorre, A. Evans, E. Meyer: Neural mechanisms underlying melodic perception and memory for pitch, J. Neurosci. **14**, 1908–1919 (1994)

31.215 R. Zatorre, P. Belin: Spectral and temporal processing in human auditory cortex, Cerebral Cortex **11**, 946–953 (2001)

31.216 T. Pasternak, M. Greenlee: Working memory in primate sensory systems, Nature Rev. Neurosci. **6**, 97–107 (2005)

31.217 B. Postle: Working memory as an emergent property of the mind and brain, Neuroscience **139**, 23–38 (2006)

31.218 A. Rakowski: Short-term memory for pitch. In: *Recent Trends in Hearing Research*, ed. by H. Fastl, S. Kuwano, A. Schick (BIS, Oldenburg 1996) pp. 99–128

31.219 R. Kochmann: Über musikalische Gedächtnisbilder, Zeitschr. Angew. Psychol. **22**, 329–351 (1923)

31.220 E. Altenmüller: How many music centers are in the brain?, Annals N.Y. Acad. Sci. **930**, 273–280 (2001)

31.221 C. Pantev, A. Engelien, V. Candia, T. Elbert: Representational cortex in musicians. Plastic alterations in response to musical practice, Annals N.Y. Acad. Sci. **930**, 300–314 (2001)

31.222 G. Musacchia, M. Sams, E. Skoe, N. Kraus: Musicians have enhanced subcortical auditory and audiovisual processing of speech and music, Proc. Natl. Acad. Sci. **104**, 15894–15898 (2007)

31.223 G. Bidelman, A. Krishnan, J. Gandour: Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch, Eur. J. Neurosci. **33**, 530–538 (2011)

31.224 G. Bidelman, A. Krishnan: Neural correlates of consonance, dissonance, and the hierarchy of pitch in the human brainstem, J. Neurosci. **29**, 13165–13171 (2009)

31.225 M. Tramo, P. Cariani, B. Delgutte, L. Braida: Neurobiological functions for the theory of harmony in western tonal music, Annals N.Y. Acad. Sci. **930**, 92–116 (2001)

31.226 R. Meddis, E. Lopez-Poveda: Auditory periphery: From pinna to auditory nerve. In: *Computational Models of the Auditory System*, ed. by R. Meddis, E. Lopez-Poveda, R. Fay, A. Popper (Springer, New York 2010) pp. 7–38

31.227 R. Hut, M. Boone, A. Giesolf: Cochlear modelling as time-frequency analysis tool, Acustica **92**, 629–636 (2006)

31.228 E. Terhardt, G. Stoll, M. Seewann: Algorithm for extraction of pitch and pitch salience from complex tone signals, J. Acoust. Soc. Am. **71**, 679–688 (1982)

31.229 R. Patterson, K. Robinson, J. Holdsworth, D. McKeown, C. Zhang, M. Allerhand: Complex sounds and auditory images, Adv. Biosci. **83**, 429–443 (1992)

31.230 E. Terhardt, G. Stoll, M. Seewann: Pitch of complex signals according to virtual-pitch theory: Tests, examples, and predictions, J. Acoust. Soc. Am. **71**, 671–678 (1982)

31.231 E. Terhardt: Fourier transformation of time signals: Conceptual revision, Acustica **57**, 242–256 (1985)

31.232 F. Wightman: The pattern-transformation model of pitch, J. Acoust. Soc. Am. **54**, 407–416 (1973)

31.233 M. Cohen, S. Grossberg, L. Wyse: A spectral network model of pitch perception, J. Acoust. Soc. Am. **98**, 862–879 (1995)

31.234 R. Patterson, M. Allerhand, C. Giguère: Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform, J. Acoust. Soc. Am. **98**, 1890–1894 (1995)

31.235 R. Patterson, T. Irino: Modeling temporal asymmetry in the auditory system, J. Acoust. Soc. Am. **104**, 2967–2979 (1998)

31.236 E. Lopez-Poveda, R. Meddis: A human nonlinear cochlear filterbank, J. Acoust. Soc. Am. **110**, 3107–3118 (2001)

31.237 X. Zhang, M. Heinz, I. Bruce, L. Carney: A phenomenological model for the responses of auditory nerve fibers: I. Nonlinear tuning with compression and suppression, J. Acoust. Soc. Am. **109**, 648–670 (2001)

31.238 W.D. Keidel (Ed.): *Physiologie des Gehörs. Akustische Informationsverarbeitung* (Thieme, Stuttgart 1975)

31.239 R. Meddis, L. O'Mard: Virtual pitch in a computational physiological model, J. Acoust. Soc. Am. **120**, 3861–3869 (2006)

31.240 A. de Cheveigné: Pitch shifts of mistuned partials: A time-domain model, J. Acoust. Soc. Am. **106**, 887–897 (1999)

31.241 K. Davis, K. Hancock, B. Delgutte: Computational models of inferior colliculus neurons. In: *Computational Models of the Auditory System*, ed. by R. Meddis, E. Lopez-Poveda, R. Fay, A. Popper (Springer, New York 2010) pp. 129–176

31.242 T. Chi, P. Ru, S. Shamma: Multiresolution spectrotemporal analysis of complex sounds, J. Acoust. Soc. Am. **118**, 887–906 (2005)

31.243 J. Grose, J. Hall, E. Buss: Across-channel spectral processing. In: *Auditory Spectral Processing*, Intern. Rev. Neurobiol., Vol. 70, ed. by M. Malmierca, D. Irvine (Elsevier, Amsterdam 2005) pp. 87–119

31.244 R. Meddis, R. Delahaye, L. O'Mard, C. Sumner, A. Fantini, I. Winter, D. Pressnitzer: A Model of signal processing in the cochlear nucleus: Comodulation masking release, Acustica **88**, 387–398 (2002)

31.245 P. Boersma: Accurate short-term analysis of the fundamental frequency and the harmonic-to-noise ratio of a sampled sound, Proc. Inst. Phonetic Sci. Univ. Amsterdam **17**, 97–110 (1993)

31.246 D. Hermes: Measurement of pitch by subharmonic summation, J. Acoust. Soc. Am. **83**, 257–264 (1988)

31.247 P. Boersma, D. Weenink: *Praat. Doing Phonetics by Computer* (Institute of Phonetics Univ. Amsterdam, Amsterdam 2012)

31.248 A. Schneider: Inharmonic sounds: Implications as to pitch, timbre, and consonance, J. New Music Res. **29**, 275–301 (2000)

31.249 A. Schneider: Complex inharmonic sounds, perceptual ambiguity, and musical imagery. In: *Musical Imagery*, ed. by R.I. Godøy, H. Jørgensen (Swets Zeitlinger, Lisse 2001) pp. 95–116

31.250 A. Schneider, K. Frieler: Perception of harmonic and inharmonic sounds: Results from ear models. In: *Computer Music Modeling and Retrieval. Genesis of Meaning in Sound and Music*, ed. by S. Ystad, R. Kronland-Martinet, K. Jensen (Springer, Berlin 2009) pp. 18–44

31.251 A. Schneider, M. Leman: Sound, pitches and tuning of a historic carillon. In: *Studies in Musical Acoustics and Psychoacoustics*, ed. by A. Schneider (Springer, Cham 2017) pp. 247–298

31.252 A. Schneider, M. Leman: Sonological and psychoacoustic characteristics of carillon bells. In: *The Quality of Bells (Proc. 16th Meet. FWO Res. Soc. Foundations Music Res., Sept. 2002)*, ed. by M. Leman (IPEM, Univ. Ghent, Ghent 2002)

31.253 R. Patterson, R. Milroy, M. Allerhand: What is the octave of a harmonically rich note?, Contemp. Music Rev. **9**, 69–81 (1993)

31.254 W. Wundt: Über psychologische Methoden, Philosophische Studien **1**, 1–38 (1882)

31.255 E. Zwicker: *Psychoakustik* (Springer, Berlin 1982)

31.256 C. Wier, W. Jesteadt, D. Green: Frequency discrimination as a function of frequency and sensation level, J. Acoust. Soc. Am. **61**, 178–184 (1977)

31.257 J. Meyer: Zur Tonhöhenempfindung bei musikalischen Klängen in Abhängigkeit vom Grad der Gehörschulung, Acustica **42**, 189–204 (1979)

31.258 W. Stebbins, D. Moody: How monkeys hear the world: Auditory perception in nonhuman primates. In: *Comparative Hearing: Mammals*, ed. by R. Fay, A. Popper (Springer, New York 1994) pp. 97–133

31.259 H. Fletcher: Auditory patterns, Rev. Mod. Phys. **12**, 47–66 (1940)

31.260 H. Fletcher: Loudness, masking, and their relation to the hearing process and the problem of noise measurement, J. Acoust. Soc. Am. **9**, 275–293 (1938)

31.261 B. Scharf: Critical bands. In: *Foundations of Modern Auditory Theory*, Vol. I, ed. by J. Tobias (Academic, New York 1970) pp. 157–202

31.262 E. Zwicker, G. Flottorp, S. Stevens: Critical bands with loudness summation, J. Acoust. Soc. Am. **29**, 548–557 (1957)

31.263 B. Moore: Frequency analysis and masking. In: *Hearing*, ed. by B. Moore (Academic, San Diego 1995) pp. 161–205

31.264 B. Moore: Basic psychophysics of human spectral processing. In: *Auditory Spectral Processing*, International Review of Neurobiology, Vol. 70, ed. by M. Malmierca, D. Irvine (Elsevier, Amsterdam 2005) pp. 49–86

31.265 E. Zwicker, B. Scharf: A model of loudness summation, Psych. Rev. **72**, 3–26 (1965)

31.266 D. Howard, J. Angus: *Acoustics and Psychoacoustics*, 2nd edn. (Focal, Oxford 2001)

31.267 B. Moore, B. Glasberg: Suggested formulae for calculating auditory-filter bandwidths and excitation patterns, J. Acoust. Soc. Am. **74**, 750–753 (1983)

31.268 A. Schneider, A. von Ruschkowski, R. Bader: Klangliche Rauhigkeit, ihre Wahrnehmung und Messung. In: *Musical Acoustics, Neurocognition and Psychology of Music*, ed. by R. Bader (P. Lang, Frankfurt 2009) pp. 103–148

31.269 C. Jurado, C. Pedersen, B. Moore: Psychophysical tuning curves for frequencies below 100 Hz, J. Acoust. Soc. Am. **129**, 3166–3180 (2011)

31.270 A. Mayer: Researches in acoustics, No. IX, Philos. Mag. **37**(125), 259–288 (1894)

31.271 A. Spanias, T. Painter, A. Venkatraman: *Audio Signal Processing and Coding* (Wiley, Hoboken 2007)

31.272 D. Pressnitzer: *Perception de rugosité psychoacoustique: d'un attribut élémentaire de l'audition à l'écoute musicale*, Thèse de doctorat (Université Paris 6, Paris 1998)

31.273 P. Daniel, R. Weber: Psychoacoustical roughness: Implementation of an optimized model, Acustica **83**, 113–123 (1997)

31.274 E. Terhardt: On the perception of periodic sound fluctuations (roughness), Acustica **30**, 201–213 (1974)

31.275 F. Födermayr, W. Deutsch: Zur Forschungsstrategie der vergleichend-systematischen Musikwissenschaft, Musicol. Austriaca **17**, 163–180 (1998)

31.276 P. Vassilakis: Auditory roughness as a means of musical expression. In: *Perspectives in Systematic Musicology*, ed. by R. Kendall, R. Savage (Dept. Ethnomusicology, UCLA, Los Angeles 2005) pp. 119–144

31.277 W. Sethares: Local consonance and the relationship between timbre and scale, J. Acoust. Soc. Am. **94**, 1218–1228 (1993)

31.278 W. Sethares: *Tuning, Timbre, Spectrum, Scale* (Springer, London 1998)

31.279 R. Plomp, W. Levelt: Tonal consonance and critical bandwidth, J. Acoust. Soc. Am. **38**, 548–560 (1965)

31.280 E. Meyer, G. Buchmann: Die Klangspektren der Musikinstrumente, Sitzungsber. Preuss. Akad. Wiss. Math. Phys. Kl. **XXXII**, 735–778 (1931)

31.281 R. Ritsma: Frequencies dominant in the perception of the pitch of complex sounds, J. Acoust. Soc. Am. **42**, 191–198 (1967)

31.282 A. Houtsma: What determines musical pitch?, J. Music Theory **15**, 138–157 (1971)

31.283 A. Houtsma, J. Smurzynski: Pitch identification and discrimination for complex tones with many harmonics, J. Acoust. Soc. Am. **87**, 304–310 (1990)

31.284 G. Bruhn: *Über die Hörbarkeit von Glockenschlagtönen. Untersuchungen zum Residualproblem* (Bosse, Regensburg 1980)

31.285 E. Terhardt, M. Seewann: Auditive und objektive Bestimmung der Schlagtonhöhe von historischen Kirchenglocken, Acustica **54**, 129–144 (1984)

31.286 W.A. Hibbert: *The Quantification of Strike Pitch and Pitch Shifts in Church Bells*, Ph.D. Thesis (The Open Univ., Milton Keynes 2008)

31.287 A. Schneider: Virtual pitch and musical instrument acoustics. The case of idiophones. In: *Musik im virtuellen Raum. KlangArt-Kongress 1997*, ed. by B. Enders, J. Stange-Elbe (Universitätsverlag Rasch, Osnabrück 2000) pp. 397–417

31.288 G. Tartini: *Trattato di musica seconda la vera scienza dell'armonia* (G. Manfre', Padua 1754), German translation with commentary by A. Rubeli: *Traktat über die Musik gemäß der wahren Wissenschaft von der Harmonie*. Düsseldorf: Gesellschaft zur Förderung der Syst. Musikwiss. 1966

31.289 G. Smoorenburg: Combination tones and their origin, J. Acoust. Soc. Am. **52**, 615–632 (1972)

31.290 G. Smoorenburg: Audibility region of combination tones, J. Acoust. Soc. Am. **52**, 603–614 (1972)

31.291 R. Plomp: Detectability threshold for combination tones, J. Acoust. Soc. Am. **37**, 1110–1123 (1965)

31.292 C. Stumpf: Beobachtungen über Kombinationstöne, Z. Psychol. **55**, 1–142 (1910)

31.293 J. Roederer: *The Physics and Psychophysics of Music: An Introduction*, 3rd edn. (Springer, New York 1995)

31.294 P. Dutilleux: Verstärkung von Differenztönen (f_2-f_1), Bericht Tonmeistertagung Karlsruhe **19**, 798–806 (1996)

31.295 C. Stumpf: Geschichte des Consonanzbegriffs. Erster Theil. In: *Abhandl. der Philos.-philol. Classe der königl. Bayer. Akad. der Wiss.*, Vol. 21 (Verl. der Akad., München 1901)

31.296 A. Barker: *The Science of Harmonics in Classical Greece* (Cambridge Univ. Press, Cambridge 2007)

31.297 I. Düring: *Die Harmonielehre des Klaudios Ptolemaios*, Göteborgs Högskolas Årsskrift T. 36, no. 1 (Wettergren Kerbers, Göteborg 1930)

31.298 C. Stumpf: *Konsonanz und Dissonanz* (Barth, Leipzig 1898)

31.299 C. Stumpf: *Die Sprachlaute* (J. Springer, Berlin 1926)

31.300 A. Schneider: Foundations of systematic musicology. A study in history and theory. In: *Systematic and Comparative Musicology: Concepts, Methods, Findings*, ed. by A. Schneider (P. Lang, Frankfurt 2008) pp. 11–61

31.301 A. Schneider: 'Verschmelzung', tonal fusion, and consonance: Carl Stumpf revisited. In: *Music, Gestalt, and Computing. Studies in Cognitive and Systematic Musicology*, ed. by M. Leman (Springer, Berlin 1997) pp. 117–143

31.302 C. Stumpf: Neueres über Tonverschmelzung, Z. Psychol. **15**, 280–303 (1897)

31.303 C. Stumpf: *Erkenntnislehre*, Vol. 1 (Barth, Leipzig 1939)

31.304 A. Beurmann, A. Schneider, E. Lauer: Klanguntersuchungen an der Arp-Schnitger-Orgel zu St. Jacobi, Hamburg, Syst. Musikwiss. Syst. Musicol. **6**, 151–187 (1998)

31.305 A. Schneider: Music theory: Speculation, reasoning, experience. In: *Musiktheorie / Musikwissenschaft. Geschichte–Methoden–Perspektiven*, ed. by T. Janz, P. Sprick (Olms, Hildesheim 2010) pp. 53–97

31.306 A. Schneider: Change and continuity in sound analysis: A review of concepts in regard to musical acoustics, music perception, and transcription. In: *Sound–Perception–Performance*, ed. by R. Bader (Springer, Cham 2013) pp. 71–111

31.307 R. Plomp: Hoe wij horen. Over de toon die de muziek maakt [How we hear. On the tone that makes up music. 2 booklets and cd] (Plomp, Breukelen 1998)

31.308 D. Hall, J. Hess: Perception of musical interval tuning, Music Percept. **2**, 166–195 (1984)

31.309 J. Vos: Spectral effects in the perception of pure and tempered intervals: Discrimination and beats, Percept. Psychophys. **35**, 173–185 (1984)

31.310 J. Vos: Purity ratings of tempered 5ths and major thirds, Music Percept. **3**, 221–258 (1986)
31.311 H. Husmann: *Eine neue Konsonanztheorie* (Müller-Thiergarten, Heidelberg 1953)
31.312 G. Bidelman, M. Heinz: Auditory-nerve responses predict pitch attributes related to musical consonance-dissonance for normal and impaired hearing, J. Acoust. Soc. Am. **130**, 1488–1502 (2011)
31.313 F. Opelt: *Allgemeine Theorie der Musik auf den Rhythmus der Klangwellenpulse gegründet* (Barth, Leipzig 1852)
31.314 T. Lipps: *Psychologische Studien*, 2nd edn. (Dürr, Leipzig 1905)
31.315 J. Boeyens, D. Levendis: *Number Theory and the Periodicity of Matter* (Springer, Dordrecht 2008)
31.316 M. Ebeling: Neuronal periodicity detection as a basis for the perception of consonance: A mathematical model of tonal fusion, J. Acoust. Soc. Am. **124**, 2320–2329 (2008)
31.317 H. von Helmholtz: Zählen und Messen, erkenntnistheoretisch betrachtet. In: *Wissenschaftliche Abhandlungen*, Vol. 3, ed. by H. von Helmholtz (Barth, Leipzig 1895) pp. 356–391
31.318 W. Sethares: *Tuning, Timbre, Spectrum, Scale*, 2nd edn. (Springer, London 2004)
31.319 D. Benson: *Music: A Mathematical Offering* (Cambridge Univ. Press, Cambridge 2006)
31.320 B. Repp: Categorical perception: Methods, issues, findings. In: *Speech and Language: Advances in Basic Research and Practice*, Vol. 10, ed. by N. Lass (Academic, Orlando 1984) pp. 243–335
31.321 M. Studdert-Kennedy, A. Liberman, K. Harris, F. Cooper: Motor theory of speech perception: A reply to Lane's critical review, Psych. Rev. **77**, 234–249 (1970)
31.322 S. Harnad (Ed.): *Categorical Perception. The Groundwork of Cognition* (Cambridge Univ. Press, Cambridge 1987)
31.323 E. Husserl: *Erfahrung und Urteil. Untersuchungen zur Genealogie der Logik*, 5th edn. (Meiner, Hamburg 1976), ed. by L. Landgrebe
31.324 G. Lakoff: *Women, Fire, and Dangerous Things. What Categories Reveal About the Mind* (Univ. Chicago Press, Chicago 1987)
31.325 D. Dennett: *Consciousness Explained* (Penguin, London 1993)
31.326 A. Bregman: *Auditory Scene Analysis* (MIT Press, Cambridge 1990)
31.327 C. Krumhansl: *Cognitive Foundations of Musical Pitch* (Oxford University Press, New York 1990)
31.328 M. Leman: *Music and Schema Theory* (Springer, Berlin 1995)
31.329 M. Leman: A model of retroactive tone-center perception, Music Percept. **12**, 439–471 (1995)
31.330 M. Leman: Naturalistic approaches to musical semiotics and the study of causal musical signification. In: *Music and Signs. Semiotic and Cognitive Studies in Music*, ed. by I. Zannos (ASCO, Bratislava 1999) pp. 11–38
31.331 J. Kazem-Bek: Informationstheorie und Analyse musikalischer Werke, Archiv für Musikwiss. **35**, 62–75 (1979)
31.332 W.D. Keidel: *Biokybernetik des Menschen* (Wissenschaftliche Buchgesellschaft, Darmstadt 1989)
31.333 A. Beurmann, A. Schneider: Struktur, Klang, Dynamik. Akustische Untersuchungen an Ligetis *Atmosphères*, Hamburger Jahrb. Musikwiss. **11**, 311–334 (1991)
31.334 W.R. Garner: *The Processing of Information and Structure* (Erlbaum, Potomac 1974)
31.335 P. Zimbardo: *Psychologie*, 5th edn. (Springer, Berlin 1992), ed. by S. Hoppe-Graff, B. Keller
31.336 K. Johnson: *Acoustic and Auditory Phonetics*, 3rd edn. (Wiley-Blackwell, Malden 2012)
31.337 P. Kuhl: Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech sound categories, J. Acoust. Soc. Am. **70**, 340–349 (1981)
31.338 H. Repp, A. Liberman: Phonetic category boundaries are flexible. In: *Categorical Perception*, ed. by S. Harnad (Cambridge Univ. Press, Cambridge 1987) pp. 89–112
31.339 L. Holt, A. Lotto: Speech perception as categorization, Attention, Percept. Psychophys. **72**, 1218–1227 (2010)
31.340 A. Rakowski: Intonation variants of musical intervals in isolation and in musical context, Psychol. Music **18**, 60–72 (1990)
31.341 A. Rakowski: Context-dependent intonation variants of melodic intervals. In: *Music, Language, Speech, and Brain*, ed. by J. Sundberg, L. Nord, R. Carlson (Macmillan, London 1991) pp. 203–211
31.342 J. Siegel, W. Siegel: Categorical perception of tonal intervals: Musicians can't tell sharp from flat, Percept. Psychophys. **21**, 399–407 (1977)
31.343 E. Burns, D. Ward: Categorical perception – Phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals, J. Acoust. Soc. Am. **63**, 456–468 (1978)
31.344 N.A. Garbuzow: *Muzikant, Issledowatel', Pedagog (Musician, Scientist, Teacher). Collected Works* (Izdatel'stwo 'Muzika', Moscow 1980), ed. by O. Sachaltuewa, O. Sokolova
31.345 J. Fricke: Klangbreite und Tonempfindung. Bedingungen kategorialer Wahrnehmung aufgrund experimenteller Untersuchung der Intonation, Musikpsychologie **5**, 67–87 (1988)
31.346 D. Howard, S. Rosen, V. Broad: Major–minor triad identification and discrimination by musically trained and untrained listeners, Music Percept. **10**, 205–220 (1992)
31.347 E. Burns, S. Campbell: Frequency and frequency-ratio resolution by possessors of absolute and relative pitch: Examples of categorical perception?, J. Acoust. Soc. Am. **96**, 2704–2719 (1994)
31.348 F. Sixtl: *Meßmethoden der Psychologie. Theoretische Grundlagen und Probleme*, 2nd edn. (Beltz, Weinheim 1982)
31.349 W. Torgerson: *Theory and Method of Scaling* (Wiley, New York 1958)
31.350 R. Shepard: Toward a universal law of generalization for psychological science, Science **237**, 1317–1323 (1987)

31.351 R. Nosofsky: Similarity scaling and cognitive process models, Ann. Rev. Psych. **43**, 25–53 (1992)
31.352 I. Kant: *Kritik der Reinen Vernunft*, 2nd edn. (Hartknoch, Riga 1787)
31.353 E. Husserl: *Erfahrung und Urteil. Untersuchungen zur Genealogie der Logik* (Academia, Prag 1939), ed. by L. Landgrebe; 5th edn. Meiner, Hamburg 1972
31.354 N. Rescher: *Cognitive Systematization. A Systems-Theory Approach to A Coherentist Theory of Knowledge* (Blackwell, Oxford 1979)
31.355 L. Thurstone: A law of comparative judgment, Psych. Rev. **101**, 266–270 (1994)
31.356 L. Thurstone: A law of comparative judgment, Psych. Rev. **34**, 273–286 (1927)
31.357 D. Laming: *Mathematical Psychology* (Academic, London 1973)
31.358 R. Luce: Thurstone and sensory scaling: Then and now, Psych. Rev. **101**, 271–277 (1994)
31.359 G. Miller: The magical number seven, plus or minus two: Some limits on our capacitzy for processing information, Psych. Rev. **63**, 81–97 (1956)
31.360 I. Pollack, L. Ficks: Information of elementary multidimensional auditory displays, J. Acoust. Soc. Am. **26**, 155–158 (1954)
31.361 R. Shiffrin, R. Nosofsky: Seven plus or minus two: A commentary on capacity limitations, Psych. Rev. **101**, 357–361 (1994)
31.362 H.P. Reinecke: *Experimentelle Beiträge zur Psychologie des musikalischen Hörens* (Sikorski, Hamburg 1964)
31.363 J. Fricke: *Intonation und musikalisches Hören* (Electronic, Osnabrück 2012)
31.364 H. Touma: *Maqam Bayati in the Arabian Taqsim*, Ph.D. Thesis (Das Arabische Buch, Berlin 1980)
31.365 H. Touma: *Die Musik der Araber*, 3rd edn. (Heinrichshofen, Wilhelmshaven 1998)
31.366 N. Jairazbhoy: *The Rāgs of North Indian Music* (Faber Faber, London 1971)
31.367 K. Signell: *Makam. Modal Practice in Turkish art music* (DaCapo, New York 1986)
31.368 W. Van der Meer: *Hindustani Music in the 20th Century* (Nijhoff, The Hague 1980)
31.369 J.-C. Chabrier: Éléments d'une approche comparative des échelles théoriques arabo-irano-turques, Rév. Musicol. **71**, 39–78 (1985)
31.370 I. Zannos: *Ichos und Makam. Vergleichende Untersuchungen zum Tonsystem der griechisch-orthodoxen Kirchenmusik und der türkischen Kunstmusik* (Orpheus, Bonn 1994)
31.371 B. Bozkurt, O. Yarman, K. Karaosmanoğlu, C. Akkoş: Weighting diverse theoretical models on Turkish maqam music against pitch measurements. A comparison of peaks automatically derived from frequency histograms with proposed scale tones, J. New Music Res. **38**, 45–70 (2009)
31.372 M. Panteli, H. Purwins: A quantitative comparison of chrysanthine theory and performance practice of scale tuning, steps, and prominence of the octoechos in byzantine chant, J. New Music Res. **42**, 205–221 (2013)
31.373 B. Snyder: *Music and Memory* (MIT Press, Cambridge 2000)
31.374 J. Sundberg: *The Science of Musical Sound* (Academic, London 1991)
31.375 A. Schneider: On categorical perception of pitch and the recognition of intonation variants. In: *Brain, Mind, and Physics*, ed. by P. Pylkkänen, P. Pyllkö, A. Hautamäki (IOS, Amsterdam 1997) pp. 250–261
31.376 A. Schneider: Über Stimmung und Intonation, Syst. Musikwiss. Syst. Musicol. **6**, 27–49 (1998)
31.377 B. Hagerman, J. Sundberg: Fundamental frequency adjustment in barbershop singing, STL-QPRS KTH Stockholm **1**(1980), 28–42 (1980)
31.378 W. Thies: Intonationsmessungen an einem Vokalquartett, Syst. Musikwiss. Syst. Musicol. **6**, 51–72 (1998)
31.379 H. Hafke-Dys, A. Preis, D. Trojan: Violonists' perceptions of and motor reactions to fundamental frequency shifts introduced in auditory feedback, Acustica **102**, 155–158 (2016)
31.380 N.A. Garbuzow: *Zonnaja Priroda zbukowycotnogo Slucha (The zonal nature of tonal hearing)* (Izdatel'stwo Akad. Nauk USSR, Leningrad 1948)
31.381 M. Schouten, A. van Hessen: Modeling phoneme perception. I: Categorical perception, J. Acoust. Soc. Am. **92**, 1841–1855 (1992)
31.382 A. Fokker: *Rekenkundige Bespiegeling der Muziek* (Noorduijn, Gorinchem 1945)
31.383 A. Fokker: *New Music with 31 Notes* (Verlag für Systematische Musikwissenschaft, Bonn 1975)
31.384 D. Ward, E. Burns: Absolute pitch. In: *The Psychology of Music*, ed. by D. Deutsch (Academic, Orlando 1982) pp. 431–451
31.385 E.-M. Heyde: *Was ist absolutes Hören? Eine musikpsychologische Untersuchung* (Profil, München 1987)
31.386 A. Takeuchi, S. Hulse: Absolute pitch, Psychol. Bull. **113**, 345–361 (1991)
31.387 D. Levitin, S. Rogers: Absolute pitch: Perception, coding, and controversies, Trends Cogn. Sci. **9**, 26–33 (2005)
31.388 D. Deutsch: Absolute pitch. In: *The Psychology of Music*, 3rd edn., ed. by D. Deutsch (Elsevier, Amsterdam 2013) pp. 141–182
31.389 O. Abraham: Das absolute Tonbewusstsein. Psychologisch-musikalische Studie, Sammelbände der Intern. Musikges. **3**, 1–86 (1901)
31.390 A. Wellek: *Das absolute Gehör und seine Typen*, 2nd edn. (Francke, Bern, München 1970)
31.391 D. Ward: Absolute pitch. In: *The Psychology of Music*, 2nd edn., ed. by D. Deutsch (Academic, Orlando 1999) pp. 265–298
31.392 A. Costall: The relativity of absolute pitch. In: *Musical Structure and Cognition*, ed. by P. Howell, I. Cross, R. West (Academic, London 1985) pp. 189–208
31.393 K. Miyazaki: The speed of musical pitch identification by absolute pitch possessors, Music Percept. **8**, 177–188 (1990)
31.394 K. Miyazaki: Perception of musical intervals by absolute pitch possessors, Music Percept. **9**, 413–

426 (1992)
31.395 K. Miyazaki: Absolute pitch as an inability: Identification of musical intervals in a tonal context, Music Percept. **11**, 55–72 (1993)
31.396 G. Révész: *Erwin Nyiregyhazi. Psychologische Analyse eines musikalisch hervorragenden Kindes* (Veit, Leipzig 1916)
31.397 C. Stumpf: Akustische Versuche mit Pepito Arriola, Beitr. Akust. Musikwiss. **4**, 105–115 (1909)
31.398 P. Gregersen: Instant recognition: The genetics of pitch perception, Am. J. Hum. Gen. **62**, 221–223 (1998)
31.399 P. Gregersen, E. Kowalsky, N. Kohn, E.W. Marvin: Absolute pitch: Prevalence, ethnic variation, and estimation of the genetic component, Am. J. Hum. Gen. **65**, 911–913 (1999)
31.400 S. Baharloo, P. Johnston, S. Service, J. Gitschier, N. Freimer: Absolute pitch: An approach for identification of genetic and nongenetic components, Am. J. Hum. Gen. **62**, 224–231 (1998)
31.401 S. MacDougall-Shackleton, S. Hulse: Concurrent absolute and relative pitch processing in European starlings (Sturnus vulgaris), J. Comp. Psych. **110**, 139–146 (1996)
31.402 S. Hulse, A. Takeuchi, R. Braaten: Perceptual invariances in the comparative psychology of music, Music Percept. **10**, 151–184 (1992)
31.403 S. Trehub, L. Trainor: Listening strategies in infancy: The roots of music and language development. In: *Thinking in Sound: The Cognitive Psychology of Human Audition*, ed. by S. McAdams, E. Bigand (Oxford Univ. Press, New York 1993) pp. 278–327
31.404 P. Kuhl, F.M. Tsao, H.M. Liu, Y. Zhang, B. de Boer: Language–Culture–Mind–Brain. Progress at the margins between disciplines, Annals N.Y. Acad. Sci. **935**, 136–174 (2006)
31.405 K. Miyazaki, Y. Ogawa: Learning absolute pitch by children: A cross-sectional study, Music Percept. **24**, 63–78 (2006)
31.406 A. Ellis, A. Mendel: *Studies on the History of Musical Pitch* (Knuf, Amsterdam 1968)
31.407 B. Haynes: *A History of Performing Pitch. The Story of "A"* (Scarecrow, Lanham 2002)
31.408 P. Brady: Fixed-scale mechanism of absolute pitch, J. Acoust. Soc. Am. **48**, 883–887 (1970)
31.409 R. Zatorre: Absolute pitch: A model for understanding the influence of genes and development on neural and cognitive function, Nature Neurosci. **6**, 692–695 (2003)
31.410 F. Russo, D. Windell, L. Cuddy: Learning the "special note": Evidence for a critical period of absolute pitch acquisition, Music Percept. **21**, 119–127 (2003)
31.411 K. Miyazaki: Absolute pitch identification: Effects of timbre and pitch region, Music Percept. **7**, 1–14 (1989)
31.412 A. Wellek: *Musikpsychologie und Musikästhetik* (Akademische Verlagsgesellschaft, Frankfurt 1963)
31.413 K. Frieler, T. Fischinger, K. Schlemmer, K. Lothwesen, K. Jakubowski, D. Müllensiefen: Absolute memory for pitch: A comparative replication of Levitin's 1994 study in six European labs, Musicae Scientiae **17**, 334–349 (2013)
31.414 J. Andres: Grundbegriffe der multivariaten Datenanalyse. In: *Handbuch Quantitative Methoden*, ed. by E. Erdfelder, R. Mausfeld, T. Meiser, G. Rudinger (Beltz/Psychologie Verlagsunion, Weinheim 1996) pp. 169–184
31.415 S. Stevens: On the theory of scales of measurement, Science **103**, 677–680 (1946)
31.416 S. Stevens: Sensation and psychological measurement. In: *Foundations of Psychology*, ed. by E. Boring, H. Langfeld, H. Weld (Wiley, New York 1948) pp. 250–268
31.417 T. Rossing: *The Science of Sound* (Addison-Wesley, Reading 1990)
31.418 A. Schneider: Research on tone systems, tunings, and intonation: Concepts, methods, and findings. In: *Proc. VII Intern. Symp. Syst. Comparative Musicol./III Intern. Conf. Cogn. Musicol., Jyväskylä* (2001) pp. 156–164
31.419 F. Fernando-Marandola: New perspectives on interactive field experiments, Yearbook Trad. Music **34**, 163–186 (2002)
31.420 R. Bader: Buddhism, animism, and entertainment in Cambodian melismatic chanting smot. History and tonal system. In: *Systematic Musicology: Empirical and Theoretical Studies*, ed. by A. Schneider, A. von Ruschkowski (Lang, Frankfurt 2011) pp. 283–305
31.421 G. Manley, A.N. Popper, R. Fay (Eds.): *Evolution of the Vertebrate Auditory System* (Springer, New York 2004)
31.422 P. Szöke: Ist das hinter dem Horizont der Tonkunst verborgene säkulare Rätsel des Ursprungs der Musik lösbar?, Syst. Musikwiss. **2**, 71–108 (1994)
31.423 N. Wallin: *Biomusicology: Neurophysiological, Neuropsychological, and Evolutionary Perspectives on the Origins and Purposes of Music* (Pendragon, Stuyvesant 1991)
31.424 C. Stumpf: *Anfänge der Musik* (Barth, Leipzig 1911)
31.425 A. Daniélou: *Traité de Musicologie Comparée* (Hermann, Paris 1959)
31.426 M. Kolinski: Classification of tonal structures, illustrated by a comparative chart . . ., Stud. Ethnomusicol. **1**, 38–76 (1961)
31.427 C. Sachs: *The Wellsprings of Music* (Nijhoff, The Hague 1962)
31.428 L. Vikar: Archaic Types of Finno-Ugrian melody, Studia Musicol. Acad. Scient. Hung. **14**, 53–91 (1972)
31.429 D. McAllester: North America/native America. In: *Worlds of Music. An Introduction to Music of the World's Peoples*, ed. by J. Titon, J. Koetting, D. McAllester, D. Reck, M. Slobin (Schirmer, Collier Macmillan, New York 1984) pp. 12–63
31.430 P. Toner: Melody and the musical articulation of Yolngu identities, Yearbook Trad. Music **25**, 69–95 (2003)
31.431 W. Graf: Zur Rolle der Teiltonreihe in der Gestaltung klingend tradierter Musik. In: *Festschrift für Kurt Blaukopf*, ed. by I. Bontinck (Universal Edition, Wien 1975) pp. 48–66

31.432 W. Deutsch, F. Födermayr: Zum Problem des zweistimmigen Sologesanges mongolischer und Turkvölker. In: *Von der Vielfalt musikalischer Kultur*, ed. by R. Schumacher (Müller-Speiser, Salzburg-Anif 1992) pp. 133–145

31.433 O. Elschek: *Fujara. The Slovak Queen of European Flutes* (Music Centre, Bratislava 2006)

31.434 G. Kubik: *Theory of African Music*, Vol. I (Heinrichshofen, Wilhelmshaven 1994)

31.435 H. Powers: *Mode. The New Grove Dict. Music and Musicians*, Vol. 12 (Macmillan, London 1980) pp. 376–450

31.436 A. Szabó: *The Beginnings of Greek Mathematics* (Reidel, Dordrecht 1978)

31.437 O. Busch: *Logos Syntheseōs. Die Euklidische Sectio Canonis, Aristoxenos, und die Rolle der Mathematik in der antiken Musiktheorie* (Staatliches Institut für Musikforschung, Berlin 1998)

31.438 A. Barker: *Greek Musical Writings: Harmonic and Acoustic Theory*, Vol. 2 (Cambridge Univ. Press, Cambridge 1989)

31.439 B. Van der Waerden: Die Arithmetik der Pythagoreer, Math. Annalen **120**, 676–700 (1949)

31.440 B. Münxelhaus: *Pythagoras Musicus. Zur Rezeption der Pythagoreischen Musiktheorie als Quadrivialer Wissenschaft im lateinischen Mittelalter* (Verlag für Systematische Musikwissenschaft, Bonn 1976)

31.441 K.J. Sachs: Musikalische Elementarlehre im Mittelalter. In: M. Bernhard, A. Borst, D. Illmer, A. Riethmüller, K.J. Sachs: *Rezeption des Antiken Fachs im Mittelalter*, Geschichte der Musiktheorie, Vol. 3, ed. by F. Zaminer (Wissenschaftliche Buchgesellschaft, Darmstadt 1990) pp. 105–161

31.442 I. De Geer: *Earl, Saint, Bishop, Skald and Music. The Orkney Earldom of the 12th Century* (Institutionen för Musikvetenskap, Uppsala Univ. 1985)

31.443 C. Huffman: *Archytas of Tarentum. Pythagorean, Philosopher and Mathematician King* (Cambridge Univ. Press, Cambridge 2005)

31.444 G. of Arezzo: *Guidonis Aretini micrologus*, Corpus scriptorum de musica, Vol. 4 (Am. Inst. Musicol, Rome 1955), ed. by J. Smits van Waesberghe

31.445 K.J. Sachs: *Mensura fistularum. Die Mensurierung der Orgelpfeifen im Mittelalter*, Vol. 2 (Musikwissenschaftliche Verlagsgesellschaft, Murrhardt 1980)

31.446 N. Phillips: Notationen und Notationslehren von Boethius bis zum 12. Jahrhundert. In: M. Huglo, C. Atkinson, C. Meyer, K. Schlager, N. Phillips: *Die Lehre vom einstimmigen liturgischen Gesang*, Geschichte der Musiktheorie, Vol. 4, ed. by F. Zaminer (Wissenschaftliche Buchgesellschaft, Darmstadt 2000) pp. 293–623

31.447 H. Klotz: *Über die Orgelkunst der Gotik, der Renaissance und des Barock*, 2nd edn. (Bärenreiter, Kassel 1975)

31.448 G. Zarlino: *Istitutioni Harmoniche*, rev. 3rd edn. (F. Senese, Venezia 1573)

31.449 B. Ramos de Pareja: *Practica musica* (Bologna 1482), repr. ed. by J. Wolf, Breitkopf & Haertel, Leipzig 1901

31.450 P. Barbieri: *Enharmonic Instruments and Music 1470–1900* (Levante, Latina 2008)

31.451 M. Lindley: Stimmung und Temperatur. In: C. Dahlhaus, S. Dostrovsky, J.T. Cannon, M. Lindley, D.P. Walker: *Hören, Messen und Rechnen in der frühen Neuzeit*, Geschichte der Musiktheorie, Vol. 6, ed. by F. Zaminer (Wissenschaftliche Buchgesellschaft, Darmstadt 1987) pp. 109–331

31.452 F.J. Ratte: *Die Temperatur der Clavierinstrumente* (Bärenreiter, Kassel 1991)

31.453 E. Blackwood: *The Structure of Recognizable Diatonic Tunings* (Princeton Univ. Press, Princeton 1985)

31.454 R. Rasch: Description of regular 12 tone musical tunings, J. Acoust. Soc. Am. **73**, 1023–1035 (1983)

31.455 M. Praetorius: *Syntagma Musicorum II: de Organographia* (Holwein, Wolffenbüttel 1619)

31.456 A. Schneider, R. von Busch: Zur Verwendung der Viertelkomma-Mittelton-Stimmung auf historischen Orgeln: Einige empirische Daten und musikalisch-akustische Kriterien, Acta Organologica **34**, 437–454 (2015)

31.457 L. Fogliano: *Musica Theorica docte simul ac Dilucide Pertractata* (G. Nicolini da Sabbio, Venezia 1529)

31.458 F. Rempp: Elementar- und Satzlehre von Tinctoris bis Zarlino. In: F.A. Gallo, R. Groth, C. Palisca, F. Rempp: *Italienische Musiktheorie im 16. und 17. Jahrhundert*, Geschichte der Musiktheorie, Vol. 7, ed. by F. Zaminer (Wissenschaftliche Buchgesellschaft, Darmstadt 1989) pp. 39–220

31.459 J. Barbour: *Tuning and Temperament* (Michigan State Univ. Press, East Lansing 1951), repr. Da Capo, New York 1972

31.460 R. Rasch: Why were enharmonic keyboards built? From Nicola Vicentino (1555) to Michael Bulyowski (1699), Schweizer Jahrb. Musikwiss. N.F. **22**, 25–93 (2003)

31.461 A. Werckmeister: *Musicalische Temperatur* (Calvisius, Quedlinburg 1691), repr. ed. by R. Rasch, Diapason, Utrecht 1983

31.462 R. Rasch: *Introduction to A. Werckmeister, Musicalische Temperatur* (repr. Diapason, Utrecht 1983) pp. 8–51

31.463 J. Barnes: Bach's keyboard temperament. Internal evidence from the well-rempered Clavier, Early Music **7**, 236–249 (1979)

31.464 H. Kellner: Das wohltemperirte Clavier: Implications de l'accord inégal pour l'œuvre et son autographe, Rév. Musicol. **71**, 143–157 (1985)

31.465 M. Lindley: A quest for Bach's ideal style of organ temperament. In: *Stimmungen im 17. und 18. Jahrhundert. Vielfalt oder Konfusion?*, ed. by G. Fleischhauer, M. Lustig, W. Ruf, F. Zschoch (Stiftung Kloster Michaelstein, Blankenburg 1997) pp. 45–67

31.466 B. Billeter: *Bachs Klavier- und Orgelmusik* (Amadeus, Winterthur 2010)

31.467 F.J. Ratte: Die Temperatur als Mittel der musikalischen Rhetorik am Beispiel des Orgelbüchleins von Johann Sebastian Bach. In: *Die Sprache der Musik. Festschrift für Klaus Wolfgang Niemöller*,

ed. by J. Fricke (Bosse, Regensburg 1989) pp. 401–411
31.468 M. Jira: *Musikalische Temperaturen und musikalischer Satz bei J.S. Bach* (Schneider, Tutzing 2000)
31.469 S. Martínez Ruiz: *Temperament in Bach's Well-Tempered Clavier. A historical survey and a new evaluation according to dissonance theory*, Ph.D. Thesis (Univ. Autònoma Barcelona, Barcelona 2011)
31.470 J. Norrback: *A passable and good temperament. A new methodology for studying tuning and temperament in organ music* (Göteborg Univ. Inst. Musicol., Göteborg 2002)
31.471 F. Kuttner: Prince Chu Tsai-yu's life and works. A re-evaluation of his contribution to equal temperament theory, Ethnomusicology **19**, 163–206 (1975)
31.472 K. Robinson: *A Critical Study of Chu Tsai-yü's Contribution to the Theory of Equal Temperament in Chinese Music* (Steiner, Wiesbaden 1980)
31.473 J. Neidhardt: *Gäntzlich erschöpfte, mathematische Abtheilungen des diatonisch-chromatischen, temperirten Canonis Monochordi*, 2nd edn. (Eckart, Königsberg, Leipzig 1734)
31.474 J. Mattheson: *Grosse General-Bass-Schule*, 2nd edn. (Kißner, Hamburg 1731)
31.475 H. Glareanus: *Dodekachordon* (Petri, Basle 1547), http://imslp.org/wiki/Dodecachordon_(Glareanus%2C_Henricus)
31.476 J. Mattheson: *Der vollkommene Capellmeister* (Herold, Hamburg 1739)
31.477 D. Heinichen: *Neu erfundene und gründliche Anweisung, wie ein Music-liebender auff gewisse vortheilhaftige Arth könne zu vollkommener Erlernung des General-Basses ... gelangen* (Schiller, Hamburg 1711)
31.478 D. Heinichen: *Der General-Bass in der Composition* (Dresden 1728)
31.479 M. Liang: *Music of the Billion. An Introduction to Chinese Musical Culture* (Heinrichshofen, New York 1985)
31.480 M. Drobisch: Nachträge zur Theorie der musikalischen Tonverhältnisse, Abhandl. der Math. Phys. Klasse der Königl. Sächs. Ges. der Wiss. **5**, 1–40 (1857)
31.481 M. Drobisch: Über musikalische Tonbestimmung und Temperatur, Abhandl. der Math. Phys. Klasse der Sächs. Ges. der Wiss. **2**, 1–120 (1855)
31.482 D. Hall: *Musical Acoustics. An Introduction* (Wadsworth, Belmont 1980)
31.483 E. von Hornbostel: Melodie und Skala, Jahrbuch der Musikbibliothek Peters **19**, 11–23 (1913)
31.484 M. Lindley: A systematic approach to chromaticism, Syst. Musicol. **2**, 155–194 (1994)
31.485 F. Brentano: *Psychologie vom empirischen Standpunkt*, Vol. 1 (F. Meiner, Leipzig 1924), ed. by O. Kraus
31.486 H. Riemann: *Musikalische Syntaxis. Grundriß einer harmonischen Satzbildungslehre* (Leipzig, Breitkopf & Haertel 1877), repr. Sändig, Wiesbaden 1971
31.487 A. von Oettingen: *Das Harmoniesystem in dualer Entwicklung* (Gläser, Dorpat, Leipzig 1863)
31.488 A. von Oettingen: Das duale System der Harmonie, Annalen der Naturphilos. **2**, 62–75, 375–403 (1903); **3**, 241–269 (1904); **4**, 116–136, 301–338 (1905); **5**, 449–503 (1906)
31.489 H. Riemann: Ideen zu einer ‚Lehre von den Tonvorstellungen', Jb. Peters **21/22**, 1–26 (1914/1915); **23**, 1–26 (1916)
31.490 C. Stumpf: Konsonanz und Konkordanz, Z. Psychol. **58**, 321–355 (1911)
31.491 M. Cordes: *Nicola Vicentinos Enharmonik. Musik mit 31 Tönen* (Akad. Druck- und Verlagsanstalt, Graz 2007)
31.492 E. Prame: Vibrato extent and intonation in professional western lyric singing, J. Acoust. Soc. Am. **102**, 616–621 (1997)
31.493 J. Brown, K. Vaughn: Pitch center of stringed instrument vibrato tones, J. Acoust. Soc. Am. **100**, 1728–1735 (1996)
31.494 A. Schneider: Klanganalyse als Methodik der Popularmusikforschung, Hamburger Jahrb. Musikwiss. **19**, 107–129 (2002)
31.495 A. Schneider: Sound, pitch, and scale: From 'tone measurements' to sonological analysis in ethnomusicology, Ethnomusicology **45**, 489–519 (2001)
31.496 A. Sutton, E. Suanda, S. Williams: Java. In: *The Garland Encyclopedia of World Music*, Vol. 4, ed. by T. Miller, S. Williams (Garland, New York 1998) pp. 630–728
31.497 J. Herbart: Psychologische Bemerkungen zur Tonlehre, Königsberger Archiv **1**, 158–192 (1812), repr. in J. Herbart: Sämtliche Werke in chronol. Reihenfolge, Bd. 3, Langensalza 1888
31.498 J. Herbart: *Psychologie als Wissenschaft, neu gegründet auf Erfahrung, Metaphysik und Mathematik*, Vol. I (Hartknoch, Königsberg 1824)
31.499 W. Wundt: *Grundriß der Psychologie* (Engelmann, Leipzig 1896), 15th edn. 1928
31.500 E. Mach: *Die Analyse der Empfindungen*, 4th edn. (Fischer, Jena 1903)
31.501 F. Opelt: *Ueber die Natur der Musik* (Herrmann Langbein, Leipzig 1834)
31.502 J. Yasser: *A Theory of Evolving Tonality* (American Library of Musicology, New York 1932), repr. Da Capo, New York 1975
31.503 M. Drobisch: Ueber die mathematische Bestimmung der musikalischen Intervalle, Abhandl. bei Begründung der Königl. Sächs. Ges. der Wiss. **3**, 89–128 (1846)
31.504 C. Ruckmick: A new classification of tonal qualities, Psych. Rev. **36**, 172–180 (1929)
31.505 A. Wellek: Der Raum in der Musik, Archiv ges. Psychol. **91**, 395–443 (1934)
31.506 A. Wellek: Die Aufspaltung der „Tonhöhe" in der Hornbostelschen Gehörpsychologie und die Konsonanztheorien von Hornbostel und Krueger, Z. Musikwiss. **16**, 537–553 (1934)
31.507 R. Shepard: Structural representations of musical pitch. In: *The Psychology of Music*, ed. by D. Deutsch (Academic, Orlando 1982) pp. 343–390
31.508 R. Shepard: Geometrical approximations to the structure of musical pitch, Psych. Rev. **89**, 305–333 (1982)

31.509 C. Krumhansl, E. Kessler: Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys, Psychol. Rev. **89**, 334–368 (1982)

31.510 S. Gathercole, A. Baddeley: *Working Memory and Language* (Erlbaum, Hove, Hillsdale 1993)

31.511 M. Leman: An auditory model of the role of short-term memory in probe-tone ratings, Music Percept. **17**, 481–509 (2000)

31.512 C. Neuhaus: Auditory Gestalt perception and the dissociation between pitch and time: ERP studies on processing musical sequence structure. In: *Systematic and Comparative Musicology: Concepts, Methods, Findings*, ed. by A. Schneider (P. Lang, Frankfurt 2008) pp. 171–195

31.513 C. Neuhaus: The perception of melodies: Some thoughts on listening style, relational thinking, and musical structure. In: *Sound–Perception–Performance*, ed. by R. Bader (Springer, Cham 2013) pp. 195–215

31.514 J.S. van Waesberghe: *A Textbook of Melody. A Course in Functional Melodic Analysis* (American Institute of Musicology, Rome 1955)

31.515 R. Gjerdingen: Categorization of musical patterns by self-organizing neuron-like networks, Music Percept. **7**, 339–370 (1990)

31.516 J. Bharucha, P. Todd: *Music and Connectionism* (MIT Press, Cambridge 1991)

31.517 J. Bharucha: Tonality and expectation. In: *Musical Perceptions*, ed. by R. Aiello, J. Sloboda (Oxford Univ. Press, New York 1994) pp. 213–239

31.518 M. Leman, F. Carreras: Schema and Gestalt: Testing the hypothesis of psychoneural isomorphism by computer simulation. In: *Music, Gestalt, and Computing. Studies in Cognitive and Systematic Musicology*, ed. by M. Leman (Springer, Berlin 1997) pp. 144–168

31.519 B. Tillmann, J. Bharucha, E. Bigand: Implicit Learning of tonality: a self-organizing approach, Psych. Rev. **107**, 885–913 (2000)

31.520 P. Toiviainen: *Modelling Musical Cognition with Artificial Neural Networks* (Univ. Jyväskylä, Jyväskylä 1996)

31.521 H. Riemann: *Geschichte der Musiktheorie*, 2nd edn. (Hesse, Berlin 1920)

31.522 H. Riemann: *Vereinfachte Harmonielehre oder Die Lehre von den tonalen Funktionen der Akkorde* (Augener, London 1893)

31.523 R. Imig: *Systeme der Funktionsbezeichnung in den Harmonielehren seit Hugo Riemann* (Gesellschaft zur Förderung der systematischen Musikwissenschaft, Düsseldorf 1970)

31.524 C. Dahlhaus: *Untersuchungen zur Entstehung der harmonischen Tonalität* (Bärenreiter, Kassel 1968)

31.525 H. Powers: From psalmody to tonality. In: *Tonal Structures in Early Music*, ed. by C. Collins Judd (Garland, New York 1998) pp. 275–340

31.526 C. Collins Judd: Josquin's gospel motets and chant-based tonality. In: *Tonal Structures in Early Music*, ed. by C. Collins Judd (Garland, New York 1998) pp. 109–153

31.527 S. Tanaka: Studien im Gebiete der reinen Stimmung, Vierteljahrsschrift Musikwiss. **6**, 1–90 (1890)

31.528 L. Euler: De harmoniae veris principiis per speculum musicum repraesentatis, Novi Commentarii Acad. Scient. Petropolitanae **18**, 330–353 (1774), repr. in Leonhardi Euleri Opera omnia, Series III, T. 1, Teubner, Leipzig, Berlin 1926

31.529 M. Vogel: Arthur von Oettingen und der harmonische Dualismus. In: *Beiträge zur Musiktheorie des 19. Jahrhunderts*, ed. by M. Vogel (Bosse, Regensburg 1966) pp. 103–132

31.530 A. Fokker: Unison vectors and periodicity blocks in the three-dimensional (3-5-7) harmonic lattice of notes, Proc. Kon. Nederl. Akad. Wetensch. Series B 72, no. 3 (1969)

31.531 S. Karg-Elert: *Polaristische Klang- und Tonalitätslehre* (Leuckart, Leipzig 1931), repr. in: S. Karg-Elert: *Theoretische Werke* (Ewers, Paderborn 2004)

31.532 E. Groven: *Equal Temperament and Pure Tuning* (E. Groven, Oslo 1969)

31.533 I. Loe Dalaker, A. Jorunn Kydland, D. Łopatowska-Romsvik (Eds.): *'East of Noise'. Eivind Groven. Composer, Ethnomusicologist, Researcher* (Akademiska, Oslo, Trondheim 2013)

32. Perception of *Timbre* and *Sound Color*

Albrecht Schneider

This chapter deals with perception of *timbre* or *sound color*. Both concepts can be distinguished in regard to terminology as well as to their historical and factual background even though both relate to some common features for which an objective (acoustic) basis exists. Sections of this chapter review in brief developments in (traditional and electronic) musical instruments as well as in research on timbre and sound color. A subchapter on sensation and perception of timbre offers a retrospective on classical concepts of tone color or sound color and reviews some modern approaches from Schaeffer's *objet sonore* to semantic differentials and multidimensional scaling. Taking a functional approach, acoustical features (such as transients and modulation) and perceptual attributes of timbre as well as interrelations between *pitch* and *timbre* are discussed. In a final section, fundamentals of sound segregation and auditory streaming are outlined. For most of the phenomena covered in this chapter, examples are provided including sound analyses obtained with signal processing methods.

32.1 Timbre and Sound Color: Basic Features

32.1.1 Terminology: *Timbre* and *Sound Color*

Timbre is a French word that denotes a stamp (e.g., *timbre fiscal* = revenue stamp), a brand, a sound or a sound color. The French Encyclopédie has an entry *timbre* (T. 16, 1765, 333) that contains several musically relevant annotations, namely *timbre* as referring to snare strings on a drum skin, timbre as the resonant state of a bell, timbre of a voice or of a musical instrument. In his article on *son* (tone, sound; Encyclopédie, T. 15, 1765, 345; for historical aspects see [32.1, 2]), Rousseau had used the term timbre as covering a third property of sounds besides tone height (*le degrè d'élevation entre le grave & l'aigu*) and intensity expressed as the degree *de véhémence entre le fort & le foible*. Timbre then is the quality of a sound that results from an evaluation in regard to dullness – shrillness or softness – brightness (*du sourd à l'éclatant, ou de l'aigu de doux*; the scaling of timbre thus is from sourd and doux to éclantant and aigu). The term timbre as it occurs in textbooks on orchestration [32.3, 4] denotes an integral quality of sound as produced by, and attributed to, certain instruments. These were classified, first by *Mahillon* [32.5–8] and then by *Hornbostel* and *Sachs* [32.9], according to physical principles of sound production, in the first place. The classification offered by Mahillon, based on the huge collection of instruments housed in the Brussels Conservatory of music, comprised:

1. Autophones
2. Instruments à membranes
3. Instruments à vent
4. Instruments à cordes.

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_32

Hornbostel and Sachs have the same four classes (labeled idiophone, membranophone, chordophone, aerophone), however, in a different order (and with significant differences of grouping within each class) that reflects the physical principles involved more clearly. In general, idiophones viewed as vibrating bodies consist of three-dimensional structures (rods, bars, plates, shells), membranes such as drum skins are – ideally – two-dimensional (neglecting their thickness), and thin strings have been treated as one-dimensional in the seminal work of Bernoulli and Euler so that longitudinal, transversal and torsional vibration can be covered by second-order differential equations [32.10, 11]. Finally, aerophones such as flutes, for the lack of shearing forces between molecules in air columns, only allow longitudinal vibration.

Referring to the principles and materials of sound production, *Gevaert* [32.3, p. 5] for example states that the timbre of wind instruments (instruments à vent) is determined by the geometry (length, diameter, bore profile) of the tube, on the one hand, and by the mechanism by which the air inside the tube is set to vibration, on the other. Gevaert correctly points to a regular sequence of pulses (*battements*) necessary for making the air column vibrate, and he also mentions three basic types of pulse generator (edge tone, valves formed by either beating reeds or the lips pressed into a mouthpiece) as are used for sound production in flutes, reeds, and brass instruments respectively. Further, he attributes the *timbre moelleux* (soft timbre) of the French horn to the conical shape of the mouthpiece. *Gevaert* [32.3, p. 18] found that bowed strings are *the soul of instrumental music* since they have a *timbre pénétrant et riche*. *Berlioz* [32.4, p. 21] judged that the *sons harmoniques* of the lowest (fourth) string of a violin *ont quelque chose du timbre du Flûte; ils sont préférable pour chanter une mélodie lente*. The point of interest here is that timbre is regarded as a unique and integral quality (e.g., *timbre du Flûte*), though this may be limited to a certain register or to tones played on a certain string. In this respect, *Forsyth* [32.12, p. 480], in his textbook on orchestration, argues the D (second) string of the cello, *of all the soft, silky sounds in the orchestra, it is the softest and silkiest*.

The unique and consistent timbre attributed to certain instruments or to their parts such as individual strings or pipe ranks (as in organs) is expressed, in the English language, by terms like tone quality, tone color or sound color. Likewise, these terms are used in German (*Klangfarbe*, *Tonfarbe*). Timbre is often used synonymously for *sound color* though there seem to be some differences between the phenomena covered by these terms in that *sound color* predominantly refers to a certain spectral structure while timbre, at least in more recent research, can cover spectral as well as temporal aspects (Sect. 32.1.2). Fundamental to both terms is the experience that sounds can have a distinct sensory quality that, though perhaps not independent of other qualities, in particular pitch and loudness, cannot be accounted for as a *function* of either pitch or loudness alone or as a simple combination of both. Consequently, there seems to be information encoded in the sound structure that gives rise to sensations of *timbre* in addition to information pertaining to *pitch* and to *loudness*.

32.1.2 Objective Basis of *Timbre* and *Sound Color*

In line with the realist and causal perspective taken in Chaps. 30 and 31, timbre and sound color are regarded as sensory experiences deeply rooted in natural foundations. These are physical, on the one hand, and anatomical as well as physiological, on the other. In an evolutionary perspective, various animal species (from insects and amphibians to birds and mammals) show remarkable diversity in regard to organs suited to perform sound production as well as sound perception (articles in [32.13–15]). Many vocalizations serve to communicate information (*Tembrock* [32.16] and articles in *Witzany* [32.17]). The degree of structural and functional complexity reached in birdsong [32.18, 19] and in whale songs [32.20] is particularly striking in regard to the large and diverse song repertoire of certain species involving learning and memory as cognitive capacities as well as communication networks operated by two or more members of a certain species. In addition, interspecies sound communication is known from bioacoustical observations. The songs of birds and whales comprise complex sound patterns (take, for example, songs of the nightingale and of the humpback whale), which vary considerably in spectral content over time. In this respect, it can be said that these sounds make use of timbral qualities as a means of communication. There are parallels in human speech and singing in that different phonation (resulting in different spectral energy distribution and formant structure) can convey different emotional states as well as the intentions of a person speaking or singing [32.21].

Sound production and sound perception in animals (including humans) is based on acoustical principles [32.21–23] even though these are implemented into organic biosystems. In fact, one can give a description of the vocal tract of humans in terms of anatomical structure and functional aspects of muscles, nerves, etc. in the phonation process. Further, one may go into the acoustics of phonation in terms of generator and resonator geometry, air flow and pulse sequence

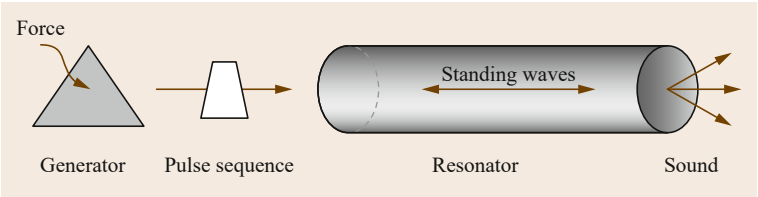


Fig. 32.1 Model of an aerophone (pulse generator coupled to a tube-like resonator)

generation at the vocal folds as well as resonance phenomena taking place in the mouth cavity [32.21, 24] etc. Quite obviously, there are also parallels between the structure and function of the human singing voice and a host of musical instruments classed as aerophones. *Gevaert* [32.3] rightly attributed the timbre of wind instruments to a combination of a valve-like pulse generator and a resonator in which an air column is set to vibration. Taking the general model of such a generator coupled to a resonator, the scheme in Fig. 32.1 can be drawn.

In this basic model, a source (a person breathing air) drives a generator by supplying a force, which in this case is air flowing at a certain speed and with a certain pressure. The flow is periodically interrupted either by a small jet of air being bend inward and outward of a sharp edge (e.g., the labium of an organ flue pipe), or by valves that are (partially or completely) opened and shut to produce pulses of air released into the resonator. A single reed (as in the clarinet) or double reeds (as in the oboe) or the lips pressed into a mouthpiece (as in the trumpet, trombone, French horn, etc.) can serve as the valve in the generator. The sequence of pulses traveling through the length of the resonator (a cylindrical or conical tube) partially is reflected at the end opposite the generator so that standing waves are formed inside the tube where natural modes of vibration are excited. Without going into details, which are intricate because the generator parameters include quite many nonlinearities [32.11, 25], one can see that the behavior of such a generator-plus-resonator system depends on the geometry of its parts, on the input parameters (input impedance of the valve, blowing pressure and speed, air flow through the valve, input impedance of the tube, etc.) and on the coupling of the generator to the resonator. The interaction between the two maintains the regeneration cycle needed for continuant tones. In regard to the geometry of resonators, it has been demonstrated [32.26] that a resonator treated as a Bessel horn produces a series of natural mode frequencies so that higher mode frequencies are multiples of the lowest only if the Bessel function J^x yields either $x = 0$ (cylinder) or $x = 2$ (conical tube). The clarinet has a cylindrical tube almost closed at one end by the valve so that the resonator predominantly responds to odd harmonics; the clarinet overblows into the twelfth, which is the fifth above the octave (the third harmonic [32.27, p. 115–125]), while the oboe (where the bore is close to a cone slightly truncated near the reed generator) overblows into the octave. For the tones in between, finger holes are provided on both instruments. Since also the number of modes excited in the resonator differs between woodwinds (for the same pitch played with similar force of excitation applied), different spectral energy distributions result that are perceived as differences in *sound color* or timbre. For instance, sounds recorded from a number of woodwinds playing the same note (C_4) with moderate force of excitation have spectra that differ in the number and strength of partials (Fig. 32.2).

The sounds analyzed with a phase vocoder algorithm (equivalent to putting the sounds through a bank of band pass filters tuned to f_1 of the sounds [32.28]) are samples of the following instruments (left to right): bassoon 1, bassoon 2, bass clarinet, clarinet, oboe, concert flute. The spectral centroid (Sect. 30.2) for these sounds varies from ≈ 680 Hz (bassoon 1) and 880 Hz (bassoon 2) to 2.4 kHz (bass clarinet) and 2.55 kHz (clarinet); the centroid for the oboe sound is 1.42 kHz while, for the flute, the centroid is identical with f_1 (261.6 Hz). Thus, for a performer or listener the sounds in question differ significantly in brightness when playing the same note. What distinguishes tones produced by different instruments such as the bassoon, the clarinet, the oboe, etc. in the steady-state portion of sound after the transient part is the shape of the spectral envelope and, thereby, the distribution of spectral energy covered by the envelope. The spectra of wind instruments controlled by a valve as pulse generator have been said to approach a cyclic structure with maxima that can be interpreted as formants [32.29–31]. A condition necessary for a perfect cyclic spectrum would be a series of rectangular pulses with a duty cycle of τ/T (τ = pulse width and T = pulse period) that yields small integer ratios; amplitudes of partials then conform to a sinc function $(\sin x)/x$ where zeros are at $n\tau/T = 1, 2, 3, \ldots, k$ [32.32, p. 40 f.] (Fig. 32.3).

Spectra approaching a cyclic structure to some degree can be recorded from reed instruments such as the oboe; in Fig. 32.4, the spectrum of the tone C_4 played (*mf*) on a baroque oboe is shown. In addition, the spectral envelope calculated from a formant filter analysis

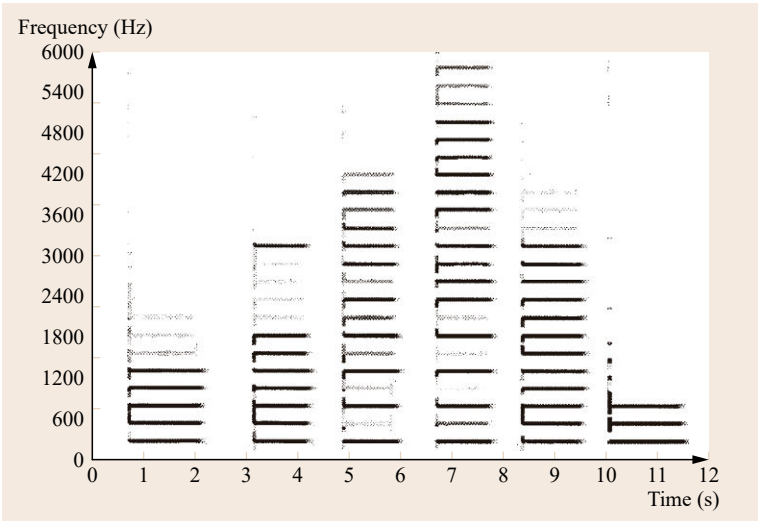


Fig. 32.2 Spectra of various woodwinds (from left to right: bassoon 1, bassoon 2, bass clarinet, clarinet, oboe, flute) all playing the note C_4 (phase vocoder analysis, base frequency = 261.6 Hz). Relative amplitude of partials indicated by grayscale (*white* = low, *black* = high amplitude)

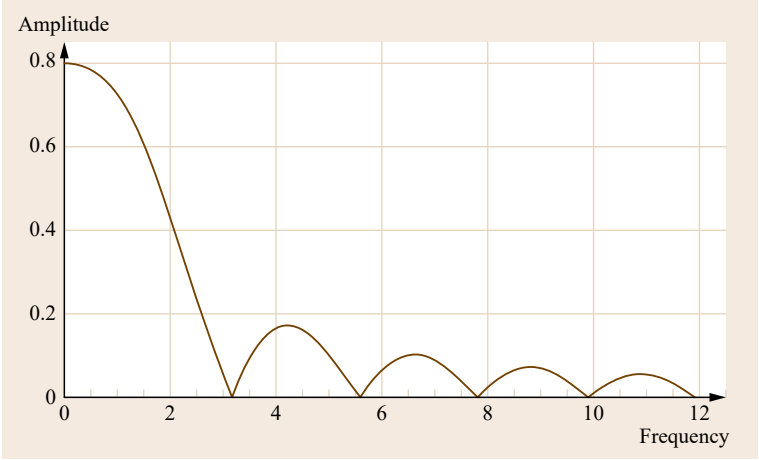


Fig. 32.3 Sinc function as a model for the envelope of a cyclic spectrum

is plotted in the same graph. One can see three groups of partials at $\approx 0-1.8$ kHz, 1.8–3.6 kHz, 3.6–6.6 kHz separated by relative minima; these groups are covered by peaks of the formant filter envelope. Also, a formant analysis performed with the Burg algorithm [32.33] yields several tracks of formant frequencies as a function of time. In this respect, one may assign a formant-like quality to this sound (as well as to sounds from other reed instruments [32.34]). Even closer approximations to a cyclic spectrum can be observed in organ reed pipes [32.35] and also in sounds recorded from plucked strings in a harpsichord where in particular the string velocity reflects the pulse train and the spectrum of the string velocity consequently is fairly periodic [32.36]. Hence, the pulse generator imposes the shape of a spectrum on the resonator, which is fixed in geometry, in organ reed pipes and in the harpsichord, while in reed instruments such as the oboe the length of the air column vibrating in the bore can be modified by means of finger holes.

The generator-plus-resonator model can also be viewed as a generator producing a source signal fed into a filter, that is, into the resonator (for source-filter processing see [32.37, 38]). In terms of linear systems [32.39, Chap. 5] and [32.40, Chap. 9], a filter of low-pass or bandpass characteristic has a certain frequency response as well as an impulse response; the transfer function $H(\omega)$ of a filter determines the amplitudes and phases of the frequency components in the output spectrum relative to the input spectrum. Relating bandwidth to response time, the filter response time τ of a symmetric bandpass to an input signal with rapid onset is

$$\tau = \frac{2\pi}{\Delta\omega} = \frac{1}{\Delta f} .$$

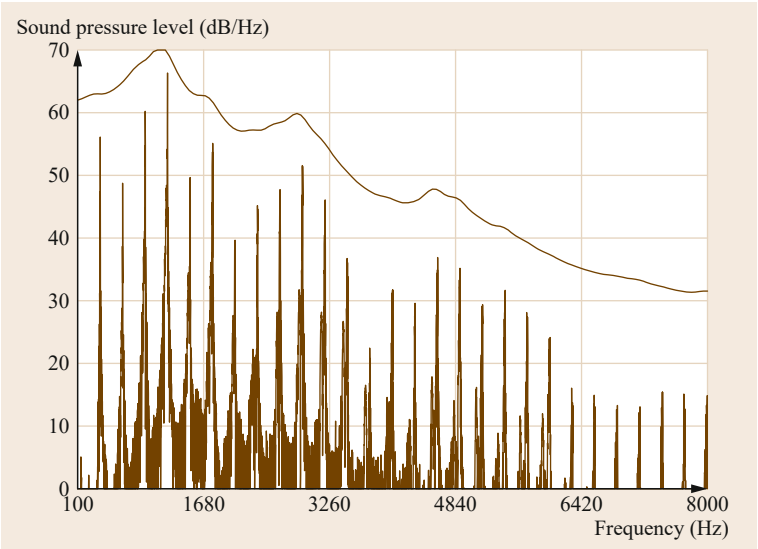


Fig. 32.4 Baroque oboe, spectrum of tone C_4 (*mf*) and formant filter envelope

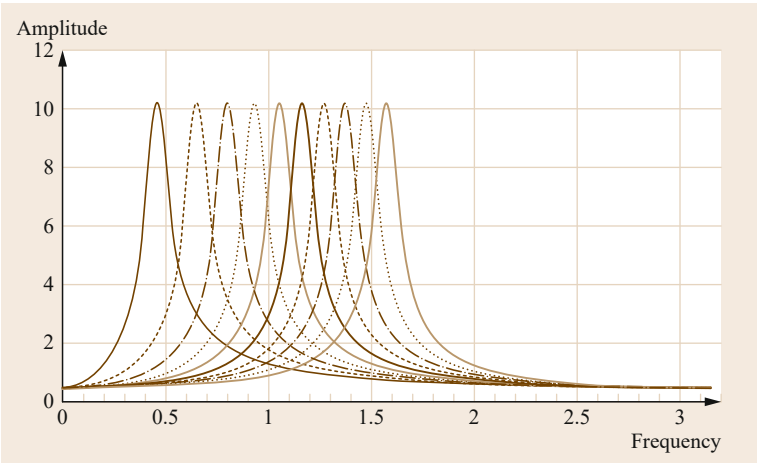


Fig. 32.5 Bank of bandpass filters as analogue of resonance peaks in instruments

Hence the response time is the inverse of the bandwidth. The transfer function of the resonator can be modeled by taking the resonances of the tube as the center frequencies of a chain of bandpass filters (Fig. 32.5).

If we conceive of a digital filter where $x(k)$ is the input signal and $y(k)$ is the output signal, convolving the input signal with the impulse response of the filter, $h(k)$, yields the output sequence $y(k) = x(k) \times h(k)$. The poles of the filter in the complex z-plane are equivalent to the resonance maxima in the resonator (e.g., the tube of an organ flue pipe). It is in physical modeling of musical instruments that such concepts have been implemented in signal processing codes.

In regard to the temporal behavior of the system, it takes some time before a pulse sequence fed into a tube resonator builds up standing waves, which is the condition for sound production and radiation. This process is particularly evident in large organ flue pipes (of 16′ and 32′ size), but can be observed also in smaller pipes (of 8′, see Figs. 30.17, 30.18 and [32.35]), in large duct flutes like the Slovak *Fujara* (of ≈ 160 cm tube length [32.41]), and even in large reed pipes where higher modes build up at their correct harmonic frequencies only after several hundred ms. In Fig. 32.6, the onset of the tone B_2 ($f_1 \sim 122.3$ Hz) recorded from a bassoon played with medium force (*mf*) is shown. One can see that mode no. 2 (corresponding to partial f_2 in the spectrum) is building up early while mode no. 1 needs more time and mode no. 3 as well as higher modes start with broad spectral lobes meaning a stable regime of vibration has not yet been established for these modes. Also, there is some noise in the transient signal in a frequency band above ≈ 1 kHz. The steady state of the tone as defined by clear spectral peaks

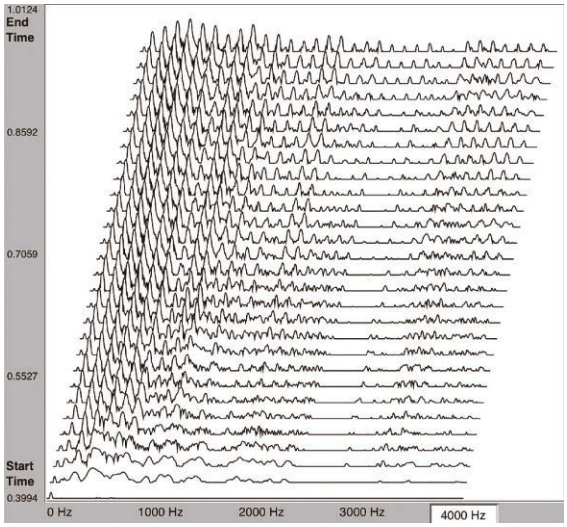


Fig. 32.6 Bassoon, onset of sound for a tone at $f_1 \sim$ 122.3 Hz; 29 spectra (FFT: 4096 pts, Hanning, hop ratio 0.2)

marking harmonics as well as by a small degree of spectral fluctuation is reached only after $\approx$ 300–400 ms. However, a clear pitch conveyed by a number of low harmonic partials as well as the period of the temporal envelope can be perceived much earlier (within $\approx$ 100 ms from onset).

The sounds from organ pipes analyzed in Sect. 30.2 and the bassoon sound (Fig. 32.6) demonstrate that, in particular in aerophones but also in plucked and bowed chordophones, a transient part precedes the steady state; the transient part results because the vibrating system (air column, string) has a certain input impedance and exhibits inertia. Since the input impedance for a cylindrical tube filled with air is

$$Z = \frac{p}{q} = \frac{p}{v\,A}\,,$$

where p is the pressure and $q = v\,A$ is the acoustical volume current (for a plane A and particle velocity v), pressure has to build up beyond a certain threshold before a stable vibration pattern with standing waves and radiation of a periodic sound signal is achieved. Consequently, the transient part contains noise in the signal in addition to the restricted number of harmonic partials resulting from such modes that respond to the excitation right from the onset. In plucked strings, as in the harpsichord, the transient comprises both a longitudinal wave preceding transversal motion as well as a short noisy segment resulting from the interaction of the plectrum with the string [32.36]. In strings excited with a small hammer, one can observe a short noisy precursor signal followed first by the longitudinal wave and then by transversal motion of the string (for measurements of a Stein-Conrad Hammerflügel from 1793 that has delicate small hammers, thin strings and a fast action see [32.42]). The precursor signals are quite short (usually, $t < 10$ ms) yet are of perceptual relevance (Sect. 32.2.3).

32.1.3 Organology, Electronics, and Timbre: Some Historical Facts

Humans evidently recognize and value different sound qualities given the immense number and diversity of musical instruments in use in various cultures and ethnic groups around the world. Exploration of the field known today as *sound color* and *timbre* was begun, on an empirical level, by musicians and craftsmen when making flute, reed, brass and string instruments. Historically, highly developed instruments such as the bronze lurs from Scandinavia [32.43] and bronze bells from China [32.44] testify instrument makers and musicians must have had a regard for sound quality already in antiquity. Timbre as related to certain materials played a significant role in Chinese tradition of instrument classification [32.45, p. 67 ff.]. Differences in sounds were also discussed in Greek writings on music theory (as in chapters of Aristoxenos' treatise on music, see [32.46]). In Roman times, several brass instruments, because of their powerful (as well as *horrible*) sound, were employed for military purposes while the first organs (the organum hydraulicum, said to have been invented, in Alexandria, by Ktsebios) appeared as instruments suited to fill a theater or arena with sound [32.47]. From the Middle Ages onward, mensuration of organ pipes perhaps led to a basic understanding of the interdependence of *pitch* and *sound color* in flue pipes. Musical instruments such as shawms, bagpipes and the hurdy-gurdy are mentioned quite frequently in medieval and Renaissance literature including remarks on sound properties. The development of bagpipes as well as the medieval bladder pipe not only provided players with a continuous sounding instrument but also with a *sonorous* sound quality (both chanter and drone(s) employed single or double reeds). Similarly, the hurdy-gurdy (which appeared in Europe around the 13th century [32.48]), offered a performance style based on continuous drone accompaniment plus distinct melodic pitch sequences as well as a specific sound (resulting from the interaction of the turning wheel with the strings, where the speed of the wheel and thereby the *attack* on the strings can be varied). Organ dispositions of the 14th century indicate that several instruments already offered a contrast of two sound concepts, namely diapason (*Prinzipal*) and organo pleno (*Grand orgue* [32.49, p. 10 ff.]). By the end of the 15th century,

the wall profile for the minor-third bell was discovered; a good example is the *Gloriosa*, Erfurt, cast by Geert (Gerhardus) de Wouw, in 1497. The minor third in this particular bell has ≈ 280 cent (the Pythagorean minor third has 294 cent), which suggests bell founders had an understanding of how to produce a certain spectrum by shaping the profile of the bell's wall. A detailed account of the instruments in use by about 1600 was provided by *Praetorius* [32.50]. The broad range of instruments (in particular reeds) developed in the Renaissance and the concept known as *Spaltklang* (split sound) realized in organ dispositions (as reported in [32.50]) as well as in ensembles demonstrates a keen sense of timbre. In modern times, *sound color* (German: *Klangfarbe*) became an essential feature in Western art music as reflected in orchestration (for a comprehensive survey, see [32.51]) growing increasingly more complex in works of the 19th and 20th century respectively. Composers demanded rare or newly invented instruments (such as the celesta employed by Tchaikovsky in *The Nutcracker* or the Heckelphone in Strauss' *Salomé*) as well as unusual ways of playing (as in many works of modern music after 1945) in order to have unique or even perplexing sounds at their disposal. It should be noted that the temporal and spectral characteristics of tones played on most orchestral instruments vary considerably in regard to dynamics (from *pp* to *ff*), to the effect that the radiation pattern and hence, the directivity of sound also changes a lot with dynamics (for a comprehensive survey of facts and data, see [32.52]). In this respect, the timbre is by no means constant (besides the variation of spectral energy distribution observed when instruments are played in different registers).

The advent of electroacoustic music brought technical sound sources (generators, oscillators) and devices to combine and modify sources (e.g., ring modulator, vocoder) into play [32.53–56]. The analogue (voltage-controlled) synthesizer and digital instruments that became available for common use in the 1970s and 1980s respectively, offered even more choices for creating complex sounds [32.56–58]. Digital synthesizers (like the Yamaha DX 7) as well as digital samplers (like the Akai S 1000 and the Emax II) in fact are music computers based on signal processing technology. The computer had been used for sound generation, by Max Mathews and other pioneers, for about two decades before digital audio became a standard in sound recording and music media (the CD was standardized as a digital format around 1979–1980 and introduced to the commercial user market in 1982). Special techniques like frequency modulation (FM) synthesis [32.59–61] allowed both to replicate the timbre of many existing instruments, among them idiophones like xylophones, gongs, etc. and to create sounds with complex, time-variant harmonic and inharmonic spectra, which were unprecedented. Such sound material led to compositions that transcend borders between the common categories of *pitch* and *timbre* (like *Stria* [32.62] by John Chowning, *Bossis* [32.63] and Sect. 32.2.4).

In addition to electronic and digital sound synthesis, all kinds of environmental and technical sounds were put to use in *musique concrète*, which, as one of its goals, considered aperiodic noises and periodic waveshapes as a continuum to be exploited for sound collage techniques [32.64, 65]. A far-reaching, more generalized concept of *sound* evolved in areas of electronic and computer music as well as in studio productions of pop and rock music when room acoustics, multichannel recording and reproduction and a host of audio effects (such as artificial delay and reverb, phasing, flanging, chorus; see articles by *Dutilleux* and *Zölzer* [32.66]) were integrated with orchestral and electronic instruments as well as with analogue and digital sound samples. In consideration of these developments, it became customary already in the 1960s to speak of, for example, *the sound of Mantovani* (multiple bowed strings deeply embedded in reverb) or *the wall of sound* produced by Phil Spector in pop music recordings, for which the recipe was to have many instruments played simultaneously in a rather small studio so that their sounds overlay and are hardly recognizable individually since Spector added, moreover, amounts of reverb and compressed the mix dynamically so that indeed a very dense *wall of sound* is audible in recordings like *Be My Baby* [32.67]. Later on, there was *psychedelic sound* in which time-axis manipulation of signals such as phasing and flanging as well as stereo panning effects figured prominently (e.g., Tomorrow: *Revolution*, 1967, Jimi Hendrix: *All Along the Watchtower*, 1968), and then *disco sound* (with huge concentration of spectral energy at the bottom end of the audible frequency range and typical patterns of percussion and bass in regard to meter and rhythm), etc. In all these *sounds*, timbre played an important if not decisive role. However, sound in this respect rather is a conglomerate of natural and artificial sound sources, effects, production and reproduction techniques while *timbre* traditionally (as in treatises on orchestration [32.12, 51]), has been assigned to single instruments or to the voice of certain male and female singers.

32.1.4 Research on *Timbre* and *Sound Color*: A Brief Retrospective

Though elements of acoustics can be traced in Greek antiquity (for example, the observation that the tension of a string determines whether the tone it pro-

duces sounds dull or sharp), a more systematic approach was pursued in the 16th and 17th century when empirical research was established along with mathematical treatment of problems. Knowledge concerning sound structure improved a lot when Beeckman and Mersenne understood the nature of harmonic vibration that led to the discovery of partials in strings (for which Sauveur gave a detailed description in the years 1700–1713 [32.68–70]). From Sauveur's published lectures, *Rameau* [32.71] saw that musical tones comprised a fundamental and its harmonics. By about 1800, it was clear to acousticians like Chladni that musical sound in general was a mixture of harmonic or inharmonic partials. The explanation *Chladni* [32.72, p. 241 ff.] gave was that elastic bodies can undergo very many vibrations at the same time, which would correspondingly activate many different parts of the inner ear without hampering each other. Thereby, sounds from different instruments could be perceived simultaneously. *Chladni* [32.72, Sec. 248] attributed different sound qualities to the different materials of elastic bodies consisting of organic or inorganic material (e.g., wood, brass, iron) and the microstructure of vibration inside such bodies as well as in the media (fluids, solids) through which sound propagates. *Opelt* [32.73, Sec. 7] held that the quality of sounds (e.g., strings, trumpet, flutes) or the *so-called sound color* (*Klangfarbe*) depends on different kinds or shapes of pulses reaching our ears. Opelt argued pulses and vibrations must be complex since, in a musical instrument, all parts vibrate; that is, strings vibrate coupled to resonance plates and air columns vibrate coupled to the tubes and bells of wind instruments. The resulting sound embedded in complex tones thereby is *an aggregate of several isochronous pulse sequences* having their origin in various parts of the instrument. Moreover, the mechanism of excitation (plucking or bowing a string, hammers in keyboards, etc.) would result in different *sound colors*.

When Helmholtz, in the 1850s, began his work on musical acoustics and psychoacoustics, he had tuning forks and resonators as well as sirens for sound analysis. Sets of precise tuning forks driven electromechanically for continuant tones provided kind of an early synthesizer (Rudolph Koenig, ten forks, $f_1 = 128\,\mathrm{Hz}$ [32.74, p. 329 ff.]) for complex sounds such as vowels. Koenig, himself an acoustician who cooperated closely with Helmholtz, also constructed a mechanical wave analyzer with a set of resonators. In the 19th century, some elementary tools for recording sound waves had become available [32.75] before Edison developed his improved model of the phonograph (issued in 1888) that was used in many investigations of sound. Among the objects of study was sound production in the human voice, and in particular the nature of vowels [32.74, p. 367 ff.]. In the second half of the 19th century, the quest for finding a specific resonance mechanism that could explain the production of vowels led to theories of *formants* (a term coined by the physiologist Ludimar Herrmann in the 1880s who contributed greatly to empirical sound research [32.34]). Sound research gathered further momentum when, after about 1920, continuous recording of sound on film labeled *phonophotography* [32.76, p. 10 ff.], [32.77] and analysis of the sound wave both in regard to periodicity and spectrum became widespread as lab techniques. Pitch had been calculated as fundamental frequencies from the periods of vibration (by $f = 1/T$) even before and spectral analysis had been done with the aid of tuned resonators, notably by Helmholtz. However, spectral analysis by means of filters [32.78] not only allowed identification of spectral components but also investigation of transient behavior. Using octave sieves where the impulse response is short due to the rather broad filter bandwidth one could see modes of vibration and corresponding spectral components building up over time before a stable (quasistationary) regime was reached [32.79, 80]. Still, *sound color* rather referred to the quasistationary regime of vibration and its corresponding spectrum showing the amplitudes of spectral components at their (harmonic or inharmonic) frequencies while transients were addressed as the short section at the onset where the sound wave often lacks clear periodicity (Sect. 32.1.2; for a survey of research on transients in nonpercussive instruments, see [32.31]).

Significant progress in musical sound research was made when an improved model of the analogue Sona-Graph became available for musicologists in the 1960s. The Sona-Graph, first issued in the late 1940s for research in phonetics (*visible speech*), allowed a spectral and temporal representation of sound in a quasi-three-dimensional (3-D) format (time as abscissa, frequency as ordinate, and energy of partials indicated by degrees of a grayscale). The Sonagraph Model II offered several bandpass filter settings and was used to explore sound structures in Western and non-Western musics [32.81] including characteristics of the singing voice [32.82]. In the 1980s, digital spectrum analyzers (such as the B&K 2032 model) allowed FFT-based spectral analysis of complex sounds [32.83]. Since the 1970s, a number of special codes for sound signal analysis had been developed including linear prediction, autoregressive models, and pitch tracking [32.33, 84]. From about 1990 on, powerful workstations suited for digital signal processing (DSP) became available. Software packages (like sndan, based on phase vocoder analysis/synthesis, introduced in 1993 [32.28]) allowed users to perform high-resolution sound analysis and synthesis/resynthesis. With the new tools (both hardware and

software) it was possible to study temporal and spectral structures of sounds recorded from Western and non-Western (e.g., Indonesian *gamelan*) instruments in great detail [32.85, 86]. At the same time, approaches to sound synthesis and resynthesis were refined further when wavelets, granular synthesis, digital waveguides and other techniques had been developed (there are collections of many relevant articles in [32.28, 66, 87, 88].

Summing up this paragraph, it seems obvious that exploration of musical timbre depended significantly on the tools available to researchers for sound analysis and synthesis at a certain time and place. There is a line of progress leading from mechanical to electrical devices and finally to computer-based algorithmic modeling of sound and sound-producing instruments and to ever more fine-grained analysis. Not only did acoustical and musical sound research benefit from computerized tools. As will be seen in the following section, also psychological research into timbre and sound color took new directions when computers and software for advanced statistics became available for many institutions. Meanwhile, toolboxes for sound research have been set up including audio descriptors applicable to musical timbre [32.89].

32.2 Sensation and Perception of *Timbre* and *Sound Color*

32.2.1 Classical Concepts of Tone Color or Sound Color

Helmholtz [32.90] ascertained that the pitch of a simple or complex harmonic tone depends on the period of vibration and that sound intensity depends on its amplitude. He attributed sound color to the microstructure within each period of vibration as well as to the fine structure of the resulting sound wave. He stated that each different sound shape calls for a distinct shape of vibration whereas several different waveshapes might bring about the same sound color. Different sound colors according to Helmholtz result from different patterns of harmonic partials added to the fundamental frequency (f_1) determining the pitch of the sound. Hence, differences in the sound color of complex sounds result from the number and amplitudes of partials above f_1 while phase differences between partials, *Helmholtz* [32.90, p. 194] declared, can be neglected. This view is in line with his resonance theory of hearing, which posits the inner ear would perform a Fourier analysis whereby a number of partials would become audible as constituents of a harmonic complex. From his observations *Helmholtz* [32.90, p. 97] argued that strong partials suited to activating a resonator would also be audible as an individual harmonic (*Oberton*), and that no *Oberton* was audible in the case where no response from a resonator had been observed. Though Helmholtz considered also sounds with inharmonic partials, his main concern was the harmonic type since most of the instruments assembled in an orchestra are chordophones and aerophones.

Stumpf [32.91, p. 520 ff.] stated that, from a phenomenological perspective, human subjects assign three basic attributes to sounds, namely pitch (*Höhe*), intensity (*Stärke*), and extension (*Größe*). While sensations of pitch and intensity can be directly traced to physical properties of sound (frequency and period as well as the amplitude of the sound wave reaching the ear), extension as implying some spatial interpretation is a more complex concept that can incorporate several sound features and sensory attributes. In his book on speech sounds, *Stumpf* [32.92, Chap. 15] also elaborates on his concept of instrumental sounds in detail. He distinguished *inner* and *outer* moments of sound color, where the latter depend mostly on temporal factors (transients, envelope, modulation) while the inner structure of sounds depends mostly on spectral composition and energy distribution [32.34]. Both taken together perhaps embrace what the term *timbre* seems to denote: an intricate combination and interplay of temporal, spectral, and dynamic features of sounds that, in normal sensation and even in an analytical mode of hearing, are often difficult to analyze and hard to separate from each other. *Stumpf* [32.91] argued that, in actual sensation and perception, even basic attributes are not completely independent since the pitch of pure tones can vary to some extent with intensity, and both combined give rise to differences not only in *tone height* but also in brightness, density, and volume. *Köhler* [32.93–96] and some other researchers pointed to the similarity between pure tones played from low to high frequencies and the sequence of vowels *u-o-a-e-i*, resulting in an attribute often labeled *vowel quality* (also termed *vocality*).

The phenomenal description of tonal attributes and their interrelations had been addressed, from an empirical descriptive approach including auditory tests with musically trained and untrained subjects, by *Stumpf* [32.91, 92, 97–99] and by several of his coworkers [32.93–96, 100] as well as by other researchers. For example, *Rich* [32.101] had proposed the attribute of *volume* to mark the spatial extension or diffusion of low tones against high ones imagined by listeners in addi-

tion to other such attributes like brightness or density. *Stevens* [32.102] found that even pure tones evoke sensations of *volume* (*bigness*, *spread*, [32.103]), and that this attribute relates to both the frequency and the intensity of tones. Since pitch to a small extent [32.104] and loudness attributed to pure tones also depend on both frequency and intensity, *Stevens* [32.102] warned *that no psychological dimension need be a simple correlate of a single dimension of the stimulus*. In fact, several attributes covary positively or negatively with frequency and intensity; an increase in frequency from low to high brings about sensations that covary positively with frequency, namely pitch (if taken as *tone height*), brightness, and density. Brightness increases also when intensity is raised within certain limits, and the same holds true for density as a phenomenal attribute of sound. Further, increases in intensity for pure and complex tones in the treble frequency range ($\approx 6-10$ kHz) can turn sensation of brightness into unpleasant sharpness. While tone height, density and brightness increase with frequency, the volume of tones seems to be larger at low and smaller at high frequencies (provided constant sound pressure level (SPL)). *Lichte* [32.105] confirmed that, besides pitch and loudness, complex tones have *at least three attributes. These are brightness, roughness, and one tentatively labeled fullness* (which corresponds to *volume* in other studies).

Stimulus parameters and attributes of sensation that had been described in a range of studies [32.92, 100, 103, 107–109] were given a detailed interpretation by *Albersheim* [32.106]. Ordering tonal properties that have objective correlates in physical parameters as well as tonal attributes (these are phenomenal descriptions of sensations) in regard to their mutual relations, he derived a scheme for pure tones and another scheme for complex tones of which slightly adapted versions (to account for differences between the original German terms and English translations) are shown in Fig. 32.7a,b.

Likewise, *Albersheim* [32.106, p. 268] condensed his discussion of the stimulus parameters and sensory attributes of complex tones into the scheme shown in Fig. 32.7b.

The term mixed color denotes the phenomenal quality of a sound resulting from the composite effect of fundamental frequency and spectral energy distribution. Different spectral patterns varying in the number and strength of partials determine the sensation of brightness as well as the similarity such complex sounds may have with vowels in speech and singing, on the one hand, and timbres known from musical instruments, on the other. In addition, the phenomenal quality of a complex sound can change when it is shifted up and down a musical scale. The term *mixed color* expresses

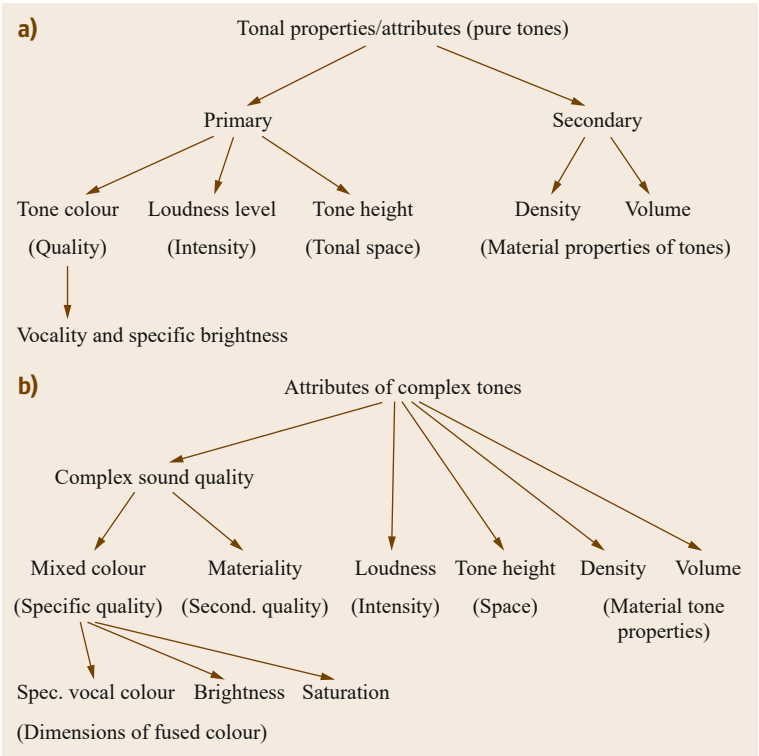


Fig. 32.7 **(a)** Tonal properties and attributes for pure tones (after [32.106]) **(b)** Properties and attributes of complex tones (after [32.106])

the combined effect of all these factors. Materiality (German: *Stofflichkeit*) denotes material properties of a sound one can *touch* and *feel* (analogous to haptic sensation), for instance, roughness or smoothness of sounds. *Albersheim* [32.106] devoted a significant part of his monograph to phenomena and categories known from the study of phonetics and voice quality in regard to their applicability to musical sound, addressing aspects like *specific vocal color* of complex harmonic tones. Vowels had been covered by *Köhler* [32.93–96, 100] and given systematic treatment by *Stumpf* [32.92] on the basis of many experimental findings.

In contrast to the elaborate descriptions of tonal attributes provided by Stumpf, Hornbostel, and Albersheim, the cognitive psychologist *Ebbinghaus* [32.110, p. 306] gave a neutral definition of sound color when he said the *sound color* (*Klangfarbe*) of tones is understood as that which distinguishes them in sensation, at identical pitch and intensity, when produced from different instruments or voices. This statement implies that there are distinct sound properties giving rise to (at least three) basic sensations: pitch, intensity, and sound color. However, the nature and perceptual *content* of sound color were left out of this definition which, much later, resurfaced in nearly identical shape as that of *timbre*, issued by the American Standard Association ([. . .] *the attribute of sensation in terms of which a listener can judge that two sounds having the same loudness and pitch are dissimilar* [32.111]).

32.2.2 Modern Approaches: From the *Objet Sonore* to Multidimensional Scaling

As is evident from the preceding sections, approaches to the classification of sounds in regard to sound color or timbre originally started from fundamentals of vibration, sound production and radiation as realized in certain instruments (including the voice). As far as sensation and perception is concerned, the basic concept was that of psychophysics where one seeks for a physical input parameter causing an output in sensation that is measurable (and possibly scalable in some unit, Sects. 30.1.2–30.1.4). This view was gradually but steadily changed when, in the descriptive and analytical studies published by Stumpf, Rich, Hornbostel, etc. the experience of subjects in perceiving sound played an increasingly greater role. Though physical and physiological facts known from measurements and observations were still taken into account, the focus in music psychology was on tonal *attributes* relevant for perception. The phenomenological paradigm reflecting the experience of musically trained subjects is clearly evident in the monograph *Albersheim* [32.106] devoted to attributes of tone and sound.

In the following, a number of approaches to the study of timbre and sound color will be discussed in regard to developments in music and media technology as well as in empirical research methodology. For obvious reasons (given the large number of publications on timbre, and on *sound* and perceptual aspects of sound in general), the survey must be selective.

Developments in Audio Technology

Technical production and reproduction of music as sound began after the phonograph and the gramophone had been introduced to the public [32.112]. Radio transmission followed in the 1920s when electric amplification and recording on the basis of vacuum tube technology was developed, and solid-state technology (the transistor was invented in 1947/48) came into use during the 1950s. By about 1930, the *Trautonium* (a kind of early synthesizer) had been invented, followed soon after by its polyphonic expansion, the so-called Mixtur-Trautonium; these instruments were used by composers of modern music (P. Hindemith, H. Genzmer). Also in the 1930s, the Hammond organ became available in several models (making use of rotating tone wheels and electromagnetic pickups for sound generation). Shortly before and during World War II, the modern tape recorder was developed (even in stereo, by AEG). The vinyl LP (12″, 33 1/3 rpm) was issued in 1949. Evolving technology offered new possibilities to creative artists engaged in what became *electronic music* [32.53]. Institutions like the Studio for Electronic Music at Cologne (part of the public radio station, NWDR, later WDR, and home turf of composers Herbert Eimert and Karl-Heinz Stockhausen) offered technical facilities including a range of sound generators, filters, ring modulators, and tape recorders as well as electronic keyboard instruments (for technical information and musical aspects, see [32.113]). *Moles* [32.114, Chap. IV] elaborated on sound according to physical (time, frequency, SPL) and psychoacoustic parameters (envelope, spectrum, level and dynamics, pitch, timbre), which can be used for both analytic description of existing *sonic objects* as well as for the creation of new ones. His main idea was that music as is performed, recorded on tape or other medium, and perceived by listeners consists of *objets sonores* that can be distinguished from one another, and can be decomposed into cells in a frequency/level plane representing the area defined by lower and upper limits of human hearing. *Moles* [32.114, p. 118–119] considered such cells (he uses terms like *cellules* as well as *quanta sonores*) as carriers of information, thus relating to concepts of communication developed in a more formal

approach by *Gabor* [32.115] and by *Shannon* [32.116]. Moles also offered rules of how to work with sonic objects. The acoustic and psychoacoustic description of sonic objects met the scientific spirit encountered in electronic music circles. The tape recorder became instrumental in collage and montage techniques employed not only in electronic music but also in *musique concrète* that made use of a broad range of natural and technical environmental sounds [32.64]. The idea to expand sound material from tones and chords as well as complex sonorities into *musical noise* had been propagated by Italian *Futuristi* [32.117] and had been pursued by some composers in Europe and overseas.

Schaeffer's Typology of Sonic Objects

An attempt to deal with the huge range of sonic and musical objects serving as material for contemporary music was the comprehensive *solfège des objets musicaux* prepared by *Pierre Schaeffer* [32.118, pp. 387–597]. *Schaeffer* condensed his considerations into a scheme [32.118, pp. 584–587], which is formally organized like a matrix (*m* rows, *n* columns). The rows contain what he calls criteria of musical perception (critères de perception musicale); the criteria are masse, dynamique, timbre harmonique, and profil mélodique, profil de masse, grain, allure. These terms refer to volume (in the sense of Stumpf, Rich, Stevens), dynamics, and spectral as well as tonal structure in sonorities. Since Schaeffer's scheme is not confined to single sounds, and in fact tries to establish a typology of musical objects, he includes basic types of melodic contour (like the *podatus* and *torculus* known from Gregorian chant and neume notation). Grain (on the level of sounds) refers to the *surface* (which can be rough or smooth), and allure means the envelope, which can be regular (with or without vibrato) or irregular, etc. The columns comprise qualifications (types, classes, genres) and evaluations according to *species* (espèce, which also can designate a sort or a kind) in regard to the *height* (hauteur), intensity, and duration of sonic objects. To be sure, the typology offered by Schaeffer was a bold attempt which, however, is based on personal experience and description rather than on systematic treatment backed by empirical data. As a typology, it serves to order phenomena according to certain aspects even if it is not consistent in every respect (the matrix has some empty cells). Like in the phenomenological description of tonal attributes [32.106], the perspective is that of the appreciative subject perceiving sound and music.

Semantic Attributes of Timbre

Though subjects experience sounds as sensations characterized by certain qualities and intensities, perception can involve verbalization when judgment based on sensory data includes an act of predication [32.119]. Since sounds and their timbres allow for verbal description (and, moreover, can convey a musical or extramusical meaning), people speak of a *rumbling thunder*, *gurgling water*, *whistling wind* as well as of a *whining violin* or a *blaring trumpet*. Most languages contain very many adjectives used to characterize certain sounds or their timbres, and one can find a host of such adjectives in reviews of concerts and recordings as well as in other publications on music. In psychology, sets of adjectives were used to characterize expressive qualities of musical pieces or phrases in order to disclose their affective mood and their *meaning*, in regard to listeners [32.120].

The study of semantic meanings that things or persons or ideas have for various people was pursued by *Charles Osgood* et al. with a methodology known as semantic differential [32.121] and also as *polarity profile*, in areas of psychology. The goal is to measure attitudes and preferences of people as they value or dislike certain things, persons, ideas, etc. Osgood et al. conceived of a *semantic space* in which various concepts (this term was used instead of *stimuli*) could be placed according to the judgments subjects give on lists of adjectives (or nouns) ordered so as to form bipolar pairs (like soft–hard, good–bad, fast–slow). In between such polar adjectives, a scale is inserted for which *Osgood* et al. [32.121, p. 85] proposed seven alternatives. These can be expressed by numbers or by verbal qualifications suited to indicating a scale. The technicalities of scale construction are in fact as important for the outcome and the reliability and validity of experiments as is the selection and the arrangement of adjectives (some critical issues are discussed in [32.122–124] and [32.125, p. 73 ff.]).

A semantic differential or polarity profile usually comprises from about 30 to 70 pairs of adjectives, which are the variables used to characterize a number of items (*concepts* in Osgood's terminology), by a sample of subjects. The resulting data (blocks of variables × items × subjects) is subjected to the calculation of descriptive statistics (means, variances, etc.) and of intercorrelation matrices, on which in many studies factor analysis (FA) with respect to PCA (principal component analysis, usually with varimax rotation) has been performed. Without going into details of FA (besides PCA, there are several other methods and models in use), it should be noted that the methodology involves a load of vector and matrix algebra as well as the representation of the variables and the factors calculated from the variables in a *k*-dimensional vector space [32.126]. To be stable as topological constructs and reliable and valid in regard to interpretation, fac-

tor models must conform to geometrical axioms and principles. *Osgood* et al. [32.121, p. 91] suggested to conceive of distances between *concepts* in the semantic space in terms of linear geometrical distances. To calculate these distances without violation of geometrical axioms (and minimizing the risk of artifacts), the raw data should be interval or ratio scaled. However, this is a condition rarely fulfilled in studies based on semantic differentials. Moreover, these seem to comprise two different kinds of scales as the adjectives selected for the polarity profile in many studies are either denotative (descriptive in regard to features and attributes of items) or connotative (associative in regard to features of items); some adjectives are in between these categories. Consider, for example, sounds recorded from various orchestral instruments (strings, woodwinds, brass, percussion) that differ in temporal envelope and spectral structure (Sects. 30.2 and 32.1.2). On the denotative level, one could offer adjectives such as soft–loud, dull–bright, smooth–rough, transparent–dense. Further, there are adjectives such as *nasal* that are widely used but may refer to more than one acoustic property [32.127]. In addition, there are adjectives that may denote timbre qualities more indirectly. For instance, the sound of a trumpet appears *shining* to many listeners; the objective basis for this is spectral energy distribution and centroid. Then there are adjectives that express connotative rather than denotative meanings associated with certain timbres like *doleful*, *exciting*, *gloomy*, etc. The problem is that the scales used for denotative attributes can be regarded as rating scales (subjects try to estimate a certain quality and/or intensity on the basis of a sensation) while the connotative adjectives rather call for an emotional appraisal or associative guess (consider, for instance, a pair like *dreamy–awake* in regard to the violin sound of Isaac Stern playing the opening measures of the Adagio in Bruch's violin concerto).

The semantic differential (often coupled with FA) has been used in the study of musical and other sounds [32.128–132]. In *Osgood* et al. [32.121, p. 36 ff.], the *standard* solution obtained from FA consists of three factors, the first of which was interpreted as evaluative, the second as a *potency variable*, and the third as *activity*. Such a solution restricted to three factors seems plausible if the factors allow for a rather wide interpretation and thus can embrace a larger number of variables (in this case, pairs of adjectives). It should be noted that *Wundt* [32.133] had proposed a three-componential model of emotions based on three polar pairs (pleasure–displeasure, excitement–inhibition, tension–relaxation) from which a number of additional emotions were derived as combinations (e.g., joy is derived from excitement, pleasure, and tension). In a factor model consisting of three independent, that is, orthogonal factors, these fit into a three-dimensional space so that the items can be conceived as *points* (relative to coordinates x, y, z in Euclidean space) and the distances between items (as well as between subjects) calculated as Euclidean distances where the linear distance between pairs of points ($a = x_i, y_i, z_i$; $b = x_j, y_j, z_j$) is interpreted as a measure of phenomenal similarity or dissimilarity respectively. Finally, three factors in many empirical studies suffice to explain a substantial percentage of the variance in the data. However, the factors derived from FA again are vectors that do not always lend themselves easily to factual or perceptual interpretation in regard to the items and/or subjects. Also, the interpretation calls for some *label* that will be attached to a factor. *Rahls* [32.129] found for 20 sounds and 47 pairs of adjectives a four-factor solution of which the first three were interpretable: the first as an *evaluative* factor, the second as one that involves *activity* and the third viewed as *potency* (see above). In *Jost*'s [32.130] study of clarinet sounds, one factor pointed to *volume* as an attribute, and a weak factor indicated the specifics of the clarinet timbre.

In contrast, *von Bismarck* [32.131] found for his sounds (sine tones, complex tones, noise) a strong factor accounting for 44% of the variance that relates to spectral energy distribution and presence of energy in higher frequency bands. This factor, which quite clearly reflects a sensory attribute [32.134], was labeled *sharpness* (German: *Schärfe*). In an experiment taking up some of Bismarck's pairs of polar adjectives as *differentials* but using dyads played from wind instruments, *Kendall* and *Carterette* [32.132, p. 455] found that *sharp is not a good discriminator across wind instrument dyads* since *sharp* in English refers to *pitch* rather than timbre. The problem encountered with such terms as *Schärfe* $\neq$ *sharpness* of course is one of semantics. Many of the terms used in description of sounds, even if denotative in essence, are taken from other spheres such as optics (brightness) or haptics (roughness, smoothness). Certain attributes are intermodal in regard to sensation as is the case with brightness [32.135] and apparently so with roughness (which is scalable both as an auditory attribute of sound and as tactual experience; *Zwicker* and *Fastl* [32.136, Chap. 11], *Stevens* and *Harris* [32.137]). Also *Schärfe*, which one could tentatively translate as *stridency* can be addressed as an intermodal attribute. The German *Schärfe* refers to the condition of a blade of a sword or knife that has been sharpened; the word also denotes the sharp outer edge of a swinging or carillon bell profile. Thus, the original meaning is in the tactual and haptic area of sensation. The term can be applied to sounds sensed as glaring, strident, penetrating and perhaps even *hissing* if they contain a strong proportion of energy in higher parts of the audible fre-

quency range. Such sounds can be easily generated in harmonic complexes when amplitudes of partials increase in proportion with harmonic number [32.138]. Another method simply is bandpass or high-pass filtering of harmonic or inharmonic complexes or noise bands so that energy is concentrated in the frequency range corresponding to higher critical bands (CBs) of the auditory system.

The problems involved in verbalizing sound and musical phenomena in general are complex and difficult to solve. In certain ways (reflected in a number of essays by *Charles Seeger* on fundamentals of musicology [32.139]), it would be recommendable, though perhaps not always practical, to communicate in the medium of music about things musical instead of using speech. From the viewpoint of empirical methodology, it seems wise to reduce semantic ambiguity in verbal descriptions of sound as far as possible, which can be done by selecting such adjectives and nouns that denote certain sound characteristics rather than using connotative adjectives (where subjects in a sample often have ideas as to their meanings that differ widely). Also, verbal descriptors should be checked in experiments for consistency relative to various sets of stimuli as well as for their strength of differentiation between items (one of the issues with semantic differentials is the lack of standardization [32.122]).

In order to avoid pitfalls that can happen with the methodology of semantic differential and FA interpretations, the use of adjectives combined with unipolar scales and a robust method of data analysis such as hierarchical cluster analysis seems advisable. This approach was chosen by *Thies* [32.140] in an attempt to find fundamental categories for a descriptive classification of sounds that could bring Schaeffer's more intuitive scheme to a systematic and formal typology. Starting from the German vocabulary which, in regard to sound and its attributes, offers some 1600 descriptive and connotative words, he selected 51 *general* classifiers (e.g., loud, soft, rough, smooth, high, low) and 382 more *specific* (like nasal, creaky, buzzing, pulsating, whispering). In experiments offering musical and environmental sounds as stimuli, subjects were asked to judge which of the classifiers applied to certain sounds, and if so, to what extent. Cluster analysis found groups of adjectives that represent basic descriptive sound categories. From these groups, pairs of adjectives (like tonal–noisy, dark–bright, soft–hard) representing fundamental spectral and temporal features and sensory attributes (tonalness, brightness, loudness, etc.) were considered as fundamental classifiers to be used in a first-level analysis and classification while, on the next level, more specific classifiers are used for a fine-grained classification and typology.

The problem of semantic meanings adjectives have with respect to properties of sounds and attributes of timbre may aggravate if viewed from an inter-language perspective. Inasmuch as languages reflect cultural norms and experiences shared by ethnic and language groups, it is likely that cultural differences manifest themselves in different concepts and terminology even though acoustic properties of sounds and also parameters of auditory perception are identical among such groups. In a comparative study with subjects having either Greek or English as their native tongue, the convergence of descriptive adjectives was tested with the verbal attribute magnitude estimation (VAME) method [32.123, 141, 142] which avoids semantic differentials by putting adjectives against their negation (e.g., dull–not dull, sharp–not sharp). This comparative study [32.143] employed various methods of data analysis (cluster analysis, FA, correlational techniques) and also feature extraction from the stimulus sounds (23 musical instrument tones). The data analysis resulted in three common dimensions labeled luminance, texture, and mass, where luminance refers to attributes like brightness and sharpness (to include, however, depth/thickness in the Greek sample) while texture was interpreted as related also to spectral energy distribution, and mass possibly referring to spectral density and flux. The results demonstrate that a set of sound stimuli presented to subjects from different culture and language groups may elicit responses that converge to a certain extent due to acoustic stimulus features while there are also differences in conceptualized attributes derived from descriptive adjectives. In a Swedish study on the timbre of the steady state of alto sax sounds also employing adjectives along with VAME, *Nykänen* et al. [32.144] found a combination of rough and sharp (seemingly included in the Swedish adjective rå ≈ raw, bleak), soft, warm as well as a vowel quality o-like to describe the sounds best. There was a correspondence to psychoacoustic parameters in that the attribute *rough* could be predicted from the model of roughness used by *Aures* [32.145, 146] combined with the model of sharpness proposed by *von Bismarck* [32.131, 134].

Similarity Ratings of Sounds

Another approach to timbre research is that of similarity ratings. Detection of similarity in objects conveyed as sensory input is an experience basic to humans and other species. There are several theories of similarity some of which emphasize extraction of features from sensory input that are used for comparison while others underpin the cognitive evaluation process underlying similarity judgments [32.147, 148]. In regard to the problem of verbalization addressed in the previous paragraph, one might assume that the recognition

of similarities driven by sensory input works without verbalization while an analytic cognitive evaluation of input rather would involve such. Assuming a hierarchical model that leads from sensation to perception and apperception (Sect. 30.1.2), different levels of processing would likely be reflected in shades of awareness of such processes.

In the field of sound and music perception, *Stumpf* [32.149, Sects. 6,7] especially argued that different degrees of similarity as perceived by subjects can be expressed as differences in distance, which implies similarity is scalable and differences in similarity can be translated into some distance measure. Scaling of similarity has been a paradigm for a long time, for which a standard methodology usable in psychophysics and other areas was developed [32.150, 151]. Since phenomenal similarities between a set of objects often rest on several or even many properties, the geometric concept according to which perceptual *similarity* ≡ *proximity* has led to a multidimensional perspective where objects can be represented as points in a k-dimensional space. The interpoint distances relative to a number of dimensions then express the *similarity* or *dissimilarity* between n objects on m dimensions ($n \gg m$). In addition, one can calculate and geometrically represent similarity judgments of different subjects in a sample to compare their perceptions. The methodology to carry out such empirical studies is multidimensional scaling (MDS), also termed multidimensional similarity structure analysis [32.152, 153]. There are some ideas and concepts shared by both FA and MDS, namely the idea that a subject's responses to complex stimuli comprising several or many variables can be explained and/or predicted from a small number of factors or dimensions as well as the concept to represent relations between high-dimensional (multivariate) stimuli in a rather low-dimensional geometrical space. This space, spanned as factor configuration or MDS model derived from empirical data in several steps, is assumed to reflect perceptual and also cognitive evaluations of subjects. It should be noted that the metric-dimensional approach to *similarity* perception has been challenged on methodological and factual grounds [32.147] and has also been defended as a basis valid for the study of similarity as well as recognition and identification processes [32.154–156]. Though MDS (and, likewise, FA) can be used for data reduction from a larger number of variables presented to subjects in experiment to a few metavariables labeled *factor* or *dimension*, its main function is to help shape hypotheses concerning the dimensionality of complex structures. Like in FA, MDS models can be derived in a precise manner (for Euclidean space and distance functions) if the input consists of interval- or ratio-scaled data. However, similarity judgments are ordinal as subjects find two objects (say, two sounds recorded from any two instruments at the same pitch and presented at the same loudness) are *not similar*, *somewhat similar*, *similar*, *very similar*, *highly similar* or even perceptually *identical*. To allow for a so-called nonmetric MDS, one can assign numbers to such degrees and turn them into proximities from which, in an iterative process (minimizing errors and adjusting interpoint distances between objects or other items), the *final* distances in a k-dimensional space are calculated. The distances of n objects relative to m dimensions are regarded to reflect the perceptions and judgmental decisions of the subjects. The type of metric (Euclidean, city block, Minkowski [32.152]) that is chosen for the model expresses different cognitive decision strategies. If the similarity judgment is decomposable into several more or less separate evaluations of similarity along dimensions $(a, b, c, \ldots, k)$, a so-called city-block metric reflects the additive process. In case a direct and overall estimate of the similarity between objects is performed, a Euclidean distance model sums the differences on the contributing dimensions. When dimensions have different salience for individual subjects, a weighted Euclidean distance model seems apt (though this technique requires certain precautions). A Minkowski r-metric can be chosen if, in a decision process with respect to similarity, one stimulus feature is predominant so that the dimension to which it relates seems supreme against all others. The goodness-of-fit that various models yield with respect to their metric, number of dimensions, error score, as well as the coefficient of determination in relation to the dataset, can be used in a cautious interpretation of the evaluation and decision processes involved in the perception of phenomenal similarity.

Since about 1970, computer software for metrical and nonmetrical MDS has been available, which has spurred many experiments also in the field of sound and music perception [32.157, 158]. Some of the now *classical* experiments on timbre were performed at Stanford [32.159–161]. Grey worked with 16 synthesized sounds that emulated those of orchestral instruments since one task of his research project was to explore computer-based analysis-synthesis techniques. Data from an experiment with 20 musically sophisticated subjects judging the similarity of the 16 sounds were subjected to nonmetric MDS, which did yield three dimensions interpretable in terms of acoustic properties of sounds. The first dimension quite obviously relates to spectral energy distribution and spectral envelope; the second was interpreted as relating

> *to the form of the onset-offset patterns of tones, especially with respect to the presence of synchronic-*

ity in the collective attacks and decays of upper harmonics. [32.159, p. 61]

The third dimension was also viewed as temporal and related to the energy distribution in the attack of tones.

Studies of sound and timbre in most cases have confined their MDS analyses to three or even two dimensions [32.132, 159–168], [32.169, Chap. 6] one of which seems almost notoriously related to spectral energy distribution, centroid and brightness [32.170]. The second often has been interpreted as related to onset characteristics (hard or soft attack, percussive or rather mellow sound) or to the temporal envelope in total. A third dimension sometimes has been viewed in regard to spectral flux which, in most sounds, is considerable at the onset of sound because of the transient portion and becomes small once the steady state is reached. Spectral flux (SF) can be calculated from digitized audio signals where SF expresses the rate of change in spectral composition from one frame of analysis to the next [32.171]. Some studies could not confirm the interpretation of a third dimension as expressing SF [32.158, 167] though in fact most natural sounds exhibit some or even considerable variation in both spectral frequencies and amplitudes over time. In case composite waveshapes are used as stimuli, one dimension can be interpreted as spectral energy distribution (sparse harmonics–rich harmonics) and another in terms of the *surface quality* of sounds; since a sawtooth with many harmonics in zero phase has steep slopes in every period, the sound appears comparatively *rough* while a triangular wave appears quite *smooth* [32.168].

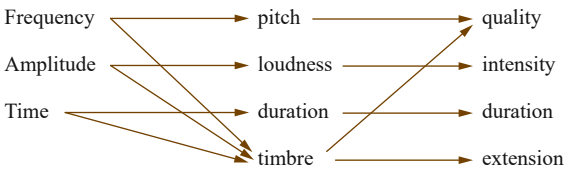
Incorporating a third dimension in a MDS model can be useful to account somewhat better for the variance in the similarity data; however, one might be confronted with a certain trade-off since, with a third dimension in a MDS model, as with a third factor in FA, one often may explain a higher percentage of the variance, on the one hand, while the interpretation of a third dimension or factor can be arduous, on the other. Not only does each factor or dimension need to be given a *label* (the verbalization problem is encountered again, if on another level), but its interpretation must also account for the factual *content* and perceptual effect of stimuli weighted in the light of the *k*-dimensional model. In general, interpretability seems to diminish with the number of factors or dimensions respectively. This can be possibly explained in regard to conditions under which similarity judgments are mostly made in real life, on the one hand, and processes of categorization, on the other. Concerning conditions, one has to take the huge amount of environmental information arriving at our senses as possible input into account. From an evolutionary perspective, *survival* makes very fast recognition of objects imperative. In a first and probably pre-attentive step, recognition involves some overall and provisional categorization in which comparison of a possibly new object to previously learned *prototypes*, *schemata* (or whatever the template is called) is performed. For an instantaneous comparison which, with respect to *survival* behavior, may trigger further sensory processes as well as motor responses, restriction to a few salient features seems pertinent. Adopting either a feature detection approach or that of dimensional metric as in MDS, reduction of the processing load is inevitable in fast estimates. The goal can be achieved successfully through exclusion of irrelevant as well as dimensional reduction of usable sensory input [32.172]. Even in many instances where the time factor is less critical, overall judgments of similarity apparently still seem restricted to a few salient features or a small number of dimensions respectively. Apparently, such suffice for a rough perceptual ordering of things into categories; listeners hearing music in a concert or from the radio can perform at least a tentative classification of sounds representing, for example, strings, brass, and percussion instruments. This can be done with reference to very few acoustical features and perceptual attributes, a fact corroborated from computer-based classification of musical sounds with respect to sources [32.173]. Many of the two-dimensional MDS models of *timbre similarity* in fact relate to the temporal and the spectral envelope as fundamental components of complex sounds; the two most common dimensions *centroid/brightness* and *onset/envelope* (see above) apparently suffice to categorize a host of sounds according to their timbre, at least in a gross mode. A closer inspection then may reveal further details needed for a finer classification. In this respect, timbre can be viewed as an emergent perceptual quality [32.174]. Of course, learning plays a role in such processes as recognition and identification of timbres where trained musicians usually perform faster and more reliably than nonmusicians. For example, it needs no expertise to assign a certain impulsive sound one hears to a class of instruments (say, membranophones) while one must have some experience to judge what the type of drum from which the sound comes could be (size, single- or double-headed, beaten with hand or stick), and it is certainly demanding to decide whether a drum sound correctly identified as that of a frame-drum belongs to a Moroccan *bendir* or to an Irish *bodhran*. However, in experiments involving synthesized FM sounds (TX 802) most of which emulated familiar instruments, differences in performance between subjects in three subclasses of musical expertise (professionals, amateurs, nonmusicians) were not just as great as one might have expected [32.167].

32.2.3 Acoustical Features and Perceptual Attributes of Timbre

Schouten [32.175] saw five acoustical parameters defining timbre:

1. Tonalness versus noise
2. Spectral envelope
3. Time (or temporal) envelope
4. Change in spectral envelope or in pitch
5. Onset characteristics, which he labeled *acoustic prefix* of sounds.

These in fact are interrelated in a number of ways, as are the four fundamental qualities of pitch, loudness, duration, and timbre that are constitutive for auditory perception. It is evident that one cannot perceive a *pitch* from a tone unless it does have a duration and sound intensity. Likewise, loudness as based on intensity and frequency [32.176] and variable also with the duration of stimuli presented to subjects calls for a sound that must have some frequency content with nonzero duration and amplitude(s). In this respect, parameters are not totally separable even though one can study the relative weight they can have in sensation and perception of complex sounds. The idea that parameters might be separable and decomposable into *dimensions* perhaps owes to classical concepts of psychophysics (Sect. 30.1.4) where sensations were studied with respect to quality, intensity, duration, extension, etc. Taking acoustical sound parameters, on the one hand, and perceptual dimensions, on the other, one can set up the following scheme:



Of course, pitch can be explained also in the time domain (Sect. 31.4), and loudness depends not only on physical intensity of the sound but to some degree on spectral energy distribution as well as on presentation time and temporal integration (Chap. 33). Further, timbres can be perceived as distinct qualities in case they offer salient features.

The view according to which pitch is separable, in principle, from timbre is not identical with that advanced by *Helmholtz* [32.90] on separability of pitch and sound color (Sect. 32.2.1). Taking the Fourier-based resonator model of Helmholtz, a harmonic complex simply is decomposed, on the basilar membrane (BM) level, into partials along a log frequency axis where the lowest partial (f_1) accounts for the sensation of pitch, and the remainder of the spectrum for *sound color*. The concept indeed implies that a sound, well-defined as to its pitch through f_1 (as well as its inverse, the period $T = 1/f$), gains some additional *coloring* from the spectrum above f_1. In this respect, sound color is a more or less stable quality corresponding to the steady state of a sound [32.177]. Further, sound color can be viewed in regard to *extension* where the number and strength of partials define spectral width, density, and centroid (for a given f_1 of, say, 110 Hz, the centroid shifts upwards in frequency with the number of additional partials). This effect is easily demonstrable with an analogue synth where a tone (say, A_2), when played with a sawtooth wave, gains in *color* as well as in brightness from sweeping the cutoff frequency of a low-pass filter toward higher frequencies. In early works on orchestration [32.3] the term *timbre* also referred to a sound quality largely determined by spectral structure while the temporal aspects as yet were of little concern. Since the 1920s and 1930s, when onset characteristics and temporal fluctuations of sounds could be studied with electroacoustic equipment at hand, transients in particular in aerophones and chordophones have been a topic of research (for a survey, see [32.31]). With computerized sound analysis, transients were investigated as dynamic time-frequency structures from which information relevant for auditory perception can be drawn [32.178]. In the following, the parameters discussed by *Schouten* [32.175] will be examined further in the light of research data.

Transients and Onset

The transient part of a sound reflects a vibrating system immediately after excitation, which can be effected by a single impulse as in many idiophones (e.g., xylophones, gongs, bells, cymbals) and membranophones as well as in plucked strings, or by a sequence of pulses as in aerophones (Sect. 32.1.2) and in bowed chordophones where excitation continues with energy supply to the generator. The transient portion of a sound can be defined as that from the absolute onset of vibration (the time point or sample where the amplitude is nonzero) up to a point where vibration becomes either periodic with small fluctuations (as in aerophones and chordophones) or where the peak amplitude is reached and the decay of the envelope begins (as in idiophones and membranophones excited by a single impulse). The transient portion of a sound is the most *interesting* part in terms of information (see [32.25] for concepts describing information structure of sounds) since information depends on entropy and the rate of change per time unit. *Shannon* [32.116] gives a formal proof that white noise has the maximum possible entropy. In comparison to the almost periodic regime of vibration in the steady state, the onset often lacks clear periodicity and

can even appear *chaotic* (as indicated by the limit cycle for the time series of a vibration or sound in a phase space). The onset of many sounds recorded from natural instruments includes noise in the transient portion; a well-known phenomenon observed in organ flue pipes is the *spitting* response to air pulse sequences before a standing wave is established (Sect. 30.2 and examples in [32.179, Chap. 5]).

For the listener, the rapid changes that the wave shape and spectral content undergo in a short time result in a high rate of information delivered to the perceptual system. In contrast, the flow of information saturates once the steady state is reached. For example, for identification of clarinet sounds it was found that scores do not improve after 0.5–0.6 s of presentation time [32.180]. In regard to detection and identification of onsets and transients as well as of changes in spectral envelope over time, there are temporal and dynamic thresholds as well as several integration constants that have been reported from experimental findings (for an overview of research, see [32.31]). Further, there are forward and backward masking effects (for a detailed discussion, see [32.181]). Some of the relevant integration constants concern the very limit of auditory temporal discrimination between two events and the threshold of temporal order between events. The limit of temporal discriminability has been given as 3–5 ms (as in experiments on gap detection), however, the absolute value depends somewhat on sound level and other conditions. Other time constants concern just noticeable differences in onset asynchronicity and the threshold of temporal order. Onset asynchronicity between instruments playing notes simultaneously typically is in the range of 0–20 ms; asynchronicity within this range (and approaching its upper limit) supports identification of instruments [32.182]. Onset, in this respect, means the sensory and perceptual *event* that of course is dependent on the vibration pattern of sounds yet is not identical with the SPL measured at a certain time. The perceptual onset for tones in succession will be detected when the amplitude of vibration and the SPL of the sound radiated from a source exceed a certain threshold, which has been given as 6–15 dB below the maximum level of a tone [32.183]; this threshold is variable with respect to the absolute SPL of the stimulus and is also variable due to masking effects if several tones/sounds are played in close succession or nearly simultaneously. Two tones played in succession become clearly discriminable as *auditory events* when their perceptual onsets are about 40 ms apart; the threshold for audible single reflections of sound from walls to be perceived as *echoes* is in the range of 40–50 ms so that several early reflections occurring within a shorter time span ($t < 30$ ms) typically *smear* into one sensation. Given that ≈ 50 ms mark the time difference between events to be perceived as an orderly sequence and that complex tones in general are identified in regard to their pitch, timbre and loudness in a time window of ≈ 100–200 ms, it follows that the number of notes in music that are clearly identifiable as a melodic and diastematic structure, plus conveying a certain timbre and loudness, is limited per time unit.

Onsets can be very short as in swinging and carillon bells where, after the energy is transferred from the clapper to the bell (contact time ≤ 1 ms [32.184]), a large number of modes is excited within a few ms since the wave speed in bronze is ≈ 4400 m/s for longitudinal and ≈ 2160 m/s for transversal waves. The sound radiated from the bell or from a similar idiophone struck with a hard mallet contains many strong harmonic and inharmonic partials as a complex mixture [32.85, 86, 185] from which the auditory system must derive pitch and timbre information [32.186]. The onset of such sounds (which fall into a time window of 20–50 ms) usually is perceived as *clangorous* due to considerable spectral inharmonicity as well as high-frequency content, and assignment of a single pitch to the onset segment often is not possible. If idiophones struck with a mallet are at one end of a scale measuring the acoustical transient portion of a sound, there are some instruments on the other where the transient is long and can last for several hundred milliseconds as, for instance, in a double-bass played softly with a bow [32.187] or in large organ flue and reed pipes [32.35]. Also, some folk instruments such as the Slovak duct flute *Fujara* (of ≈ 160 cm tube length [32.41]) are slow in the buildup of modes. However, the dynamics of the onset of a sound is much dependent on playing technique as well as the dynamics of the musical context in which instruments are used. In works of modern music, notation might prescribe dynamics from *fffff* to *ppppp* and might include verbal instructions that range from *tutta la forza* to *morendo*. Also, composers in the decades after 1950 demanded instrumentalists (and also singers) to produce sounds in often uncommon ways, which could bring about, for example, quite percussive sounds from wind and string instruments. The variability of onsets due to playing technique and context notwithstanding, subjects familiar with music and orchestral instruments make use of the onset of sounds as information bearing to the classification and identification of such instruments.

For listeners, the transient part of sounds serves to distinguish various instruments as well as to mark the onset of notes. Transients such as the *spitting* of organ pipes, the prerunner sound in harpsichord strings (when the plectrum touches the string before lifting and plucking it [32.188, 189]), the *raspy* attack of a bow set to

motion on a cello string (in particular the C- and the G-string), the plosive attack in a trombone or trumpet tone, are such *prefixes* (to use Schouten's term) to the steady state that make sounds characteristic of certain instruments or families of instruments. Some instruments also have a peculiar decay or have markers at the end of a tone; for instance, in most harpsichords one can hear the jack fall down when a key is released, an effect that is prominent when full chords are played. In an ensemble performance, the *prefixes* from individual instruments can help listeners to notice the onset of the tones they play (listen, for example, to Hindemith's *Kleine Kammermusik für fünf Bläser,* op. 24,2). If attack transients and the final decay to zero amplitude are removed from natural instrument tones, identification scores drop significantly even for musically trained subjects where the effect is greatest for the attack removed [32.190]. Conversely, in an MDS study [32.166], similarity ratings for onsets from natural instruments presented in isolation did not differ much from complete tones, indicating that the onset contains most of the relevant information [32.25]. Removing perceptual cues like the transient at the onset means an increase in confusability of stimuli. Apparently, the effect can be compensated, to some extent, by playing technique. For example, instruments typically played with vibrato (flute, violin) are less affected by such removal since the vibrato then may substitute the cue needed for identification [32.191].

Time (or Temporal) Envelope

Isolated sounds from various instruments can be distinguished with respect to their temporal envelope, which can be segmented into characteristic parts (as in the attack, decay, sustain, release (ADSR) model Sect. 30.2). For instruments excited by a single impulse (many idiophones, most membranophones) a rapid attack is followed by a fast decay. Large bells have a long sustain, however, when sound is radiated on a low level for tens of seconds. In aerophones, tones die away quickly once the air supply to the generator is stopped. In plucked and bowed chordophones, sustain can be quite long after excitation has stopped because parts of the resonator and the air enclosed in the box still vibrate (due to a storage of energy). For the tone G_2 played on the open string of a cello (staccato, forte), the sustain lasts for ≈ 3 s after the bow has been lifted from the string (Fig. 32.8; the sound was recorded ≈ 0.5 m away from the instrument in a *dry* studio).

In this tone, the rapid attack and the long sustain are perceptually salient envelope features. In general, there are certain shapes of envelopes which, combined with spectral patterns, probably make up *timbral prototypes* for listeners. Changing the temporal and/or the spectral envelope means part of the information does not match a certain type of instrument. For example, reversing a natural complex sound affects both cues since, even though the total energy contained in the sound is the same when played forward or backward, the order in which relevant features become audible is reversed, which makes identification of natural sounds reversed in time difficult. The effect of time reversal of sounds was used quite extensively in the 1960s, in recordings of *psychedelic* music, where usually one electric guitar track previously recorded and then played backwards provides lead guitar lines in an otherwise normal mix (listen, for example, to The Byrds: *Thoughts and Words*, recorded in December 1966). An important factor for sensation and perception of temporal envelopes is modulation. Regular amplitude modulation (AM) can result from narrowly spaced spectral components, which would give rise to a sensation of roughness

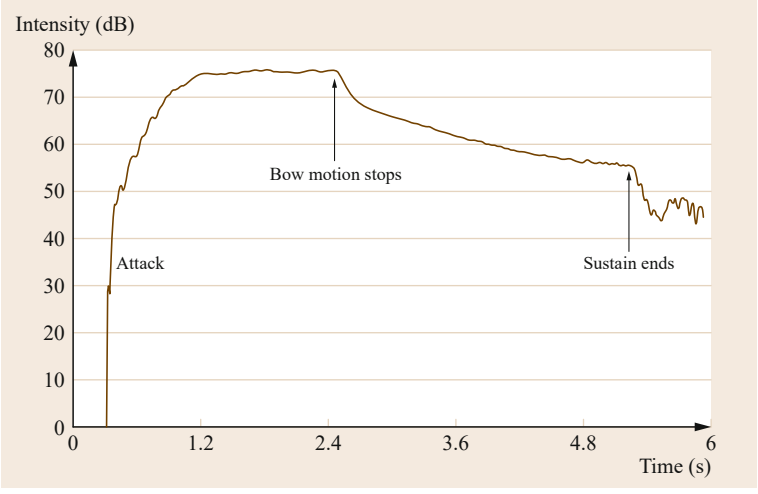


Fig. 32.8 Cello, open G-string (G_2), staccato, forte, steep attack, long sustain

(Sect. 31.6.2) in addition to the amplitude fluctuation. Quite regular AM with very little FM of f_1 and without introducing roughness can be produced also by modulating excitation parameters such as the blowing pressure in a flute [32.192].

Tonalness versus Noise

Tonalness usually is understood as a sensory attribute dependent on spectral harmonicity as well as its equivalent, periodicity of the time signal. Tonalness is one factor, besides roughness, sharpness (or stridency, Sect. 32.2.2: *Semantic Attributes of Timbre*), and loudness in a psychoacoustic model of *sensory euphony* [32.145, 146]. In regard to spectral harmonicity ≡ temporal periodicity (consequent to the Wiener–Khintchine theorem, Sect. 30.2), the steady state of a tone from an aerophone or chordophone played with medium force of excitation is basically tonal while the transient onset can contain noise in significant quantity. Some noise also results from the motion of the bow on a string or from the flow of air in wind instruments (there are playing techniques for flutes and saxophones emphasizing a *breathy* sound). In order to make synthesized sounds appear natural to listeners, their transient part needs to be shaped and noise must be added to sinusoidals in a harmonic or inharmonic complex [32.193, 194].

Spectral Envelope

Perhaps the most important component in timbre perception (and the defining criterion for *sound color*) is the spectral envelope. Additive or subtractive synthesis as was implemented in a host of electronic organs of the 1960s produces sounds which, though lacking characteristic onset and decay of natural instruments, at least are indicative of certain classes and types of instruments (those organs usually offered a selection of *flutes*, *reeds*, and *brass* sounds in more or less fair imitations of the originals). Spectral envelopes for these classes often were derived from filtering a complex (e.g., rectangular) waveshape so that concentration of energy in a band from 1−2 kHz resulted in a more nasal (*reed*) sound while emphasis on energy in low partials and in higher bands (2−5 kHz) should indicate a *brass family* sound. Flutes simply were imitated by low-pass filtering with the cutoff frequency set for a *mellow* sound. Since such synthetic steady-state sounds after a short while are experienced as *static* by listeners (for the lack of fresh information, see above), most electronic organs offered some AM (*Tremolo*) unit modulating amplifier output level while some had facilities to enrich the sound (for instance, by pairs of oscillators slightly detuned against each other so that organ stops with double sets of pipes such as *vox coelestis* or *unda maris* were imitated).

The concept according to which certain instruments or families of instruments distinguish themselves by spectral structure and the shape of the spectral envelope owes much to the source-filter model (Sect. 32.1.2). According to empirical findings [32.169, Chap. 6], timbre is essentially determined by the absolute frequency position of a spectral envelope, which suggests the perceptual attribute of timbre has a physical correlate in formant-like spectral energy distribution. Given that many instruments (aerophones including the singing voice, chordophones) are driven by pulse sequences fed into a resonator [32.195], the sound radiated from the instrument depends significantly on the geometry of the resonator. The size and shape of the resonator largely determines the spectral envelope (assuming the instrument is in its normal register, and played *mf*), which is sensed as a peculiar *sound color* [32.177]. This concept of approximately constant tone or sound color is behind organ pipe stops where different stops (or ranks) of flue and reed pipes distinguish themselves by their sound color, for pipes of the same pitch (determined basically by pipe length, measured in foot, e.g., 8′, 4′, 2′). To maintain a given sound color (e.g., diapason, salicional, flute, trumpet) over several octaves, organ builders follow certain mensuration rules [32.196, Chap. 3]. For example, for two flue pipes an octave apart, the ratio of the pipe lengths is 2 : 1 while the pipe diameters should have a ratio of 1.682 : 1 (and the cross-sectional areas should be in the ratio of 2.828 : 1). That is, approximate homogeneity of sound color from one pipe tone of a given rank to the next is achieved by means of scale factors [32.11, 196, 197]. Correct scaling ideally would produce almost identical spectral envelopes for all the tones within the gamut of several octaves; the envelope then simply is shifted along the frequency axis with rising f_1 of each tone. To illustrate the case, formant filter envelopes of three organ tones recorded from the same stop are shown in Fig. 32.9. The tones are C_2, C_3 and C_4 from a trumpet 8′ stop of the historical organ at Hollern. The envelopes are fairly similar given that the microphone distance relative to the three pipes was not identical and that some reflections of sound inside the organ case may have occurred (sound levels were normalized to −6 dB fs for analysis).

A scale factor that preserves the relations within a specific geometry can also be used for peals of (swinging or carillon) bells where the scaling of size and mass determines the pitches but should (ideally) not affect the sound color. Further, the same principle can be applied to *families* of instruments that cover different registers (soprano, alto, tenor, baritone, bass) but share the same basic sound color [32.195] in the steady-state portion of sound when playing conditions are kept almost constant. A violin, a viola, a cello and a double

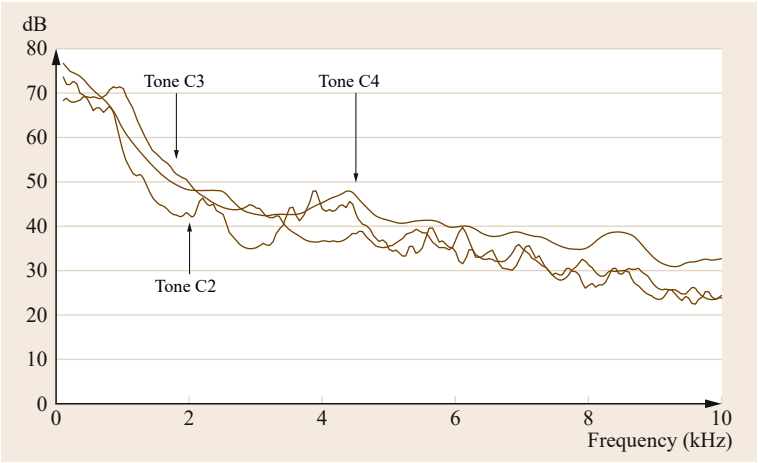


Fig. 32.9 Formant filter envelopes for three tones (C_2, C_3, C_4) of one organ stop (trumpet 8′)

bass are recognized as members of the *bowed strings family* notwithstanding differences in register and variation in spectral fine structure. If types of instruments can be distinguished (and possibly identified) by their respective spectral envelopes sensed as *sound color*, a hypothetical explanation is that the spectral energy distribution in the sound corresponds to a specific excitation pattern along the BM that is learned as representing a certain sound source, and is stored as a template or profile in long-term memory (LTM). However, one has to take the variation in sound color into account, which may occur in different registers of a single instrument, or even within one octave of its playing range. Also, factors such as the strength of excitation (usually from *ppp* to *fff*), dependence of spectral energy distribution and of the directivity pattern of radiation on sound level as well as room acoustics (absorption, reverberation) all influence the *timbre* one perceives of a given instrument [32.52, 198], [32.169, Chap. 6]. The problem of how homogeneous timbral qualities are relative to the tones of a musical scale within one octave, and more so if scales extend into another octave, has been studied empirically. Research on this topic involving MDS was done by *Marozeau* et al. [32.199] who concluded that pitch differences of tones within one octave had but little effect on timbre dissimilarity judgments. With respect to wind instrument sounds, data from musically naive subjects suggested timbre is perceived as identical for tones within one octave [32.200]. A replication of the experiment showed, however, that musicians can make reliable judgments beyond that range [32.201].

Change in Spectral Envelope and Pitch

A close inspection of many musical sounds reveals that both the period length of the complex waveshape and the frequencies and amplitudes of spectral components vary with time. For the steady state of complex harmonic sounds recorded from aerophones and chordophones played without vibrato, the variance of period length T (ms) or its inverse, the fundamental f_0, can be quite small, so that no pitch modulation is audible (in line with autocorrelation function (ACF) analysis). Pitch shifts, however, will be encountered in case tones are played with vibrato, which is customary nowadays for flute and violin performances, especially for the repertoire of the *romantic* era. Vibrato is also used extensively in belcanto singing (Fig. 31.27). Vibrato is performed, for instance on a violin or other bowed string instrument, rolling the cup of the finger up and down on a string whose vibrating length is thereby varied periodically. Vibrato thus produces FM; the modulation frequency usually applied by violinists is about 5−8 Hz and modulation of f_1 and higher partials can reach or even exceed ±35 cent. Vibrato can also give rise to AM of harmonic partials when their frequencies move in and out of the narrow resonance zones of the resonator [32.202]. Consequently, the spectral components in general show periodic AM along with the FM from the vibrato, however, individual components can be affected differently dependent on their position relative to the resonance zones. In sum, the pattern of FM plus AM resulting from vibrato may deviate somewhat from strict periodicity. To illustrate the case, a small segment from the recording of a professional violinist playing the note c#5 (over a chord provided from soft piano accompaniment) is shown in Fig. 32.10.

One can see that the violin partials undergo FM while they exhibit AM to different degrees; some of the tracks marking partial frequencies shift between strong and weaker amplitudes as indicated by black and gray color respectively. Players of the modern concert flute learn a special technique of breathing and chest muscle control [32.203] that enables them to produce FM vibrato and AM tremolo effects. Periodic variation of

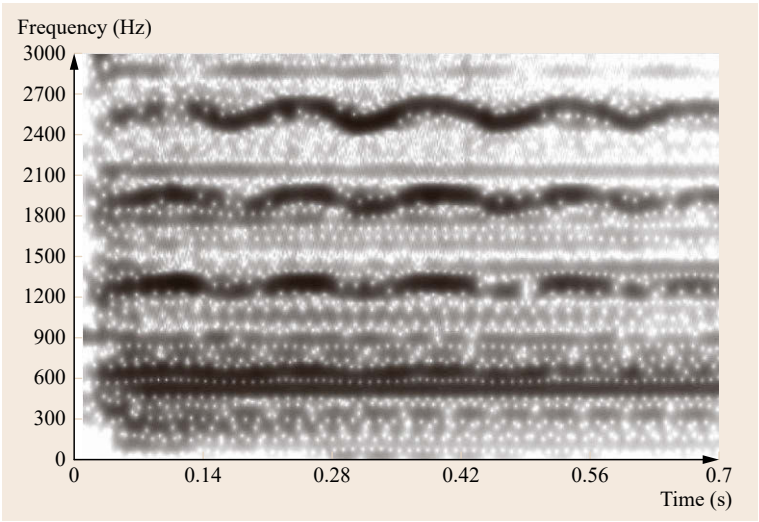


Fig. 32.10 Violin, note C#5 ($f_1 \approx$ 554 Hz), vibrato, FM and AM effects

blowing pressure results in a tremolo without significant FM [32.192]. Changes in blowing pressure means the force of excitation is varied over time, which results in different numbers of natural modes that are excited in the resonator. Also, amplitudes of partials contained in the sound radiated from the instrument vary as a function of blowing pressure. In general, higher levels at the input of the system produce more partials so that the spectrum broadens toward higher frequencies. Consequently, the spectral centroid also shifts upwards in frequency, which can be sensed as an increase in brightness. Further, subjects are sensitive to changes in timbre evoked by phase shifts between harmonic partials [32.162]. The effect is strongest for harmonic complexes where partials are either in sine or in cosine phase as opposed to partials alternating in sine and cosine phase; the strength of the timbral difference depends on f_1 of the complex and varies markedly between subjects. For $f_1 = 294.4$ Hz the maximal effect of phase on timbre for a sample of eight subjects was equal to changing the SPL by about 2 dB. A possible explanation is that the peak amplitude per period and the crest factor are higher for harmonic partials in cosine phase compared to partials in alternating phase. For harmonic complexes consisting of five partials with $f_1 = 200$ Hz and amplitudes decreasing by -6 dB with harmonic number, the difference is ≈ 1.7 dB.

Summing up this section, *timbre* was found to depend on several temporal and spectral properties of sounds, which often combine into complex spectrotemporal patterns that are analyzed into constituents in perception in order to distinguish individual instruments or other sound sources. Modeling sensory analysis as carried out in the auditory periphery, the concept of CBs usually is implemented by chains of bandpass filters suited to perform spectral analysis as a basis for perception of pitch (Sect. 31.5). Such an approach can be followed also for timbre and loudness though in particular the transient part of sounds requires high temporal resolution, which means a conventional Fourier-based analysis may fall short of the performance of the auditory system, which is superior in regard to the *uncertainty relation* $\Delta f \Delta t$ [32.204]. For the steady state of most sounds, spectral envelope and centroid have proved to be good descriptors of sensory attributes (see above and also *Barthet* et al. [32.205] for clarinet tones). Viewed in terms of CBs, patterns of spectral energy distribution code pitch and timbral information as well as loudness (*Pollard* and *Jansson* [32.206], Sect. 32.2.4 and Chap. 33). Applying time constants relevant for auditory perception, timbre can be approached as a sequence of *windows* or *frames* that contain spectral energy distributions. If there is not much change from one frame to the next, a more or less stable spectral profile evolves, which can represent a *sound color* that in turn may indicate a certain instrument or *family* of instruments [32.195]. However, identification of instruments or other sound sources is facilitated when temporal cues are offered along with spectral information. Experiments have shown that attack time is an important cue where either soft onsets or steep slopes (Fig. 32.8) help to categorize sounds. In addition, the overall shape of the envelope can indicate whether sounds are percussive rather than continuant. Further, modulation (AM, FM; regular, quasiperiodic, or irregular) can be used as a cue for timbre perception and categorization. Sounds undergoing modulation (AM and/or FM) might appear *raspy* or *blurred*; some sounds appear *shimmering* or *clangorous* or *ringing* due to spectral inharmonicity and modulation. The actual

effect on sensation depends on modulation parameters (depth, frequency) as well as on how many spectral components are modulated and if modulation is conjoined for these components or not.

32.2.4 Interrelation of *Pitch* and *Timbre*

In concepts of tone perception developed in psychophysics [32.74, Chaps. 9–11], [32.176, 207] the pure tone figures prominently since it allows the establishment of a close correspondence between physical and sensory magnitudes. In regard to pure tones, it is feasible to address pitch as depending on the frequency of vibration, and loudness proportional in some way to the vibration amplitude and intensity of radiated sound. Duration does not pose a problem assuming time is a linear process (equivalent to the constant motion of a mass point in 3-D-space [32.208]). This leaves sound color or timbre as the perceptual quality that relates to several physical parameters (Sect. 32.2.3). As has been argued above, sound color constitutes a quality complementary to pitch if one adopts the perspective of Helmholtz based on Fourier and Ohm, which allows the decomposition of a harmonic complex into the fundamental (f_1) carrying all or at least most of the pitch information, and the remainder of the spectrum conveying the sound color that may characterize a certain instrument or family of instruments. Such a view might hold for the steady state of a harmonic complex where f_1 is dominant and the amplitudes of the other partials roll off at a certain rate (dB/oct) in the spectrum so that the pitch is unambiguously related to the fundamental. However, there are musical sounds in particular from idiophones and also membranophones where this model fails to capture relevant structures. For example, in *Western* swinging bells and carillon bells there are some partials close to harmonic frequency ratios while other spectral components are clearly inharmonic ([32.185] and Sect. 30.2). Typically, complex bell sounds give rise to more than one spectral or virtual pitch [32.104, Chap. 11], [32.186], and these sounds often show marked AM due to interaction of narrowly spaced components. Spectral inharmonicity and spectral modulation are features also found in non-Western idiophones, in particular in Javanese and Balinese *gamelan* [32.85, 86, 209, 210]. The combination of pitch ambiguity, spectral inharmonicity and modulation, which is characteristic of sounds from bells, gong chimes and other metallophones, is not compatible with the additive (f_1 pitch + spectral sound color) concept sketched above for harmonic complexes. Rather, such sounds are sensed as spectrotemporal conglomerates that are not easily analyzable into constituents by ear even though musically experienced subjects can assign pitches to many sounds by singing or humming a tone (or several if they sense more than one pitch per sound). Also, subjects make comments on the *clangy* or *metallic* onset and the *shimmering* decay of such sounds, which thus can be described verbally in terms of timbral attributes. However, for many complex inharmonic sounds there is even less a demarcation between the perception of *pitch* and that of *timbre* than might exist for harmonic complexes.

The issue of whether pitch and timbre are two distinct or two interrelated qualities that subjects perceive when listening to sounds such as synthesized harmonic complexes or tones played on certain instruments has been discussed on the basis of empirical data [32.211–213], [32.104, Chaps. 10–12]. In regard to criteria elaborated by *Garner* [32.172] for *integral* and *separable* dimensions of perception, there seem to be indications for both points of view. From common experience, one could argue that musically trained subjects are capable of assigning two labels to musical tones presented in isolation, one denoting a pitch and the other denoting a timbre that is characteristic of a certain instrument or at least a type or a *family* of instruments. The two judgments implied in such a labeling task are *categorical* in that the stimuli have to be ordered into discrete categories. Such a task can be accomplished if the cues available to subjects intending to identify both the pitch (in terms of musical denominations, e.g., F_3, B_4) and the instrument type (e.g., tenor sax, trombone, cello, electric bass) are unambiguous and salient. However, in experiments on possible interactions of pitch, timbre, and loudness involving synthetic sounds, subjects seemed unable to attend to one dimension or attribute while disregarding the other (two experiments coupled timbre and loudness as well as timbre and pitch [32.211]). If the task is discrimination of small changes in either pitch (with respect to f_0) or timbre (defined by spectral centroid) of synthetic harmonic complexes, musically trained subjects showed smaller difference limens (DLs) for pitch than nonmusicians but were similar in their respective spectral centroid DLs [32.213]. In addition, performance differed significantly between congruent (changes in f_0 and spectral centroid were in the same direction) and noncongruent conditions (which, in natural sounds such as produced from wind and string instruments, is unlikely). While some experiments suggest pitch and timbre can be perceived as independent of each other [32.212], interference of pitch and timbre has been reported as well. A possible explanation for conflicting evidence may be sought in different experimental designs, stimuli, and tasks. If the stimuli are tones from familiar musical instruments (or synthesized tones close in timbre to natural sounds), in particular

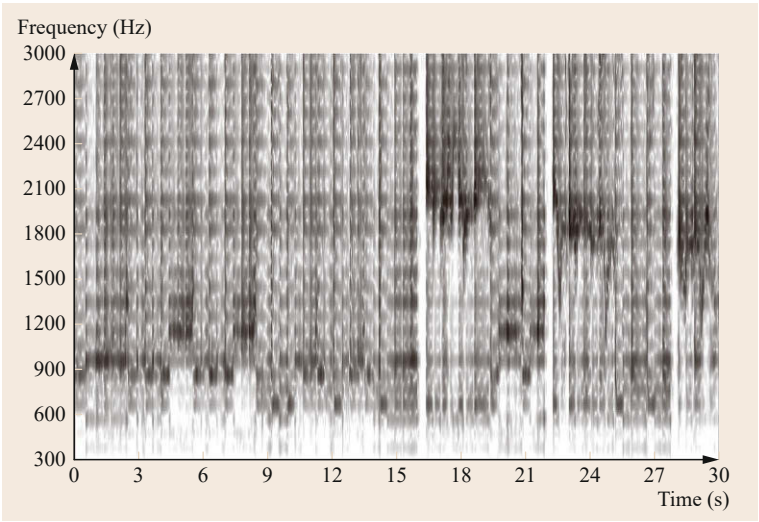


Fig. 32.11 Spectrogram 300 Hz – 3 kHz, Jew's harp melody, Opal Shuluu (Tuva)

musically trained subjects will have little difficulty in tracking both pitch changes and watching for changes in timbre (for example, shifts in centroid and brightness evoked through spectral filtering or by means of phase shifts between partials). Subjects thus can focus on one parameter at a time and quickly shift their attention between two parameters; such a strategy may be effective for essentially *independent* processing of pitch and timbre information. However, the structure of sound stimuli and of the sensory input used for processing has to be taken into account (Sects. 30.2, 31.1–31.6). For complex harmonic sounds such as radiated from aerophones and chordophones, BM filtering and auditory neural processing conveys pitch information both in place and in temporal code. Resolved partials and groups of unresolved partials all contribute to periodicity pitch while they also convey information concerning spectral energy distribution and spectral profile, which result in perception of sound color and the changes it may undergo in the course of presentation. Major changes in spectral structure and energy distribution can also have effects on perceived pitches. For instance, attenuation of the odd harmonics as well as phase shifts of partials in a complex can bring about a shift in perceived tone height by an octave or even several octaves while not affecting the *chroma* component of pitch ([32.214] and experiments reported in [32.215]).

The interdependence of pitch and timbre in harmonic complex tones is evident from music realized with instruments where a generator is put into the mouth that functions as resonator as well as with styles of vocal music where the mouth cavity serves as a bandpass filter. Such techniques can be observed in musical genres of various cultures where the Jew's harp (also: jaw's harp, trump) or the mouth bow are in use, or where styles of overtone singing are practiced. Figure 32.11 shows an excerpt of a melody played by Opal Shuluu (Tuva) on a Jew's harp [32.216, Track 31]. The spectrogram shows that each vertical sonority contains quite many components in the most relevant band (300 Hz–3 kHz), many of which are nearly harmonic and appear as multiples of a virtual pitch (autocorrelation, AC) at ≈ 96.1 Hz; the melody is filtered out by changes in the resonator (mouth cavity) so that certain components become more prominent in the spectral energy distribution per time frame. It is a typical mixture of melody against a drone, or, put in terms of Gestalt psychology [32.217, Chap. 7], of a figure against a ground.

With inharmonic sounds, separation of pitch and timbre is much more difficult since there is no strict periodicity relevant for f_0 pitch perception, and spectral structure may be also quite irregular. This hampers pitch perception for individual sounds from bells, gong chimes, or other metallophones [32.218]. Further, sequences of inharmonic sounds representing a scale or a melody may differ considerably in spectral energy distribution from one sound to the next as can be observed, for example, with bells in historic carillons. Spectrograms of sounds (cut to the initial 3 s) recorded from bells no. 1–3 of the Brugge carillon (Joris du Mery 1744, [32.186]) in Fig. 32.12 demonstrate that, notwithstanding essential components of a minor third bell that can be identified in all three segments, there is no homogeneous *sound color* since distribution of spectral energy varies markedly between these sounds. Further, the spectral component lowest in frequency (*arrows* in Fig. 32.12 marking the so-called *hum note*) in each of these three bell sounds is weak in intensity, which implies it will hardly elicit an individual spectral

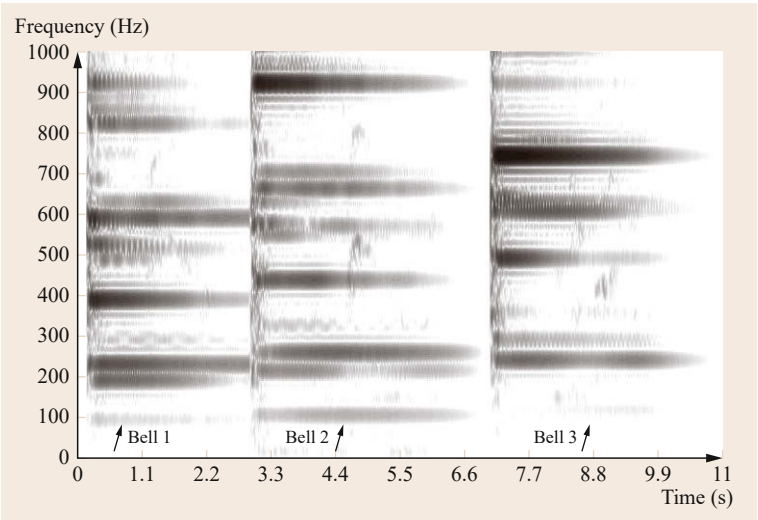


Fig. 32.12 Brugge carillon, sound segments from bells no. 1–3, spectral components 0–1 kHz, *fundamentals* (hum notes) marked by *arrows* are weak in these sounds

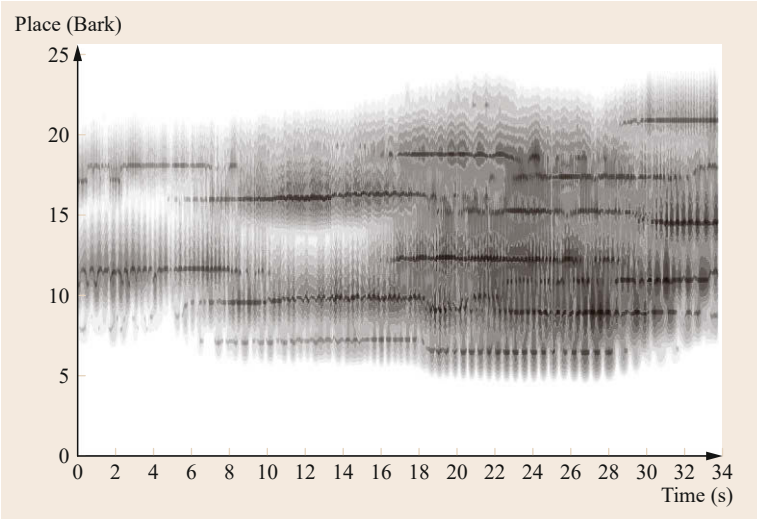


Fig. 32.13 John Chowning: excerpt from *Phoné*, left channel, cochleagram

pitch. Unlike the fundamental f_1 in a harmonic complex, which is reinforced by the sum of partials contributing to f_0 ($f_1 \equiv f_0$), virtual pitches in inharmonic complexes must not converge with the lowest spectral component. In bell sounds, the most prominent virtual pitch (the strike note) often is close to, but not identical with, the second spectral component. If this is fairly strong, it can be sensed as a spectral pitch besides the strike note. Pairs of adjacent spectral and virtual pitches cause ambiguity in perception.

The ambiguous nature of inharmonic sounds has been explored in compositions of computer music where textures of harmonic and inharmonic complexes are interwoven as in works of John Chowning who made use of the FM sound generation technique he had developed [32.59, 219]. In one such work, *Phoné* (1980/81), sequences of complex sounds, which at times resemble formant structures of the human singing voice, are distributed on two stereo channels (in the CD mix [32.220]). Though the listener can detect tonal elements in the complex sounds from both channels, spacing of spectral components is quite dense and energy distribution covers much of the audio bandwidth. A cochleagram (Fig. 32.13) shows an excerpt from *Phoné* (left channel only); one can see several strong components per time unit that are relevant in regard to spectral and/or virtual pitches interspersed with broad bands of energy. The perception resulting from both channels is a complex mixture of *timbral* elements with faint pitch structures in between. The vertical structure of these complex sonorities thus differs from traditional compositions where, typically, several *voices* can be

distinguished, which interact in polyphonic settings or combine into chord-like formations.

Electronic music based on analogue technology had already lifted the boundaries between *voices* and *timbre*. With digital technology, one can create complex sonorities of many kinds. Also, interpolation between various spectra chosen from natural sources can be effected (spectral interpolation is different from cross-fades between sounds as are available in most samplers, *Arfib* et al. [32.221]). The approach to sound generation based on digital FM opened new perspectives since, taking the basic concept of FM, that is, carrier and modulator frequencies, one can combine several modules that produce either carrier or modulator frequencies in a network that can generate arbitrary spectra, which again are modulated in frequency (or in phase) and in amplitude over time. Such sounds in fact can contain several inharmonic and/or harmonic complexes at any given time. Implementation of this concept became available in digital synthesizers suited to daily use in the studio, and even on stage, in 1983 (DX 7 and follow-up models like TX 802, TX 81Z). With digital FM sound generation one can set parameters so that octaves in a scale do have frequency ratios other than 2 : 1 (e.g., the *golden mean*, 1.618 : 1, which Chowning employed in *Stria*, 1976/77, [32.63, 219]). Further, FM technology enables composers to create scales, melodic sequences and chord-like sonorities that appear paradoxical in certain respects, and may be perceived like *auditory illusions* [32.222]. In compositions like *Stria* and *Phoné*, textures and mixtures of complex sounds may still evoke perceptions of *pitch* and *timbre*, however, the flow of sonic objects that are heard, moreover, moving in a 3-D space (the original version of *Stria* is quadraphonic), transcends traditional categories of *tone*, *pitch*, *voice* (in the sense of harmony and counterpoint), and *timbre*.

Composing with complex sounds rather than with musical tones is an approach realized in electronic and computer music as well as in such orchestral works where *clusters* of tones are equivalent to complex spectral structures. In regard to perception and musical syntax, an issue much debated is whether sequences of complex harmonic and/or inharmonic sounds may constitute a *scale* similar in function to a scale of pitches we perceive when hearing a sequence of harmonic complex tones. Some exploratory studies suggested a hierarchical organization of timbre similar to that of pitch would be feasible, at least in principle [32.164, 165, 223].

The issue whether sound sequences could be constructed in which shifts in timbre is the parameter equivalent to pitch shifts like in a musical scale, apparently was nourished from some sketchy ideas on the possibility of a *Klangfarbenmelodie* that *Schönberg* [32.224, p. 471] had added to his textbook of harmony. Schönberg seems to suggest that sounds might be varied in *sound color* so that identifiable sequences more or less analogous to pitch sequences would result. Schönberg did not go into detail except noting that, as he saw it, *Klangfarbe* was a more general concept comprising *Klanghöhe* as one dimension, referring to a sensation of relative height evoked by one complex sound when compared to another. His idea then was to vary *Klanghöhe* in such a way that sequences similar to a melody would be perceived. To be sure, *Klanghöhe* is not identical with *tone height* defined by linear frequency in two-componential models of pitch. A likely interpretation of Schönberg's short remarks is that *Klanghöhe* can be taken as equivalent to the spectral centroid. Consequently, operations on sounds that would shift the centroid up and down like on a pitch scale while maintaining the shape of the spectral envelope might be suited to create a *Klangfarbenmelodie*. Schönberg's own approach to this concept as manifest in his op. 16/III was that he had five tones in a complex sonority changing so as to produce noticeable changes in brightness over time [32.225]. Sensory brightness of a sound is largely dependent on the centroid resulting from spectral energy distribution.

If the spectral envelope would be identical for all the tones played by one particular instrument in different registers (see above), a near-constancy of sensational quality could be expected for listeners as the spectrum is virtually shifted up or down in frequency without changing the amplitude relations between partials. There have been considerations of how operations such as transposition and inversion (known from operations on melodic pitch sequences such as canons) could be applied to sounds with respect to the sensory and perceptual quality of *sound color* [32.177]. The problem, however, remains that shifts of the spectrum, while maintaining a more or less identical sound color, might not be perceived as such, yet rather as interacting with pitch structures since pitch, in particular for musically trained subjects, appears to be the more fundamental perceptual quality. Consider for example the sounds from 31 diapason pipes per octave as are available on the organ that has been built for the Huygens–Fokker tone system [32.226]. Playing the tones of this quite unusual scale, one after another (the difference in f_1 between adjacent pipes is 38.71 cent), musically trained listeners will perceive a sequence of tones distinctive in pitch and almost homogeneous in sound color. If one conceives of timbral sequences based on complex inharmonic FM sounds (see above) meant to constitute a *timbre scale*, it is likely that listeners with a musical background may still be inclined to infer pitch relations from such sequences though they may also perceive

a concept of *order* among sounds varying in certain timbral parameters.

Of course, in music and musical instruments there are interrelations between pitch and timbre in many respects. For example, mixture stops in pipe organs were introduced to expand the spectral width and to strengthen the brightness of organ sounds as well as to reinforce the pitch of tones that were played with fundamental stops (such as the diapason or Prinzipal). There is a combination of spectral and virtual pitch effects if the tuning of the keyboard matches that of pipes in mixture stops as close as possible. Also, interrelations of pitch and timbre have played a significant role in regard to musical composition and orchestration since composers knew from experience which instruments would *fuse* well in a homophonic texture and which instruments would be suited to support perception of individual voices in a polyphonic setting. Viewed from acoustics and psychoacoustics, there are factors such as formant-like concentration of spectral energy, partial spectral masking between instruments and sensation of roughness caused by spectral interaction in dissonant sonorities that need to be considered [32.181, 227]. Spectral fusion versus spectral roughness as well as emphasis of formants in singing styles are also relevant in vocal music, as for example folk music idioms can differ significantly in this respect [32.228].

32.2.5 Sound Segregation and Auditory Streaming

The performance of the auditory system in mammals is striking in several respects:

1. Sensory processing is fast and quite precise with respect to locating sound sources and extracting features from complex sounds.
2. The auditory system is capable of integrating related information into entities that become perceivable as *objects*.
3. The auditory system can distinguish between several concurrent sound sources and objects so that these can be identified and categorized accordingly.

This section will provide basics on hearing conditions in environments and will then survey some of the principles underlying formation of auditory objects, on the one hand, and their segregation when occurring simultaneously, on the other. Since in particular auditory stream segregation has been the subject of comprehensive monographs [32.217, 229], the following paragraphs will only cover some of the relevant points.

Listening to music in a concert hall or in front of an audio system (which may be stereophonic, quadraphonic, or ambiophonic) means a quasicontinuous stream of sound waves propagating through a medium (Sect. 30.3) reaches both ears of a subject who may try to identify sound sources and *objects* within this stream. In this respect, binaural hearing is the normal situation. Depending on the environment (which may be a concert hall, an open air concert, or one's living room), the ratio of sound energy emitted directly from the source and the sound reflected from hard surfaces may vary. In open spaces unbounded by reflecting surfaces, a free field condition prevails. If music is presented on stage in a concert hall, listeners typically have the orchestra or band in front so that most of the sound energy is transmitted directly, with a certain portion of lateral energy reflected from the side walls; in addition, energy might be reflected from a hard ceiling (for room acoustics and their effects on auditory perception, see [32.230–232]). In effect, in rooms bounded by hard surfaces, there is a mixture of direct sound and diffused sound between which a delay can be measured. The interaural cross-correlation [32.230, Chap. 3] between sound signals fed into both ears is a parameter suited to measure the degree of diffuseness. There are several more parameters (such as clarity, reverb time, coloration, distinctiveness or *definition*) that are of relevance for perception. For sound sensed binaurally in a free field (or in other spaces with negligible reverberation), mammals can use the interaural time difference (ITD) and the interaural level difference (ILD) as primary cues for spatial hearing and source localization [32.233, 234]. A model widely accepted is that the interaural time difference (ITD) is processed at the level of the inferior colliculus (IC), in pooled neurons [32.235]. Acuity is extremely high in that the just-noticeable difference (JND) for ITD seems to be close to 10 µs for a 500 Hz tone. In addition, the interaural level difference (ILD) serves as a cue for localization [32.236]. Though both ITD and ILD are restricted to certain conditions and ranges, taken together they can provide sufficient information to subjects for localizing sound sources. In regard to prerecorded music reproduced from audio systems, listeners are mostly confronted with a stereophonic setup where sources are *panned* on a left–right axis while 5.1 and other surround sound systems simulate a 360° panorama. Wave field synthesis [32.237] can even improve spatial representations of sources from prerecorded music that are perceived as if distributed in a natural 3-D environment.

Patterns of sound waves entering the auditory system binaurally can be extremely complex according to musical and physical parameters encoded. Consider, for example, performances of symphonic works rich in harmonic textures and instrumentation where also the dynamic range may vary considerably over time. Hence, listening to music requires fast processing

with sufficient resolution to allow for feature extraction and overall categorization of the input. Fast extraction of stimulus features necessary for sensory and motor responses in afferent-efferent feedback loops affords distributed processing along stations of the auditory pathway (AuP) (Sects. 31.2–31.4). As expected, there are indications that much of the processing relevant for pitch and other features is done subcortically, and in parallel up to the IC [32.238–240]. However, if processing is hierarchical and distributed, one would expect some neural network capable of integrating related information so that one perceives *objects* or even *complex wholes* and not just a bunch of features. This problem has been discussed quite extensively, in psychology and neuroscience, as one of *binding* [32.241, 242]. Though many studies on *binding* are concerned with visual perception as well as with language, formation of auditory objects also calls for a neural system suited to integrate spatial and temporal information gained from the various stages of analysis in the cochlea and along the ascending AuP. In the following, temporal and spectral criteria relevant for *fusion* as well as for *fission* of components in individual sounds and for fusion or segregation of concurrent sounds will be reviewed. Cues for identification of sound sources such as instruments and voices as well as for perceiving sonic objects embedded in a quasicontinuous stream of waves are temporal, spectral, and dynamic.

Fusion and *Fission* of Spectral Components in Individual Sounds

Evidently, we can hear a harmonic complex tone played on an aerophone or chordophone (e.g., oboe, trumpet, cello, sax) as a coherent whole notwithstanding that a number of low partials of such tones will be resolved in the BM filter bank, and also groups of higher partials will be segregated according to the CB auditory filter model [32.243–245] In a nonanalytic listening situation, individual harmonic partials and groups of partials falling into different CBs will be perceived as *fused* into one complex that, because harmonic partials join into a common period, gives rise to a distinct pitch and can convey a sensation of a certain timbre. The section of a sound that appears as *fused* into one object is the steady state while at the onset individual harmonics may be audible because, in particular in aerophones, some modes can reach a stable regime of vibration earlier than the bulk of modes making up the spectrum [32.204]. Of course, one can adopt an analytic stance and try to *hear out* some of the low partials in, for example, a harmonic complex comprising ten partials ($f_1 = 200$ Hz) with amplitudes rolling off at 3 dB/oct. Also, a nonharmonic component included in a harmonic spectrum most likely will be detected (if strong enough in level) as a separate component not fitting to the main body of partials constituting the perceptual object (a complex harmonic tone). Consider, for example, a harmonic spectrum where the third partial is detuned to a frequency ratio of 2.55 to $f_1 = 100$ Hz while the other components are in small integer frequency ratios. Using ten components with amplitudes decreasing at -3 dB/oct, the waveshape plotted in Fig. 32.14 results. The basic periodicity of 10 ms corresponding to f_1 as well as to f_0 resulting from partials 1, 2, 4–10 is still dominant; however, the inharmonic component obstructs a regular waveshape to repeat per period.

Separation in this case is effected by means of two concurrent pitch percepts, one based on the spectrum and periodicity (f_1 and f_0) of the harmonic complex, the other on the frequency and period of the inhar-

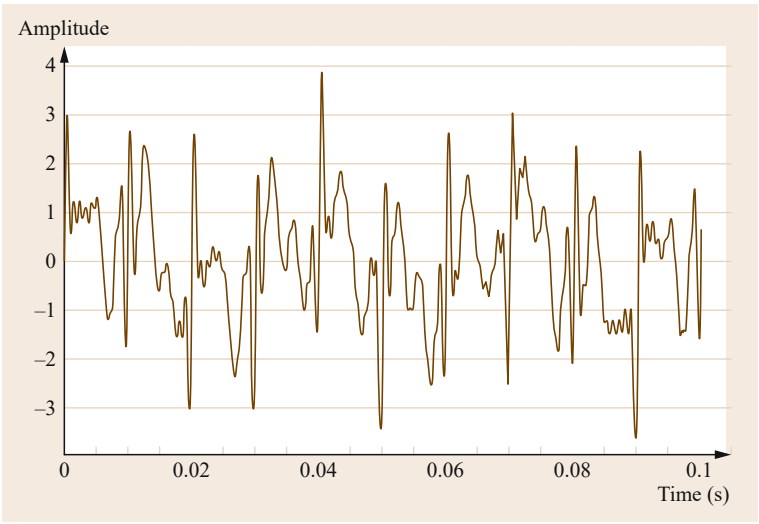


Fig. 32.14 Waveshape, harmonics 1, 2, 4–8 of $f_1 = 100$ Hz; partial no. 3 detuned to 255 Hz

monic component that *stands out* against the complex. Such a situation basically occurs with bell sounds where several partials may *fuse* quite well while in particular the minor third often is clearly detectable as a strong component (Fig. 32.12). Most sounds from carillon and swinging bells as well as such produced from gong chimes or by means of FM technique (see above) are ambiguous not because they would lack pitch and timbre information but because this information does not integrate easily into representing a single, coherent sound object. Rather, subjects in experiments tend to assign two or even more pitches to one complex inharmonic sound and in addition may find the timbre variable with time [32.85, 86, 209, 246–248].

Onset Synchrony versus Asynchrony of Concurrent Parts and Sounds in Music

Theoretical onset times and durations of notes in works of music can be calculated from notations in case a certain metronome value is prescribed. Absolute tempo (measured in beats per minute, BPM) in many recordings of pop music is evident from drum computer or sequencer tracks. In music performances not bound by notation and/or absolute tempo set by a machine, onset times of various instruments may be measured relative to some *time keeper* (for instance, a regular sequence of fast notes the drummer plays on the ride cymbal). Jazz and also rock musicians seem to have opinions of what may contribute to a good *groove* in a musical performance. One aspect often alluded to is that onsets of certain instruments should be slightly ahead or behind those of the time keeper. For instance, the bass may *drag* the tempo a tiny bit ahead of the time keeper while the snare drum is a tiny bit late, and then is perceived as somewhat *heavier* relative to the beat marked by a regular sequence of pulses (e.g., the attack of strokes on the ride cymbal). In such perceptions, two factors seem to be of relevance. One relates to temporal order and precision relative to an audible or imagined pulse, the other perhaps to effects of masking. If all onsets were synchronized so that sounds from different instruments in an ensemble would converge as much as possible, it would be difficult to distinguish their notes, in particular if instruments similar in sound color (see above) are playing notes in consonant chords. Of course, this is a situation wanted in certain musical contexts such as homophonic settings where maximum fusion of parts (as well as of partials) is desired. However, in polyphonic or multipart music, it is often necessary for listeners to follow individual parts (or *voices*) in order to apprehend musical structure as based on themes, motives, and voice-leading. For example, many works written for string or saxophone quartet require listeners capable of perceiving individual parts as well as the interplay of such parts in simultaneous chords or other vertical sonorities. Experimental data suggest that even musically trained subjects have difficulties in keeping track with multipart music if the number of voices exceeds three and when timbres for all voices or parts are relatively homogeneous [32.249]. Since timbre may not suffice to distinguish sources within *families* of instruments such as bowed strings or saxophones, and partial spectral masking can occur in particular if parts are relatively close in the pitches of their respective notes [32.181], temporal and dynamic factors come into play as a means to keep parts or voices apart. As has been reported by *Rasch* [32.182], subjects showed higher scores in correctly detecting the direction of intervals in quasisimultaneous concords formed of two complex tones when their onsets differed by 0–20 ms. *Rasch* [32.250] also tested the role of onset asynchrony in small ensembles where he found that those instruments that play the main melodic line tend to lead by about 30–50 ms. Onsets as in performed music seem to vary considerably relative to a grid one may calculate from notation, or may impose from a reference instrument. Onset asynchrony helps the listener to segregate voices in polyphonic music. In polyphonic keyboard works of the Baroque era (such as fugues written by D. Buxtehude, N. Bruhns, J.S. Bach), different voices (assigned to the right and left hand of the performer respectively) often do not begin aligned but with one voice starting on the beat and another kept apart by a quaver or semiquaver rest preceding the entry of that line.

Coordinated Modulation of Spectral Components as Marker of a Source

There are many reports on the effects modulation has on source segregation and identification. In particular, coordinated modulation of spectral components so that their frequencies vary in parallel has been stressed [32.251]. In experiments with synthesized vowels presented in combinations at different pitches, subjects judged the prominence of a modulated target vowel higher than unmodulated vowels [32.252]. In a musical situation such as when a solo violinist is backed by an orchestra in a violin concerto, there might neither be much difference in averaged (root mean square, rms) sound level between the solo violin and the orchestra nor in the timbre of the solo violin as compared to the string section of the orchestra. Parameters suited to effect acoustic and auditory segregation of the solo violin against the orchestra then can be strong onset attacks in nonlegato phrases and the use of substantial vibrato in legato phrases comprising long-held notes. In fact, this is observed in many performances. Detection of modulation has been found an efficient cue for computerized scene analysis [32.253].

Segregation of Harmonic Compounds as in Single Consonant Chords

As has been explained in Sect. 31.6.4, there are certain conditions for perceiving harmonic complex tones and combinations of such tones in terms of consonance, fusion, and *Verschmelzung*, which means sounds such as Stumpf's *ideal concord* (Fig. 31.15) are perceived as highly coherent while being also apprehended as a configuration of tones related by certain intervals. In order to apprehend structural relations between several pure or complex tones, a listener should be able to segregate a chord or other harmonic compound into its constituents. This requires that a sound is present for a certain time, and that the listener has some experience in analytic listening. The task is to find how many pure or complex tones are contained in a chord or other sonority, and to identify the relations between the components. Consider for example the final chord in a work of organ music like the *Praeambulum primi toni a 5 in d* by Matthias Weckmann where five voices join into a long-held chord comprising the notes D_2, D_3, A_3, $F\#_4$, A_4, D_5 (the note D appears tripled to emphasize the key of the piece). Since works like this *Praeambulum* are usually played with several stops at $16'$, $8'$ and probably also at $4'$ and $2'$, spectral energy relevant for pitch and timbre perception as contained in many partials should cover a frequency range from D_1 to at least three octaves above D_5, that is, ≈ 4.7 kHz. In the recording used here for analysis (Wilde–Schnitger organ of 1683, Lüdingworth near Cuxhaven), one of the stops is a $16'$ dulcian reed pipe that produces very many harmonic partials per tone. Figure 32.15 shows a spectrogram 0–2 kHz and the f_0 of the chord derived from AC and subharmonic summation (SHS) analyses as well as the output of a Bark filter analysis with the center frequencies spaced closer (ratio 0.85) than in the usual Bark scale to emulate CB bandwidths as observed in auditory peripheral filtering. These data-driven analyses demonstrate that segregation of constituents according to features (pitch-relevant partials, other harmonics, onsets, temporal fluctuations of spectral components in frequency and amplitude) is possible, at least to a certain extent.

Hearing the chord live in the church, one may try to use spatial information since the pipes sounding the chord are distributed in the organ (with the pedal stops housed in two towers flanking the main organ case). In regard to pitches, which are often the strongest cue for auditory analysis, one can hear the *fundamental* at $D_1 \sim 38.8$ Hz (the organ is tuned to $A_4 \sim 466.3$ Hz), which is the f_1 spectral component of two of the $16'$ pipes employed as well as a periodicity pitch (f_0) that is confirmed by both AC and SHS analyses (Fig. 32.15). Of the remaining tones, several can be identified by their pitch intervals relative to D_1, especially A_3, $F\#_4$, and D_5. Neglecting small fluctuations in frequency and amplitude (Fig. 32.15), the D-major chord to one's ear offers a high degree of *Verschmelzung* notwithstanding small tuning deficiencies (the organ is in meantone tuning, implying the $F\#_4$ is a just major third but the A_3 and A_4 are flat by 5.5 cent respectively). Auditory analysis in this case is not too difficult because the chord lasts for almost 5.5 s.

Segregation of Auditory Streams

In 1649 and 1654, *Jacob van Eyck*, a famous Dutch carillonist and flutist, published two volumes of a collection of some 150 tunes, many of which were part of the popular repertoire of his time [32.254]. These tunes, which can be performed with a single soprano

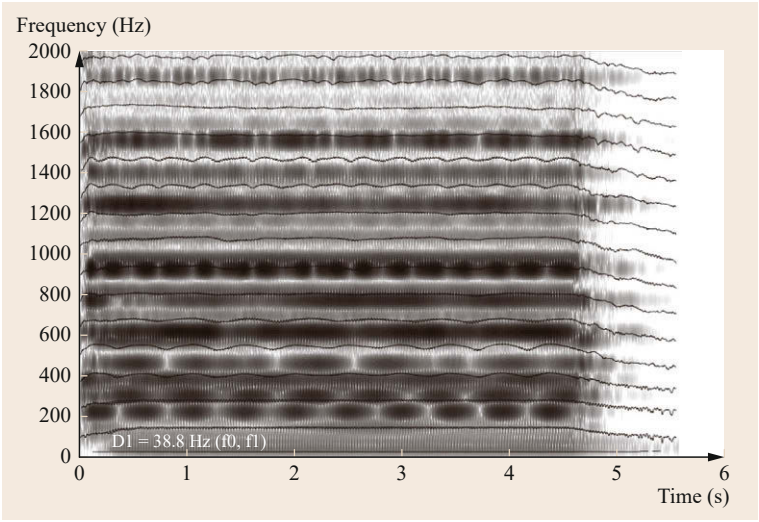


Fig. 32.15 M. Weckmann, *Praeambulum*, final D-chord, spectrogram 0–2 kHz, AC and SHS pitch analysis, Bark filter analysis 0–2 kHz (15 bands), extraction of f_0 (identical with f_1), detection of strong components and of (frequency, amplitude) modulation effects

recorder, are elaborated in variations, which include embellishments such as figurative elements as well as short notes inserted into the somewhat longer notes of the tune. In effect, if played at a lively or even fast speed by a skilled flutist (as van Eyck was), the impression on a listener is that the music, though not truly polyphonic (which would rather call for two or more independent voices), is not monophonic either. The type of setting found in van Eyck's anthology has been labeled pseudopolyphony; it was a technique employed also by J.S. Bach in his works for violin solo and cello solo (BWV 1001-1006; 1007-1012) where, however, simultaneous intervals and even full chords are included as these are playable on a bowed string instrument that offers four strings. In Fig. 32.16, a small section from the performance of one of the tunes adapted by van Eyck (*Wat zalmen op den Avond doen? Van Eyck* [32.254]; Marion Verbruggen, recorder) is shown, which demonstrates the style of a pseudopolyphony where two voices played on a single recorder seem to interlock in time.

As Fig. 32.16 demonstrates, most of the main notes making up the tune are in a higher register, and are played with more force than the very short notes that fall between them. Thus, there are two frequency regions divided at ≈ 950 Hz; the very first note/tone (~ 841 Hz, $A_5^{\flat}$) can be related to both the upper and the lower stream, which are divided by about an octave (tone no. 2 is at 1109.3 Hz, $D_6^{\flat}$, tone no. 3 is at 552.4 Hz, $D_5^{\flat}$, etc.).

The perception of pseudopolyphonic structures such as found in works of van Eyck and Bach depends on several parameters. One is the average interval between the upper and the lower sequence of tones, another is the tempo (measured in MM or BPM) of the performance, which determines the number of events per time unit (event density) as well as the relative duration of tones (e.g., quavers, semiquavers) to be played in a phrase. Musically trained listeners will be able to follow both the tune melody and the up-and-down motion of pitches installed between notes/tones of the upper and the lower register even if their duration can be quite short ($t \leq 200$ ms). If the tempo increases further (which can be done by digital time compression without affecting pitch) to 150% of the original version, it is more difficult to follow the pattern of up-and-down intervals. Doubling the tempo, and hence the presentation rate, hampers interval perception and leads to a percept where the tune is still present but the tones in between no longer join into kind of a *counter melody* (as is indicated by tones in the lower region of Fig. 32.16). Also the rhythmic pattern seems to have changed to a *galloping* type as has been described, for certain tone sequences, by *van Noorden* [32.255] and *Bregman* [32.229]. Of course, such a tempo would not suit any musical performance of this style of music. However, there are idioms such as the Kiganda Amadinda xylophone music of former Buganda [32.256, 257] where basically two melodic sequences are played simultaneously, in isochronous pulses, with no metric accent, at a very high speed (eighthnote $\sim$ 520–600 MM). Each sequence is of a certain length (for instance, 36 notes), and is repeated over and over. Unless one is familiar with Luganda language and the concept behind the music, it is not possible to apprehend individual sequences (which are mostly derived from songs that can be narrative in character). Rather, in particular *Western* listeners tend to perceive a number of patterns that have been explained as *inherent* in the fast tone sequences one actually hears [32.256, 258]. In this respect, such

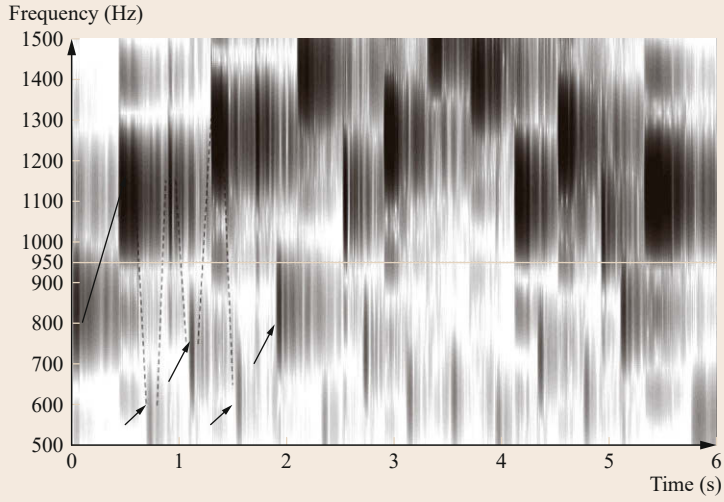


Fig. 32.16 Pseudopolyphony (after [32.254]); performance on a single recorder. The *beige line* at 950 Hz segregates two melodic streams. The tones in the lower stream (of which the first four are marked with *arrows*) are very short. The *solid line* connects the first two tones as were played and the *dashed lines* connect the tones 2, 3, 4, 5, 6, 7 in the up-and-down sequence. The decay in the tail end of the tones results from reverberation in the recording as well as from the filter response in the analysis

patterns have an objective basis. However, the Gestalt-like formations listeners perceive or imagine result from perceptual and cognitive structuring that leads to both segregation of tones within sequences into high and low pitch *streams* and to recombinations of elements into melorhythmic patterns. Since there is no objective rhythm implemented in isochronous pulse sequences produced without metric accents, the rhythmic grouping listeners realize seems to be derived from melodic patterns. The patterns listeners believe to perceive can be viewed as emerging from the very fast sequences, in an effort to group elements into coherent, musically probable formations.

The facts concerning pseudopolyphony and perception of inherent or emergent patterns briefly summarized in this section have been tested, on a much more fundamental level in a lab situation, in experiments on auditory stream segregation. An early experiment [32.259] found that, in the human hearing range up to ≈ 4 kHz, two pure tones A and B presented binaurally, each lasting 80 ms (plus 20 ms rise and decay time), appeared as a quasicontinuous up-and-down movement when their frequency difference df/f was within $\approx 15\%$ but turned into an interrupted two-tone pattern (A:B:A:B...) when the difference was larger. Seeing that the repetition rate for a pair A:B in this experiment is ≈ 5 Hz; the *trill threshold* – as the critical frequency difference was called – for 500 Hz would be close to 75 Hz, which equals 242 cent; as one might expect, this is about the width of a CB. *Bregman* and *Campbell* [32.260] reported two experiments one of which had six sine tones (2.5, 2, 1.6 kHz, 550, 430, 350 Hz) each lasting for 100 ms as stimuli. These tones were arranged in two sequences of high and low. With short duration of each tone, and high repetition rate of the sequences, subjects grouped the tones into two streams representing high and low tones. *Van Noorden* [32.255] conducted several experiments on sequential coherence of tones (as in melodies) and found that stream segregation depends on:

- The interval in pitch between two tones.
- The repetition rate. With increasing repetition rate, the interval width necessary for segregating tones at two different pitch levels into streams decreases.
- The intensity level differences between tones presented with alternating loudness, which can have an effect on perceptual grouping (termed *roll effect* by *van Noorden* [32.255]).

In addition, differences in timbre have been reported as a factor relevant in forming streams [32.261]. Finally, manipulating the phases of complexes of unresolved harmonics can lead to sequences of sounds that do not differ in power spectrum yet appear different in perception; such differences in perceptual quality may suffice for also inducing segregation [32.262].

Auditory stream segregation apparently is based on data-driven bottom-up analysis that starts at the auditory periphery. However, there are also top-down processes since segregation draws on learned schemata and on musical experience in general. It is of interest to note that quite many observations on auditory stream segregation can be viewed in regard to principles of Gestalt perception such as temporal and/or spatial *nearness* or proximity of elements, their *good continuation* as in a melody, the *common fate* partials of a harmonic complex share in FM, the *conciseness* of a certain rhythmic or tonal pattern, or its *closure*, etc. (a list of more than 100 principles of Gestalt perception was assembled by *Helson* [32.263]). A detailed discussion of experimental findings and an interpretation of many observations in terms of Gestalt perception and cognitive psychology will be found in [32.217] and [32.229]. Parallel to behavioral experiments, modeling of auditory stream segregation and development of algorithms for automated DSP-based analysis have been pursued. Though there are different approaches, many are based on peripheral auditory filtering of acoustic input and on emulating stages of neural processing [32.253, 264–267]. Separation of sources and assignment of sonic objects to *streams* or *voices* can be achieved by means of DSP algorithms operating on digitized sound files if sufficient information can be gathered from the cues named above (different onset times, different pitches and spectral patterns of concurrent sounds, identification of harmonic partials undergoing *common fate* FM as a marker of a common source, etc.). Another aspect that relates to auditory stream segregation and to auditory *scene analysis* in general is transcription of polyphonic music into conventional staff or other graphic notation [32.268–270].

Summing up this chapter on *timbre*, the term can be used to denote the spectral plus temporal features of a sound sensed by subjects as a tone quality that, however, may include dynamic changes. For example, the tone quality of a clangy metal gong sound resides largely in the modulation while the tone quality of a violin may be more dependent on spectral formant structure (and, hence, on *tone color*). In recent decades, it has been customary to address the phenomenal appearance of temporal and spectral sound features as they interrelate in sonic objects simply under the umbrella of *sound*. This is a term that has gained importance in particular in the production and perception of pop and rock music. *Sound*, in this respect, comprises various aspects having to do with the localization of sound sources in a stereo, quadraphonic or surround-sound mix as well as with the *depth* of space conveyed by the ratio of

the direct signal and (natural or artificial) reverberation. Original sound sources, in addition, can be modified by means of filters and compressors, and can be modulated in many ways. Characteristic of *sound* in pop and rock music genres is the use of special effects such as phasing, flanging, chorus, modulation of filter and spectral envelope parameters, cross-modulation between several sources, etc. applied to single or to groups of instruments (see articles by Dutilleux and Zölzer and by Evangelista in *Zölzer* [32.66], also [32.271]). Further, *sound* in pop, rock and also jazz music involves playing techniques (as is evident from rock guitar playing where one uses string bending, finger vibrato, so-called claw hammering, etc.) and also tuning of instruments. Note, for instance, the electric guitar on Pops Staples' *World in Motion* (Virgin, 1992) tuned low to C_2 instead of E_2. Moreover, the guitar apparently was played with an amp using a tremolo effect (often wrongly labeled *vibrato* in amp literature) and a 15″ loudspeaker to reinforce low frequency response and to make the sound appear *big* in volume. Further, some reverb has been put on the guitar (either in the amp or, more likely, in the mix). The *sound* resulting thus is a combination of the notes played plus the tuning of the guitar and the frequency response and other specifications of the technical devices used in the performance and recording of the song in question. *Sound*, in this respect, is a highly dynamical, time-variant construct that has objective physical and musical foundations. The actual interplay of temporal, spectral, spatial and dynamic features may undergo many changes in the course of a relatively short piece of music, which listeners may attend to; consequently, they will perceive *sound* as a dynamic process. However, one can also abstract the more recurrent and (within limits) more or less invariant features that are regarded *typical* of a certain sound (e.g., the types of distortion and of feedback in guitar sounds used in genres of hard rock productions).

References

32.1 D. Muzzulini: *Genealogie der Klangfarbe* (P. Lang, Bern 2006)

32.2 E. Dolan: *The Orchestral Revolution. Haydn and the Technologies of Timbre* (Cambridge Univ. Press, Cambridge 2013)

32.3 F.A. Gevaert: *Nouveau Traité d'instrumentation* (Lemoine, Bruxelles, Paris 1885)

32.4 H. Berlioz: *Traité d'instrumentation et d'orchestration* (Lemoine, Bruxelles, Paris 1904)

32.5 V. Mahillon: *Catalogue descriptif et analytique du Musée instrumental du Conservatoire royal de Bruxelles*, Annuaire du Conservatoire (Annoot-Braeckman, Gand 1878)

32.6 V. Mahillon: *Catalogue descriptif et analytique du Musée instrumental du Conservatoire royal de Bruxelles*, Annuaire du Conservatoire (Annoot-Braeckman, Gand 1879)

32.7 V. Mahillon: *Catalogue descriptif et analytique du Musée instrumental du Conservatoire royal de Bruxelles*, Annuaire du Conservatoire (Annoot-Braeckman, Gand 1880)

32.8 V. Mahillon: *Catalogue descriptif et analytique du Musée instrumental du Conservatoire royal de Bruxelles*, Annuaire du Conservatoire (Annoot-Braeckman, Gand 1881)

32.9 E. von Hornbostel, C. Sachs: Systematik der Musikinstrumente, Z. Ethnol. **46**, 553–590 (1914)

32.10 P. Morse, K.U. Ingard: *Theoretical Acoustics* (Princeton Univ. Press, Princeton 1986)

32.11 N. Fletcher, T. Rossing: *The Physics of Musical Instruments*, 2nd edn. (AIP/Springer, New York 1998)

32.12 C. Forsyth: *Orchestration*, 2nd edn. (Macmillan, London 1948)

32.13 B. Lewis (Ed.): *Bioacoustics. A Comparative Approach* (Academic, London 1983)

32.14 G. Manley, A.N. Popper, R. Fay (Eds.): *Evolution of the Vertebrate Auditory System* (Springer, New York 2004)

32.15 R.G. Busnel (Ed.): *Acoustic Behavior of Animals* (Elsevier, Amsterdam 1963)

32.16 G. Tembrock: *Biokommunikation. Informationsübertragung im biologischen Bereich*, Vol. 2 (Akademie-Verlag, Berlin 1971)

32.17 G. Witzany (Ed.): *Biocommunication of Animals* (Springer, Dordrecht 2014)

32.18 M. Konishi: Birdsong: From behaviour to neuron, Annual Rev. Neurosci. **8**, 125–170 (1985)

32.19 M. Naguib, K. Riebel: Singing in space and time: The biology of birdsong. In: *Biocommunication of Animals*, ed. by G. Witzany (Springer, Dordrecht 2014) pp. 233–247

32.20 L. Sayigh: Cetacean acoustic communication. In: *Biocommunication of Animals*, ed. by G. Witzany (Springer, Dordrecht 2014) pp. 275–297

32.21 J. Sundberg: *The Science of the Singing Voice* (Northern Illinois Univ. Press, DeKalb 1988)

32.22 N. Fletcher: Bird Song – A quantitative acoustic model, J. Theor. Biol. **135**, 455–481 (1988)

32.23 N. Fletcher: *Acoustic Systems in Biology* (Oxford Univ. Press, New York 1992)

32.24 G. Fant: *Acoustic Theory of Speech Production*, 2nd edn. (Mouton, The Hague 1970)

32.25 R. Bader: *Nonlinearities and Synchronization in Musical Acoustics and Music Psychology* (Springer, Berlin 2013)

32.26 A. Benade: On woodwind instrument bores, J. Acoust. Soc. Am. **31**, 137–146 (1959)

32.27 J. Roederer: *The Physics and Psychophysics of Music: An Introduction*, 3rd edn. (Springer, New York 1995)

32.28 J. Beauchamp: Analysis and synthesis of musical instrument sounds. In: *Analysis, Synthesis, and Perception of Musical Sounds*, ed. by J. Beauchamp (Springer, New York 2007) pp. 1–89

32.29 J. Fricke: Formantbildende Impulsfolgen bei Blasinstrumenten. In: *Fortschr. Akust. 4. Jahrestag. Akust. (DAGA), Braunschweig* (1975) pp. 407–411

32.30 W. Voigt: *Untersuchungen zur Formantbildung in Klängen von Fagott und Dulzianen* (Bosse, Regensburg 1975)

32.31 C. Reuter: *Der Einschwingvorgang nichtperkussiver Musikinstrumente* (P. Lang, Frankfurt am Main 1995)

32.32 E. Meyer, D. Guicking: *Schwingungslehre* (Vieweg, Braunschweig 1974)

32.33 S.L. Marple: *Digital Spectral Analysis. With applications* (Prentice-Hall, Englewood Cliffs 1987)

32.34 A. Schneider: Change and continuity in sound analysis: A review of concepts in regard to musical acoustics, music perception, and transcription. In: *Sound – Perception – Performance*, Current Research in Systematic Musicology, ed. by R. Bader (Springer, Cham 2013) pp. 71–111

32.35 A. Beurmann, A. Schneider, E. Lauer: Klanguntersuchungen an der Arp-Schnitger-Orgel zu St. Jacobi, Hamburg, Syst. Musikwiss. – Syst. Musicol. **6**, 151–187 (1998)

32.36 A. Beurmann, A. Schneider: Acoustics of the harpsichord: A case study. In: *Systematic and Comparative Musicology: Concepts, Methods, Findings*, ed. by A. Schneider (P. Lang, Frankfurt am Main 2008) pp. 241–263

32.37 D. Arfib, F. Keiler, U. Zölzer: Source-filter processing. In: *DAFX. Digital Audio Effects*, ed. by U. Zölzer (Wiley, Chichester 1996) pp. 299–372

32.38 X. Rodet, D. Schwarz: Spectral envelopes and additive + residual analysis/synthesis. In: *Analysis, Synthesis, and Perception of Musical Sounds*, ed. by J. Beauchamp (Springer, New York 2007) pp. 175–227

32.39 S. Tempelaars: *Signal processing, Speech and Music* (Swets Zeitlinger, Lisse 1996)

32.40 W. Hartmann: *Signals, Sound, and Sensation* (AIP/Springer, New York 1998)

32.41 O. Elschek: *Fujara. The Slovak Queen of European Flutes* (Music Centre, Bratislava 2006)

32.42 A. Beurmann, A. Schneider: Some Observations from a Stein-Conrad Hammerflügel from 1793. In: *Systematic Musicology: Empirical and Theoretical Studies*, ed. by A. Schneider (P. Lang, Frankfurt/M. 2011) pp. 175–184

32.43 P. Holmes: The Scandinavian bronze lurs. In: *The Bronze Lurs: 2nd Conference on the ICTM Study Group on Music Archaeology*, Vol. II, ed. by C. Lund (R. Swedish Acad. of Music, Stockholm 1986) pp. 51–125

32.44 L. von Falkenhausen: *Suspended Music. Chimebells in the culture of bronze age China* (Univ. of Cal. Press, Berkeley 1993)

32.45 M. Liang: *Music of the Billion. An Introduction to Chinese Musical Culture* (Heinrichshofen, New York 1985)

32.46 A. Barker: *Greek Musical Writings, Vol. 2: Harmonic and Acoustic Theory* (Cambridge Univ. Press, Cambridge 1989)

32.47 G. Wille: *Musica Romana. Die Musik im Leben der Römer* (Schippers, Amsterdam 1967)

32.48 M. Bröcker: *Die Drehleier. Ihr Bau und ihre Geschichte*, 2nd edn. (Verlag für Systematische Musikwissenschaft, Bonn 1977), 2 Vols.

32.49 H. Klotz: *Über die Orgelkunst der Gotik, der Renaissance und des Barock*, 2nd edn. (Bärenreiter, Kassel 1975)

32.50 M. Praetorius: Syntagma musicorum II: De Organographia (Holwein, Wolffenbüttel 1619)

32.51 C. Reuter: *Klangfarbe und Instrumentation* (P. Lang, Frankfurt am Main 2002)

32.52 J. Meyer: *Akustik und musikalische Aufführungspraxis*, 5th edn. (Bochinsky, Frankfurt am Main 2004)

32.53 W. Meyer-Eppler: *Elektrische Klangerzeugung. Elektronische Musik und synthetische Sprache* (Dümmler, Bonn 1949)

32.54 D. Ernst: *The Evolution of Electronic Music* (Schirmer, New York 1977)

32.55 Th Wells: *The Technique of Electronic Music*, 2nd edn. (Schirmer, New York 1981)

32.56 T. Holmes: *Electronic and experimental Music*, 4th edn. (Routledge, New York 2012)

32.57 A. Strange: *Electronic Music. Systems, Techniques, and Controls*, 2nd edn. (Brown, Dubuque 1983)

32.58 J. Watkinson: *The Art of Digital Audio* (Focal, London, Boston 1989)

32.59 J. Chowning: The synthesis of complex audio spectra by means of frequency modulation, J. Audio Eng. Soc. **21**, 526–534 (1973)

32.60 J. Chowning: The synthesis of complex audio spectra by means of frequency modulation, Comput. Music J. **1**, 46–54 (1977)

32.61 C. Roads, J. Strawn (Eds.): *Foundations of Computer Music* (MIT Press, Cambridge 1985) pp. 6–29

32.62 J. Chowning: Stria (1977), CD (Wergo/Schott, Mainz 1988)

32.63 B. Bossis: *Stria* de John Chowning ou *l'oxymoron* musical: du nombre d'or comme poétique. In: *John Chowning, Portraits polychromes*, ed. by M. de Maule (Éd. TUM, Paris 2005) pp. 87–113

32.64 P. Schaeffer: *La Musique Concrète*, 2nd edn. (Presses Univ. de France, Paris 1973)

32.65 P. Schaeffer: *Traité des objets musicaux* (Seuil, Paris 1966)

32.66 U. Zölzer (Ed.): *DAFX – Digital Audio Effects* (Wiley, Chichester 2002)

32.67 The Ronettes: Be My Baby (Philles Records, Los Angeles 1963)

32.68 H.F. Cohen: *Quantifying Music. The Science of Music at the First Stage of the Scientific Revolution* (Reidel, Dordrecht 1984) pp. 1580–1650

32.69 S. Dostrovsky, R. Cannon: Entstehung der musikalischen Akustik (1600–1750). In: *Hören, Messen und Rechnen in der frühen Neuzeit*, Geschichte der Musiktheorie, Vol. 6, ed. by F. Zaminer (Wissenschaftliche Buchgesellschaft,

Darmstadt 1987) pp. 7–79
32.70 J. Sauveur: *Collected Writings on Musical Acoustics (Paris 1700–1713)* (Diapason, Utrecht 1984), ed. by R. Rasch
32.71 J.P. Rameau: *Démonstration du principe de l'harmonie* (Pissot/Durand, Paris 1750)
32.72 F. Chladni: *Die Akustik*, 2nd edn. (Breitkopf Haertel, Leipzig 1830)
32.73 F. Opelt: *Allgemeine Theorie der Musik auf den Rhythmus der Klangwellenpulse gegründet* (Barth, Leipzig 1852)
32.74 E. Boring: *Sensation and Perception in the History of experimental Psychology* (Appleton-Century-Crofts, New York 1942)
32.75 R. Beyer: *Sound of our Time. Two hundred Years of Acoustics* (Springer/AIP, New York 1999)
32.76 C. Seashore: *The Present Status of Research in the Psychology of Music at the University of Iowa* (Univ. of Iowa Press, Iowa City 1928)
32.77 M. Metfessel: *Phonophotography in Folk Music. American Negro Songs in new notation* (Univ. of North Carolina Press, Chapel Hill 1928)
32.78 E. Meyer, G. Buchmann: Die Klangspektren der Musikinstrumente. In: *Sitzungsber. Preuss. Akad. Wiss., Math.-Phys. Kl*, Vol. XXXII (1931) pp. 735–778
32.79 H. Backhaus: Über die Bedeutung der Ausgleichsvorgänge in der Musik, Z. techn. Phys. **13**, 31–46 (1932)
32.80 F. Trendelenburg, E. Thienhaus, E. Franz: Klangeinsätze an der Orgel, Akust. Z. **1**, 59–76 (1936)
32.81 W. Graf: *Vergleichende Musikwissenschaft* (Stiglmayr, Wien-Föhrenau 1980)
32.82 F. Födermayr: *Zur gesanglichen Stimmgebung in der außereuropäischen Musik*, Vol. 1 and 2 (Stiglmayr, Wien-Föhrenau 1971)
32.83 R. Randall: *Frequency Analysis*, 3rd edn. (Bruel Kjaer, Naerum 1987)
32.84 R. McAulay, T. Quatieri: Speech analsis/synthesis based on sinusoidal representation, IEEE Trans. on Acoustics, Speech, and Signal Processing **34**, 744–754 (1986)
32.85 A. Schneider: *Tonhöhe – Skala – Klang. Akustische, tonometrische und psychoakustische Studien auf vergleichender Grundlage* (Orpheus, Bonn 1997)
32.86 A. Schneider: Inharmonic sounds: Implications as to pitch, timbre, and consonance, J. New Music Res. **29**, 275–301 (2000)
32.87 G. de Poli, A. Piccialli, C. Roads (Eds.): *Representations of Musical Signals* (MIT Press, Cambridge 1991)
32.88 C. Roads, S. Pope, A. Piccialli, G. de Poli (Eds.): *Musical Signal Processing* (Swets Zeitlinger, Lisse, Abingdon 1997)
32.89 G. Peeters, B. Giordano, P. Susini, N. Misdariis, St McAdams: The timbre toolbox: Extracting audio descriptors from musical signals, J. Acoust. Soc. Am. **130**, 2902–2916 (2011)
32.90 H. von Helmholtz: *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (Vieweg, Braunschweig 1863), 3rd edn. 1870, 6th edn. 1913
32.91 C. Stumpf: *Tonpsychologie*, Vol. 2 (Barth, Leipzig 1890)
32.92 C. Stumpf: *Die Sprachlaute* (Springer, Berlin 1926)
32.93 W. Köhler: Akustische Untersuchungen I–III, Z. Psychol. **54**, 241–289 (1910)
32.94 W. Köhler: Akustische Untersuchungen I–III, Z. Psychol. **58**, 59–140 (1911)
32.95 W. Köhler: Akustische Untersuchungen I–III, Z. Psychol. **64**, 241–289 (1913)
32.96 W. Köhler: Akustische Untersuchungen I–III, Z. Psychol. **72**, 1–192 (1915)
32.97 C. Stumpf: *Konsonanz und Dissonanz* (Barth, Leipzig 1898)
32.98 C. Stumpf: Beobachtungen über Kombinationstöne, Z. Psychol. **55**, 1–142 (1910)
32.99 C. Stumpf: Über neuere Untersuchungen zur Tonlehre, Beitr. Akust. Musikwiss. **8**, 305–344 (1914)
32.100 E. von Hornbostel: Psychologie der Gehörserscheinungen. In: *Handbuch der normalen und pathol. Physiol.*, Vol. 11, ed. by A. Bethe (Springer, Berlin 1926) pp. 701–730
32.101 G. Rich: A preliminary study of tonal volume, J. Exp. Psych. **1**, 13–22 (1916)
32.102 S. Stevens: The volume and intensity of tones, Am. J. Psych. **46**, 397–408 (1934)
32.103 S. Stevens: The attributes of tones, Proc. Natl. Acad. Sci. **20**, 457–459 (1934)
32.104 E. Terhardt: *Akustische Kommunikation* (Springer, Berlin 1998)
32.105 W. Lichte: Attributes of complex tones, J. Exp. Psychol. **28**, 455–480 (1941)
32.106 G. Albersheim: *Zur Psychologie der Ton- und Klangeigenschaften unter Berücksichtigung der Zweikomponententheorie und der Vokalsystematik* (Heitz, Leipzig, Straßburg 1939), repr. Körner, Baden-Baden 1975
32.107 G. Rich: A Study of tonal attributes, Am. J. Psychol. **30**, 121–164 (1919)
32.108 C. Ruckmick: A new classification of tonal qualities, Psych. Rev. **36**, 172–180 (1929)
32.109 A. Wellek: Die Mehrseitigkeit der „Tonhöhe" als Schlüssel zur Systematik der musikalischen Erscheinungen, Z. Psychol. **134**, 302–348 (1935)
32.110 H. Ebbinghaus: *Grundzüge der Psychologie*, Vol. 1, 4th edn. (Veit, Leipzig 1919)
32.111 ASA: *American Standard Acoustical Terminology* (ASA, New York 1960) p. 45
32.112 F. Kittler: *Gramophone, Film, Typewriter* (Stanford Univ. Press, Stanford 1999)
32.113 Elektronische Musik: *Sonderheft über elektronische Musik, Technische Hausmitteilungen des Nordwestdeutschen Rundfunks*, Jg. 6, Nr. 1/2 (NWDR, Cologne 1954)
32.114 A. Moles: *Théorie de l'information et perception ésthetique* (Flammarion, Paris 1958)
32.115 D. Gabor: Theory of communication, J. Inst. Electr. Eng. **93**, 429–457 (1946)
32.116 C. Shannon: A mathematical theory of communication, Bell System Techn. J. **27**(379–423), 623–656 (1948)

32.117 L. Russolo: *L'arte dei rumori* (Ed. Futuriste di Poesia, Milano 1913) French transl.: *L'art des bruits. Manifeste futuriste 1913* (Richard-Masse, Paris 1954)

32.118 P. Schaeffer: *Traité des objets musicaux. Nouveau Edition* (Seuil, Paris 1977)

32.119 E. Husserl: *Erfahrung und Urteil. Untersuchungen zur Genealogie der Logik*, 5th edn. (Meiner, Hamburg 1976), ed. by L. Landgrebe

32.120 K. Hevner: Experimental studies of the elements of expression in music, Am. J. Psychol. **48**, 246–268 (1936)

32.121 C. Osgood, G. Suci, P. Tannenbaum: *The Measurement of Meaning* (Univ. of Illinois Press, Urbana 1957)

32.122 S. Ertel: Standardisierung eines Eindrucksdifferentials, Z. exp. angew. Psychol. **12**, 22–58 (1965)

32.123 R. Kendall, E. Carterette: Verbal attributes of simultaneous wind instrument timbres: I. von Bismarck's adjectives, Music Percept. **10**, 445–468 (1993)

32.124 A. Schneider, D. Müllensiefen: Musikpsychologie in Hamburg. Ein Forschungsbericht, Syst. Musikwiss. – Syst. Musicol. **7**, 59–89 (2000)

32.125 H. Böttcher, U. Kerner: *Methoden der Musikpsychologie* (Edition Peters, Leipzig 1978)

32.126 S. Mulaik: *Foundations of Factor Analysis*, 2nd edn. (CRC, Boca Raton 2010)

32.127 R. Mores: Nasality in musical sounds – a few intermediate results. In: *Systematic Musicology: Empirical and theoretical Studies*, ed. by A. Schneider, A. von Ruschkowski (Lang, Frankfurt am Main 2011) pp. 127–136

32.128 St Solomon: Semantic approach to the perception of complex sounds, J. Acoust. Soc. Am. **30**, 421–425 (1958)

32.129 V. Rahls: *Psychometrische Untersuchungen zur Wahrnehmung musikalischer Klänge*, Ph.D. thesis (Univ. of Hamburg, Hamburg 1966)

32.130 E. Jost: *Akustische und psychometrische Untersuchungen an Klarinettenklängen* (A. Volk, Köln 1967)

32.131 G. von Bismarck: Timbre of steady sounds: A factorial investigation of its verbal attributes, Acustica **30**, 146–159 (1974)

32.132 R. Kendall, E. Carterette: Perceptual scaling of simultaneous wind instrument timbres, Music Percept. **8**, 369–404 (1991)

32.133 W. Wundt: *Grundriß der Psychologie*, 15th edn. (Engelmann, Leipzig 1928)

32.134 G. von Bismarck: Sharpness as an attribute of the timbre of steady sounds, Acustica **30**, 159–172 (1974)

32.135 L. Marks: On cross-modal similarity: the perceptual structure of pitch, loudness, and brightness, J. Exp. Psychol.: Hum. Percept. Perform. **15**, 586–602 (1989)

32.136 E. Zwicker, H. Fastl: *Psychoacoustics. Facts and Models*, 2nd edn. (Springer, Berlin 1999)

32.137 S. Stevens, J. Harris: The scaling of subjective roughness and smoothness, J. Exp. Psychol. **64**, 489–494 (1962)

32.138 A. Schneider: 'Verschmelzung', tonal fusion, and consonance: Carl Stumpf revisited. In: *Music, Gestalt, and Computing. Studies in Cognitive and Systematic Musicology*, ed. by M. Leman (Springer, Berlin 1997) pp. 117–143

32.139 Ch Seeger: *Studies in Musicology. Vol. I (1935–1975)* (Univ. of Cal. Press, Berkeley 1977)

32.140 W. Thies: *Grundlagen einer Typologie der Klänge* (Wagner, Hamburg 1982)

32.141 R. Kendall, E. Carterette: Perceptual scaling of simultaneous wind instrument timbres: II. Adjectives induced from Piston's 'Orchestration', Music Percept. **10**, 469–502 (1993)

32.142 J. Hajda: The Effect of dynamic acoustical features on musical timbre. In: *Analysis, Synthesis, and Perception of Musical Sounds*, ed. by J. Beauchamp (Springer, New York 2007) pp. 250–271

32.143 A. Zacharakis, K. Pastiadis, J. Reiss: An interlanguage study of musical timbre semantic dimensions and their acoustic correlates, Music Percept. **31**, 339–358 (2013)

32.144 A. Nykänen, Ö. Johansson, J. Lundberg, J. Berg: Modelling perceptual dimensions of saxophone sounds, Acustica **95**, 539–549 (2009)

32.145 W. Aures: Der sensorische Wohlklang als Funktion psychoakustischer Empfindungsgrößen, Acustica **58**, 282–290 (1985)

32.146 W. Aures: Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale, Acustica **59**, 130–141 (1985)

32.147 A. Tversky: Features of similarity, Psych. Rev. **84**, 327–352 (1977)

32.148 D. Medin, R. Goldstone, D. Gentner: Respects for similarity, Psych. Rev. **100**, 254–278 (1993)

32.149 C. Stumpf: *Tonpsychologie*, Vol. 1 (Barth, Leipzig 1883)

32.150 W. Torgerson: *Theory and Method of Scaling* (Wiley, New York 1958)

32.151 F. Sixtl: *Meßmethoden der Psychologie. Theoretische Grundlagen und Probleme*, 2nd edn. (Beltz, Weinheim, Basel 1982)

32.152 I. Borg, J. Lingoes: *Multidimensional Similarity Structure Analysis* (Springer, New York 1987)

32.153 I. Borg, P. Groenen: *Modern Multidimensional Scaling. Theory and Applications*, 2nd edn. (Springer, New York 2005)

32.154 F.G. Ashby, N. Perrin: Toward a unified theory of similarity and recognition, Psychol. Rev. **95**, 124–150 (1988)

32.155 N. Perrin: Uniting identification, similarity and preference: General recognition theory. In: *Multidimensional Models of Perception and Cognition*, ed. by F. Ashby (Erlbaum, Hillsdale 1992) pp. 123–145

32.156 R. Nosofsky: Similarity scaling and cognitive process models, Ann. Rev. Psych. **43**, 25–53 (1992)

32.157 J. Beran: *Statistics in Musicology* (Chapman Hall, Boca Raton 2004)

32.158 S. Donnadieu: Mental representations of the timbre of complex sounds. In: *Analysis, Synthesis, and Perception of Musical Sounds*, ed. by

J. Beauchamp (Springer, New York 2007) pp. 272–319
32.159 J. Grey: An Exploration of Musical Timbre. Report no. Stan-M-2 (CCRMA Dept. of Music, Stanford 1975)
32.160 J. Grey: Multidimensional perceptual scaling of musical timbres, J. Acoust. Soc. Am. **61**, 1270–1277 (1977)
32.161 J. Grey, J. Gordon: Perceptual effects of spectral modifications on musical timbres, J. Acoust. Soc. Am. **63**, 1493–1500 (1978)
32.162 R. Plomp, H. Steeneken: Effect of phase on the timbre of complex tones, J. Acoust. Soc. Am. **46**, 409–421 (1969)
32.163 J. Miller, E. Carterette: Perceptual space for musical structures, J. Acoust. Soc. Am. **58**, 711–720 (1975)
32.164 D. Wessel: Timbre space as musical control structure, Comp. Music J. **3**, 45–52 (1979)
32.165 C. Krumhansl: Why is musical timbre so hard to understand? In: *Structure and Perception of Electroacoustic Sound and Music*, ed. by S. Nielzen, O. Olsson (Elsevier, Amsterdam 1989) pp. 43–53
32.166 P. Iverson, C. Krumhansl: Isolating the dynamic attributes of musical timbre, J. Acoust. Soc. Am. **94**, 2595–2603 (1993)
32.167 St McAdams, S. Winsberg, S. Donnadieu, G. de Soete, J. Krimphoff: Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes, Psychol. Res. **58**, 177–192 (1995)
32.168 B. Markuse, A. Schneider: Ähnlichkeit, Nähe, Distanz: zur Anwendung multidimensionaler Skalierung in musikwissenschaftlichen Untersuchungen, Syst. Musikwiss. – Syst. Musicol. **4**, 53–89 (1996)
32.169 R. Plomp: *Aspects of Tone Sensation* (Academic, London 1976)
32.170 A. Caclin, St McAdams, B. Smith, S. Winsberg: Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones, J. Acoust. Soc. Am. **118**, 471–482 (2005)
32.171 S. Dixon: Onset detection revisited. In: *Proc. 9th Intern. Conf. Digital Audio Effects (DAFx-06), Montreal* (2006) pp. 133–137
32.172 W.R. Garner: *The Processing of Information and Structure* (Erlbaum, Potomac 1974)
32.173 B. Kostek: *Perception-Based Data Processing in Acoustics. Applications to Music Information Retrieval and Psychophysiology of Hearing* (Springer, Berlin 2005)
32.174 St Handel: Timbre perception and auditory object identification. In: *Hearing*, ed. by B. Moore (Academic, San Diego 1995) pp. 425–461
32.175 J. Schouten: The perception of timbre. In: *Reports 6th Int. Congr. Acoust., Tokyo*, Vol. VI (1968) pp. 35–44
32.176 J. Licklider: Basic correlates of the auditory stimulus. In: *Handbook of Experimental Psychology*, ed. by S.S. Stevens (Wiley, New York 1951) pp. 985–1039
32.177 W. Slawson: *Sound Color* (Univ. of Cal. Press, Berkeley 1985)
32.178 H. Pollard, E. Jansson: Analysis and assessment of musical starting transients, Acustica **51**, 249–262 (1982)
32.179 D. Howard, J. Angus: *Acoustics and Psychoacoustics*, 2nd edn. (Focal, Oxford 2001)
32.180 E. Jost: Über den Einfluß der Darbietungsdauer auf die Identifikation von instrumentalen Klangfarben. In: *Jahrb. Staatl. Inst. Musikforsch. (Berlin) für 1969* (1970) pp. 83–92
32.181 C. Reuter: *Die auditive Diskrimination von Orchesterinstrumenten. Verschmelzung und Heraushörbarkeit von Instrumentalklangfarben im Ensemblespiel* (Lang, Frankfurt am Main 1996)
32.182 R. Rasch: The Perception of simultaneous notes such as in polyphonic music, Acustica **40**, 21–33 (1978)
32.183 P. Vos, R. Rasch: The perceptual onset of tones, Percept. Psychophys. **29**, 323–335 (1981)
32.184 B. Lau, R. Bader, A. Schneider, P. Wriggers: Finite Element transient calculation of a bell struck by its clapper. In: *Concepts, Experiments, and Fieldwork: Studies in Systematic Musicology and Ethnomusicology*, ed. by R. Bader, C. Neuhaus, U. Morgenstern (Lang, Frankfurt am Main 2010) pp. 137–156
32.185 A. Schneider, M. Leman: Sonological and psychoacoustic characteristics of carillon bells. In: *The Quality of Bells: Proc. of the 16th Meeting of the FWO Res. Soc. Foundations Music Research*, ed. by M. Leman (Univ. of Ghent., Ghent 2002)
32.186 A. Schneider, M. Leman: Sound, pitches and tuning of a historic carillon. In: *Studies in Musical Acoustics and Psychoacoustics*, ed. by A. Schneider (Springer, Cham 2017) pp. 247–298
32.187 A. Melka: Messungen der Klangeinsatzdauern bei Musikinstrumenten, Acustica **23**, 108–177 (1970)
32.188 T. Gäumann: The pretransient of the harpsichord sound. In: *Proc. Stockholm Musical Acoustics Conf. (SMAC '03)*, Vol. I (2003) pp. 163–166
32.189 A. Beurmann, A. Schneider: Sonological analysis of harpsichord sounds. In: *Proc. Stockholm Music Acoustics Conf. (SMAC '03)*, Vol. I (2003) pp. 167–170
32.190 E. Saldanha, J. Corso: Timbre cues and identification of musical instruments, J. Acoust. Soc. Am. **36**, 2021–2026 (1964)
32.191 L. Wedin, G. Goude: Dimension analysis of the perception of instrumental timbres, Scand. J. Psych. **13**, 228–240 (1972)
32.192 N. Fletcher: Acoustical correlates of flute performance technique, J. Acoust. Soc. Am. **57**, 233–237 (1975)
32.193 X. Serra: Musical sound modelling with sinusoids plus noise. In: *Musical Signal Processing*, ed. by C. Roads, S. Pope, A. Piccialli, G. de Poli (Swets Zeitlinger, Lisse 1997) pp. 91–122
32.194 S. Levine, J. Smith III: A compact and malleable sines + transients + noise model for sound. In: *Analysis, Synthesis, and Perception of Musical Sound*, ed. by J. Beauchamp (Springer, New York 2007) pp. 145–174

32.195 R. Patterson, E. Gaudrain, Th Walters: The Perception of family and register in musical tones. In: *Music Perception*, ed. by M. Riess Jones, R. Fay, A. Popper (Springer, New York 2010) pp. 13–50

32.196 W. Adelung: *Einführung in den Orgelbau*, 3rd edn. (Breitkopf Härtel, Leipzig 1972)

32.197 N. Fletcher, T. Rossing: *The Physics of Musical Instruments* (AIP/Springer, New York 1991)

32.198 P. Goad, D. Keefe: Timbre discrimination of musical instruments in a concert hall, Music Percept. **10**, 43–62 (1992)

32.199 J. Marozeau, A. de Cheveigné, St McAdams, S. Winsberg: The dependency of timbre on fundamental frequency, J. Acoust. Soc. Am. **114**, 2948–2957 (2003)

32.200 St Handel, M. Erickson: A rule of thumb: The bandwidth for timbre invariance is one octave, Music Percept. **19**, 121–126 (2001)

32.201 K. Steele, A. Williams: Is the bandwidth for timbre invariance only one octave?, Music Percept. **23**, 215–220 (2006)

32.202 J. Meyer: Zur klanglichen Wirkung des Streicher-Vibratos, Acustica **76**, 283–291 (1992)

32.203 J. Gärtner: *Das Vibrato unter besonderer Berücksichtigung der Verhältnisse bei Flötisten*, 2nd edn. (Bosse, Regensburg 1980)

32.204 A. Schneider, R. Mores: Fourier-Time-Transformation (FTT), analysis of sound and auditory perception. In: *Sound – Perception – Performance*, ed. by R. Bader (Springer, Cham 2013) pp. 299–329

32.205 M. Barthet, Ph Depalle, R. Kronland-Martinet, S. Ystad: Acoustical correlates of timbre and expressiveness in clarinet performance, Music Percept. **28**, 135–153 (2010)

32.206 H. Pollard, E. Jansson: A tristimulus method for the specification of musical timbre, Acustica **51**, 162–171 (1982)

32.207 E. Zwicker: *Psychoakustik* (Springer, Berlin 1982)

32.208 K. Denbigh: *Three Concepts of Time* (Springer, Berlin 1981)

32.209 A. Schneider: Virtual pitch and musical instrument acoustics. The case of idiophones. In: *Musik im virtuellen Raum. KlangArt-Kongress 1997*, ed. by B. Enders, J. Stange-Elbe (Universitätsverlag Rasch, Osnabrück 2000) pp. 397–417

32.210 A. Schneider: Sound, pitch, and scale: From 'tone measurements' to sonological analysis in ethnomusicology, Ethnomusicology **45**, 489–519 (2001)

32.211 R. Melara, L. Marks: Interaction among auditory dimensions: Timbre, pitch, and loudness, Percept. Psychophys. **48**, 169–178 (1990)

32.212 C. Krumhansl, P. Iverson: Perceptual Interactions between musical pitch and timbre, J. Exp. Psych.: Human Percept. Perform. **18**, 739–751 (1992)

32.213 E. Allen, A. Oxenham: Symmetric interactions and interferences between pitch and timbre, J. Acoust. Soc. Am. **135**, 1371–1379 (2014)

32.214 R. Patterson, R. Milroy, M. Allerhand: What is the octave of a harmonically rich note?, Contemp. Music Rev. **9**, 69–81 (1993)

32.215 H.P. Hesse: Experimente zum musikalischen Intervallurteil. In: *Jahrb. Staatl. Inst. Musikforsch. (Berlin) für 1978* (1979) pp. 72–87

32.216 M. van Tongeren: *Overtone Singing. Physics and Metaphysics of Harmonics in East and West* (Fusica, Amsterdam 2002)

32.217 St Handel: *Listening. An Introduction to the Perception of Auditory Events* (MIT Press, Cambridge 1989)

32.218 A. Schneider, K. Frieler: Perception of harmonic and inharmonic sounds: Results from ear models. In: *Computer Music Modeling and Retrieval. Genesis of Meaning in Sound and Music*, ed. by S. Ystad, R. Kronland-Martinet, K. Jensen (Springer, Berlin 2009) pp. 18–44

32.219 J. Chowning: John Chowning on composition. In: *Composers and the Computer*, ed. by C. Roads (W. Kaufmann, Los Altos 1985) pp. 18–25

32.220 J. Chowning: *Phoné (1979–80)*, CD (Wergo/Schott, Mainz 1988)

32.221 D. Arfib, F. Keiler, U. Zölzer: Source-Filter Processing. In: *DAFX. Digital Audio Effects*, ed. by U. Zölzer (Wiley, Chichester 2002) pp. 299–372

32.222 J. Keuler: Problems of shape and background in sounds with inharmonic spectra. In: *Music, Gestalt, and Computing. Studies in Cognitive and Systematic Musicology*, ed. by M. Leman (Springer, Berlin 1997) pp. 214–224

32.223 F. Lerdahl: Timbral hierarchies, Contemp. Music Rev. **2**, 135–160 (1987)

32.224 A. Schönberg: *Harmonielehre*, 2nd edn. (Universal Ed., Wien 1922)

32.225 A. Schneider: Akustische und psychoakustische Anmerkungen zu Arnold Schönbergs Emanzipation der Dissonanz und zu seiner Idee der Klangfarbenmelodie, Hamburger Jahrb. Musikwiss. **17**, 35–55 (2000)

32.226 A. Fokker: *New Music with 31 Notes* (Verlag für Systematische Musikwissenschaft, Bonn 1975)

32.227 W. Voigt: *Dissonanz und Klangfarbe. Instrumentationsgeschichtliche und experimentelle Untersuchungen* (Verlag für Systematische Musikwissenschaft, Bonn 1985)

32.228 P. Boersma, G. Kovacic: Spectral characteristics of three styles of Croatian folk singing, J. Acoust. Soc. Am. **119**, 1805–1816 (2006)

32.229 A. Bregman: *Auditory Scene Analysis* (MIT Press, Cambridge 1990)

32.230 Y. Ando: *Concert Hall Acoustics* (Springer, Berlin 1985)

32.231 L. Beranek: *Concert and Opera Halls. How they sound* (Acoust. Soc. Am., Woodbury 1996)

32.232 H. Kuttruff: *Room Acoustics*, 5th edn. (Spon, London 2009)

32.233 J. Blauert: *Spatial Hearing. The Psychophysics of Human Sound Localization*, 6th edn. (MIT Press, Cambridge 2008)

32.234 Ch Brown, B. May: Comparative mammalian sound localization. In: *Sound Source Localization*, ed. by A. Popper, R. Fay (Springer, New York 2005) pp. 124–178

32.235 K. Hancock, B. Delgutte: A physiologically based model of interaural time difference discrimination, J. Neurosci. **24**, 7110–7117 (2004)

32.236 F. Wightman, D. Kistler: Sound localization. In: *Human Psychophysics*, ed. by W. Yost, A. Popper, R. Fay (Springer, New York 1993) pp. 155–192

32.237 T. Ziemer: Psychoacoustic effects in wave field synthesis applications. In: *Systematic Musicology: Empirical and Theoretical Studies*, ed. by A. Schneider, A. von Ruschkowski (Lang, Frankfurt am Main 2011) pp. 153–162

32.238 J. Eggermont: Between sound and perception: Reviewing the search for a neural code, Hearing Res. **157**, 1–42 (2001)

32.239 I. Nelken: Processing of complex stimuli and natural scenes in the auditory cortex, Curr. Opin. Neurobiol. **14**, 474–480 (2004)

32.240 I. Nelken: Processing of complex sounds in the auditory system, Curr. Opin. Neurobiol. **18**, 413–417 (2008)

32.241 C. Von der Malsburg: Binding in models of perception and brain function, Curr. Opin. Neurobiol. **5**, 520–526 (1995)

32.242 A. Roskies: The binding problem, Neuron **24**, 7–9 (1999)

32.243 B. Moore: Frequency analysis and pitch perception. In: *Human Psychophysics*, ed. by W. Yost, A. Popper, R. Fay (Springer, New York 1993) pp. 56–115

32.244 B. Moore: Frequency analysis and masking. In: *Hearing*, ed. by B. Moore (Academic, San Diego 1995) pp. 161–205

32.245 B. Moore: Basic psychophysics of human spectral processing. In: *Auditory Spectral Processing*, International Review of Neurobiology, Vol. 70, ed. by M. Malmierca, D. Irvine (Elsevier, Amsterdam 2005) pp. 49–86

32.246 E. Terhardt, G. Stoll, M. Seewann: Algorithm for extraction of pitch and pitch salience from complex tone signals, J. Acoust. Soc. Am. **71**, 679–688 (1982)

32.247 E. Terhardt, G. Stoll, M. Seewann: Pitch of complex signals according to virtual-pitch theory: Tests, examples, and predictions, J. Acoust. Soc. Am. **71**, 671–678 (1982)

32.248 E. Terhardt, M. Seewann: Auditive und objektive Bestimmung der Schlagtonhöhe von historischen Kirchenglocken, Acustica **54**, 129–144 (1984)

32.249 D. Huron: Voice denumerability in polyphonic music of homogeneous timbres, Music Percept. **6**, 361–382 (1989)

32.250 R. Rasch: Synchronization in performed ensemble music, Acustica **43**, 121–131 (1979)

32.251 St McAdams: Spectral fusion and the creation of auditory images. In: *Music, Mind, and Brain*, ed. by M. Clynes (Plenum, London 1982) pp. 279–298

32.252 St McAdams: Segregration of concurrent sounds. I: Effects of frequency modulation coherence, J. Acoust. Soc. Am. **86**, 2148–2159 (1989)

32.253 D. Mellinger, B. Mont-Reynaud: Scene analysis. In: *Auditory Computation*, ed. by H. Hawkins, T. McMullen, A. Popper, R. Fay (Springer, New York 1996) pp. 271–331

32.254 J. Van Eyck: Der Fluyten Lust-Hof, beplant met Psalmen, Pavanen, Almanden, Couranten, Balletten, Airs (Matthysz, Amsterdam 1654)

32.255 L. Van Noorden: *Temporal Coherence in the Perception of Tone Sequences*, Ph.D. Thesis (Technical Univ. of Eindhoven, Eindhoven 1975)

32.256 G. Kubik: Die Amadinda-Musik von Buganda. In: *Musik in Afrika*, ed. by A. Simon (Museum für Völkerkunde, Berlin 1983) pp. 139–165

32.257 G. Kubik: *Theory of African Music*, Vol. I (Heinrichshofen, Wilhelmshaven 1994)

32.258 U. Wegner: Cognitive aspects of Amadinda xylophone music from Buganda: Inherent patterns reconsidered, Ethnomusicology **37**, 201–241 (1993)

32.259 G. Miller, G. Heise: The trill threshold, J. Acoust. Soc. Am. **22**, 637–638 (1950)

32.260 A. Bregman, J. Campbell: Primary auditory stream segregation and perception of order in rapid sequences of tones, J. Exp. Psych. **89**, 244–249 (1971)

32.261 P. Iverson: Auditory stream segregation by musical timbre. Effects of static and dynamic acoustic attributes, J. Exp. Psych.: Hum. Percept. Perf. **21**, 751–763 (1995)

32.262 B. Roberts, B. Glasberg, B. Moore: Primitive stream segregation of tone sequences without differences in fundamental frequency or passband, J. Acoust. Soc. Am. **112**, 2074–2085 (2002)

32.263 H. Helson: The fundamental propositions of Gestalt Psychology, Psych. Rev. **40**, 13–32 (1933)

32.264 A. Fischer: *Neurophysiologisch motivierte Modelle zur akustischen Figur-Hintergrund-Trennung* (Deutsch, Frankfurt am Main 1994)

32.265 S. McCabe, M. Denham: A model of auditory streaming, J. Acoust. Soc. Am. **101**, 1611–1621 (1997)

32.266 A. Klapuri: Auditory model-based methods for multiple fundamental frequency estimation. In: *Signal Processing Methods for Music Transcription*, ed. by A. Klapuri, M. Davy (Springer, New York 2006) pp. 229–266

32.267 A. Klapuri: Multipitch analysis of polyphonic music and speech signals using an auditory model, IEEE Transactions Audio, Speech, and Language Process. **16**, 255–266 (2008)

32.268 K. Kashino: Auditory scene analysis in music signals. In: *Signal Processing Methods for Music Transcription*, ed. by A. Klapuri, M. Davy (Springer, New York 2006) pp. 299–325

32.269 M. Goto: Music scene description. In: *Signal Processing Methods for Music Transcription*, ed. by A. Klapuri, M. Davy (Springer, New York 2006) pp. 327–359

32.270 F. Cañadas Quesada, N. Ruiz Reyes, P.V. Candeas, J. Carabias, S. Maldonado: A multiple-F0 estimation approach based on Gaussian spectral modelling for polyphonic music transcription, J. New Music Res. **39**, 93–107 (2010)

32.271 A. Schneider: Klanganalyse als Methodik der Popularmusikforschung, Hamburger Jahrb. Musikwiss. **19**, 107–129 (2002)

33. Sensation of Sound Intensity and Perception of Loudness

Albrecht Schneider

This chapter is on sensation of sound intensity and perception of loudness. Since some of the relevant matter (on scaling concepts of loudness) has been presented in Chap. 30, and because a considerable portion of research on loudness is done outside musical contexts (namely, in industrial and environmental noise control as well as in audiology), this chapter condenses facts and models more than the previous two on pitch and timbre respectively. Section 33.1 of this chapter offers the physical and physiological basis of sound intensity sensation while Sect. 33.2 discusses features of some models of loudness sensation that have been established in psychoacoustics over the past decades. Since these models were originally designed for stationary sound signals and levels, and have been tested mostly in lab situations, they cannot adequately cover a range of real-world sound types found in natural or technical environments. In music genres such as techno presented in discos, or heavy metal performed in live music venues or at open air festivals to audiences at very high sound pressure levels, sound is heavily processed in regard to dynamics and spectral energy, which calls for appropriate measurement and assessment of sensory effects. Different from perception of pitch (where samples of subjects respond more or less in similar ways to certain types of sound signals), perception of loudness shows a high degree of variability even within groups of musically trained subjects reflecting their musical background and preferences (Sect. 33.3). Recent empirical evidence demonstrates that subjects judge loudness for various musical genres on a category scale (from very soft to very loud), however, the center (relative to loudness level and loudness scales) and the range of each category differ considerably, for individual subjects.

Finally, there is a concluding section (Sect. 33.4) in which some of the major topics and issues discussed in Chaps. 30–33 of Part D are summed up. In addition, a tentative model of the interrelationship of pitch, timbre and loudness perception is sketched.

33.1 Physical and Physiological Basis of Sound Intensity Sensation

Subjects usually have a sensation from physical sound intensity that they experience as *sound volume*, or as *loudness level*, that is, sensation grows as intensity increases. However, the exact form of such a correspondence is complex and has been the topic of a large number of studies. In addition to the descriptions and interpretations given above (Sect. 30.1.4, *Loudness Scaling* ...), some more details shall be discussed. From a psychophysical point of view, sensory magnitudes (Sect. 30.1.4) should reflect physical input parameters in some (linear or nonlinear) way. Sound intensity $\boldsymbol{I} = p\boldsymbol{v}$ as the product of sound pressure and particle velocity can be related to the sound power P like

$$I = \frac{P}{S} = cw \,,$$

where P is the sound power, S is the area (m^2), c is the wave speed in a medium and w is the sound energy density (Sect. 30.2). The sound power (acoustic power) radiated from a source is a measure of the sound energy that goes through a plane per short time unit like

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_33

$\Delta W = wSc\Delta t$. Humans and other mammals collect the sound power entering the ear channel from the pinna to the tympanon where, however, a strong divergence between the sound pressure level (SPL) measured in the free field and SPL at the tympanum can occur depending on the azimuth; the divergence is zero for 0° and increases both with angle and with frequency up to ≈ 15 dB [33.1] since, for higher frequencies, the wavelength does not permit waves to flow around the head. Moreover, a significant portion of the energy is reflected. The pressure reflectance is calculated from the complex ratio of incident pressure p_i to the reflected pressure p_r like

$$R(\omega) = \frac{p_r}{p_i} .$$

The energy reflectance then is the ratio of the reflected power P_r to the incident power P_i; the energy reflectance equals R squared

$$\Re(\omega) = |R|^2 .$$

For the human ear canal and impedance, values for the energy reflectance coefficient as dependent on frequency have been reported for adults to range from about 0.3 at 2–4 kHz up to 0.95 at 125 Hz and 0.7 or 0.8 at 8 kHz [33.2, 3]. Thus, the reflectance is significant at low and at high frequencies while much more of the incoming energy passes the tympanum into the middle ear in a frequency range from, roughly, 0.5–5 kHz. A reflection coefficient of $|R|^2 = 0.5$ means that 25% of the energy is reflected (100×0.5^2 %).

Contrary to the effect of the loss from reflection, the middle ear transfer gain increases the sound pressure up to a maximum of ≈ 23.5 dB at 1.2 kHz [33.4]. However, the power utilization ratio estimated for humans reaches about 0.4 (on a scale from 0–1) at ≈ 3 kHz (with a second peak at ≈ 8 kHz; [33.5, p. 129, Fig. 6]), and the middle ear efficiency likewise peaks at 1 kHz where efficiency is at 0.35 [33.5, p. 129, Fig. 7]. On the average, middle ear efficiency is about 0.1 from 100 Hz to ≈ 1.5 kHz and from there on diminishes rapidly. The graph of a function calculated as isopower contour [33.5, p. 131, Fig. 11] to represent the SPL (dB) in the free field needed to maintain a constant sound pressure at the cochlea for humans (threshold at 10^{-18} W) looks quite similar to the threshold curves known from psychoacoustic behavioral data [33.6], that is, the threshold is below 0 dB at ≈ 3 kHz while the slopes of the isopower contour rise quite steeply in particular toward low frequencies. In sum, of the sound power collected by the ears a considerable amount is reflected at the tympanum, and of the fraction that actually passes into the middle ear, less than half finally enters the cochlea. In this respect, the ear as a transducer is quite inefficient while its frequency characteristic response shows fair admittance for sound wave components in a range most useful for speech communication and music (≈ 500 Hz–4 kHz). The steep roll-off toward very low and very high frequencies can also be viewed as a self-protective mechanism.

In regard to inner ear excitation on the basilar membrane (BM) level, the traveling wave in the cochlea has to be considered as a combination of a compression wave in the fluid and a transversal motion of the BM (Sect. 31.3). Assuming cochleotopic frequency mapping, for simple stimuli such as pure tones a maximum of the pressure amplitude and of the local deflection can be expected at the place where a certain characteristic frequency (CF) is represented. In a passive linear system, excitation should be proportional to the pressure amplitude and to the BM transversal deflection. This, however, is not the case with the BM/outer hair cell (OHC) feedback system, which exhibits a strong compressive nonlinearity effective above 30–40 dB SPL. As has been found in animal experiments [33.7] (chinchilla), sensitivity of the BM is such that 0 dB SPL can effect a small BM displacement, while 1.6 nm were measured at 20 dB. *Russell* and *Nilsen* [33.8] (guinea pig) observed that, for a CF of 13.25 kHz, BM displacement as a function of SPL grew to ≈ 4 nm at 40 dB from where the slope flattens. At 80 dB SPL, displacement at the CF was ≈ 10 nm. In regard to gain (measured either as displacement amplitude divided by SPL, or as velocity divided by stimulus pressure), the cochlear amplifier is most effective at low SPL where gain according to *Ruggero* et al. [33.9] (chinchilla) reaches 66–76 dB relative to stapes motion for a 5–10 dB SPL input while, above ≈ 40 dB and up to ≈ 80 dB, nonlinear compression grows with level. The difference in peak gain for a low level input (5 or 10 dB) and an 80 dB SPL input was 47.9 dB [33.9]. Expressed as BM velocity (μm/s) in response to sine tone bursts, velocity was at ≈ 10^2 for a 20 dB input level and at ≈ 10^3 for 60–80 dB while BM displacement data give ≈ 4–5 nm at 20 dB and 10–15 nm in the 60–80 dB range respectively [33.9, Fig. 2 and 6].

Looking for spatial characteristics, displacement for a certain CF at low SPL (10–20 dB) was confined to a small BM section (< 1 mm), which can be viewed as a symmetric bandpass centered at CF while above 40 dB displacement is more extended [33.8, Fig. 1D]. This spread of displacement with SPL is significant in particular above 60 dB and has been incorporated as a feature into auditory models suited to calculate loudness (see below). In effect, the input/output ratio

at the BM/inner hair cell (IHC) level can be compared to an automatic gain control amplifier though its operating principle seems to be control of the local BM impedance/admittance (there are hypotheses saying that local resonance frequencies of the BM are almost invariant in regard to stimulus intensity; see [33.10]). In the range where outer hair cell (OHC) gain control seems to work best the compression ratio appears to be close to 0.2 dB/dB and reaches saturation at about 80–85 dB SPL, above which *passive* mechanical parameters (BM stiffness, fluid viscosity) seem to delimit the excitation process. The nonlinear compression carried out by the BM/OHC circuit reduces the dynamics of hearing from the ≈ 120–130 dB range at the input (defined by the lower and upper threshold [33.11, 12]) to a much smaller range effective at the cochlear transduction stage [33.13].

Intensity coding has often been viewed in regard to firing rates of neurons and fibers. In a well-known hypothesis known as *auditory nerve spike count*, the concept was that, with growing intensity of a stimulus such as a sine tone, the firing rate in auditory nerve (AN) fibers would increase accordingly so that the number of action potentials or spikes summed over the fibers activated per time interval would be in proportion to stimulus intensity. Sensation of loudness then could be directly related to the number of spikes in the AN. However, taking the compressive nonlinearity of BM motion into account, linear growth of spike rate with intensity would be unlikely. Also, the spread of excitation with input level and the shape of excitation patterns have to be considered [33.14]. Assuming the spread of excitation activates neurons in particular at frequencies above the CF corresponding to a sine tone as input, a tone high in frequency and level would not find many neurons left for such spread while a tone low in frequency could activate neurons toward higher CFs even at a moderate level. In effect, the hypothesis according to which loudness sensation can be related to AN spike activity in a simple way has been questioned [33.15]. This does not imply, though, that intensity coding would not be detectable from neural spike patterns. At least for pure tones, intensity coding for a wide dynamic range can be traced from neural spike counts such as the peristimulus compound action potential (PCAP, see [33.16]). Since the dynamic range of single AN fibers is limited to ≈ 30–60 dB depending on their rate of spontaneous activity and threshold [33.17], networks of fibers would be needed to cover broadband signals such as harmonic complexes at high input levels. Given that there are fibers with low spontaneous rates and higher thresholds as well as fibers with higher spontaneous rates and low thresholds, combinations of relatively few ($n = 50–100$) fibers would suffice to code about 100 dB of intensities [33.18]. In addition to firing rates, phase locking, lateral suppression and other effects may be involved in intensity coding. On the cortical level, an increase of activated cortex volume with rising sound level was observed in functional magnetic resonance imaging (fMRI) studies. There are some indications that the activation patterns may represent both the intensity of the stimulus and the perceived loudness [33.19].

Also memory traces have been hypothesized as a factor influencing intensity discrimination and identification processes [33.20]. Apparently, loudness levels as actually sensed enter short-term memory (STM). It has been observed that a sound with a ramped envelope was judged louder than the same sound with damped envelope [33.21]. A damped sound typically has an exponential decay, which leaves a low or vanishing level as the last *entry* into STM. The ramped version of the same sound is obtained from a simple reversal of the time series (in digitized sounds) so that the attack portion and samples with high amplitudes are the last *entry* into STM. A comparison of the sensations the two stimuli evoke can be expected to find the ramped version louder even though the energy in both stimuli is identical.

The compressive nonlinearity observed in dynamic auditory processing at one ear (i. e., in monaural stimulus presentation) apparently has some parallel in binaural loudness summation. In theory, loudness as sensed should double (ratio 2 : 1) if the input from both ears is added linearly, which could be achieved first on the level of the superior olivary complex (SOC) and at higher stations of the auditory pathway (AuP) (Fig. 31.2). In fact, it has been postulated, from behavioral experiments with pure tones (100, 400, 1000 Hz) presented at different levels (20–50 dB SPL), that binaural summation is linearly additive [33.22]. Other experiments found that binaural loudness summation gave a much lower ratio (between 1.3 : 1 and 1.7 : 1; see [33.23]) though these ratios seemed to be largely independent of level (tested for a 1 kHz tone over a wide dynamic range by *Marozeau* et al. [33.24]). In effect a ratio significantly smaller than 2 : 1 provides another hint that the power law for loudness $L = kI^{03}$ (Sect. 30.1.4, *Loudness Scaling* ...) needs correction even for elementary stimuli [33.25].

33.2 Models of Loudness Sensation

From experiments and calculations, *Harvey Fletcher* [33.26] argued that the total loudness N perceived from noise bands at higher centers of the AuP can be explained as an integral comprising partial loudnesses N_x from what he saw as 100 *patches* of auditory nerves distributed along the BM (for which he assumed a position coordinate, x), thus

$$N = \int_0^{100} N_x \mathrm{d}x \,.$$

Fletcher found a quantitative relation between a masking pattern expressed in dB and what he called agitation of nerve endings. His considerations include loudness summation across bands or *patches* as well as a correspondence between the bandwidth and frequency position of the stimulus and the *patch* (or *patches*) *of agitation*. In short, he described for loudness summation and loudness perception in general a concept later known as the critical band (CB), with a map of 20 bands covering 10 kHz stimulus frequency and 100 *patches* of nerve endings [33.26, p. 283, Fig. 8]. Later research on loudness perception [33.27–29] has drawn on the CB as the unit needed in particular for loudness summation across several bands determined apparently by the bandwidth of the auditory filter (AF). The CB unit that was developed from experiments employing a number of masking designs was called *Bark* (to honor Heinrich Barkhausen [33.30, p. 249 ff.]) The Bark scale usually comprises 24 or 25 CBs each of which is taken to correspond to 1.3 mm of the BM (Sect. 30.1.4, *Loudness Scaling* ...); the bandwidth of CB_c centered at f_c (kHz) can be calculated [33.31] like

$$CB_c/\mathrm{Hz} = 25 + 75[1 + 1.4(f_c/\mathrm{kHz})^2]^{0.69}$$

For example, the bandwidth of a CB centered at 0.5 kHz would be 117.3 Hz, for 1 kHz it is 162.2 Hz and for a band at 4 kHz it is 685.4 Hz. The problem with the Bark scale is that the bands below 500 seem too broad in bandwidth to account for perception of musical intervals in low register (Sect. 31.6.2) and particularly so in the range from $\approx 55-220$ Hz (A1 to A3). As an alternative, a revised scale known as ERB (equivalent rectangular bandwidth; [33.14, 32] and Fig. 31.9) has been established for which the bandwidth can be calculated like

$$\mathrm{ERB} = 24.7(4.37 f_c + 1) \,,$$

where f_c is the center frequency. The CB bandwidths calculated with this formula are considerably smaller than the Bark bands, and hence represent smaller sections of the BM (0.86 mm per ERB). Smaller CBs below 500 Hz than offered by the Bark(z) scale fit closer to perception of musical intervals and scales. Though both scales refer to a *rectangular* bandwidth as far as nonoverlapping, abutting CBs may be assumed, there is a technical definition of ERB in models of the AF as are central in the ERB concept of Moore and Glasberg. In research efforts of the 1940s and 1950s on audition, the CB in the range above 0.5 kHz to $\approx$ 10 kHz was already viewed as kind of a filter [33.26, 33] that could be approximated by a 1/3 octave bandpass. Behavioral data from masking experiments showed the actual bandwidth is smaller than this (1/3 octave = 400 cent) since, in the range 0.5−10 kHz, even for the Bark scale the width of the CBs is from 366 cent at maximum to 257 cent at about 2 kHz [33.34, Table 1]. Taking the CB as a bandpass filter, one can model the BM as a bank of 24 or 25 Bark filters (as in Fig. 32.13), or likewise, as a bank of 35−40 ERB filters; the absolute number of filters employed in a BM model depends on filter parameters such as the slopes of the upper and lower skirts as well as the overlap of adjacent filters. Single filters are parametrized so as to be equivalent in function and performance to the AF (or to several AFs should their characteristics not be uniform along the BM). The filter bank concept of the AF has become part of many computerized auditory models [33.35–38]. The AF according to behavioral data is asymmetric; there are several models and implementations ([33.14] and [33.39, Chap. 9]) that converge on certain features, the most important being dependency of filter shape on sound level. Without going into details of auditory filter parameters [33.14, 40, 41], it is clear that an algorithm representing the AF must vary bandwidth as measured in absolute frequencies with center frequency to account for the nearly identical bandwidth of CBs along the BM (1/3 octave bandpass filter banks with 31 ISO standardized center frequencies provide an analogy of constant-Q filtering). Further, the slope of the upper and lower skirts of the filters should adapt to changing input levels to account for the effect known as spread of excitation (see below). The so-called roex (rounded exponent) filter models the AF incorporating these characteristics; another solution (closely related in design to the roex) is the gammatone filter defined by its impulse response (cf. [33.30, Chap. 10] and [33.42], for a Mathematica™ implementation of the ear model from *Patterson* et al. [33.37]).

The following analysis might illustrate BM filter operation: pop music recordings of the so-called dance floor genre often are produced with four very strong

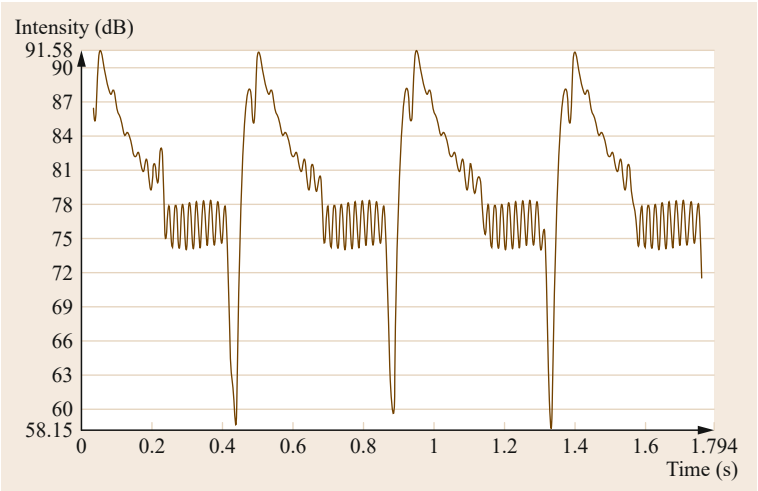


Fig. 33.1 Intensity of four beats (one measure, ≈ 1.8 s, *Fight for your right to party*)

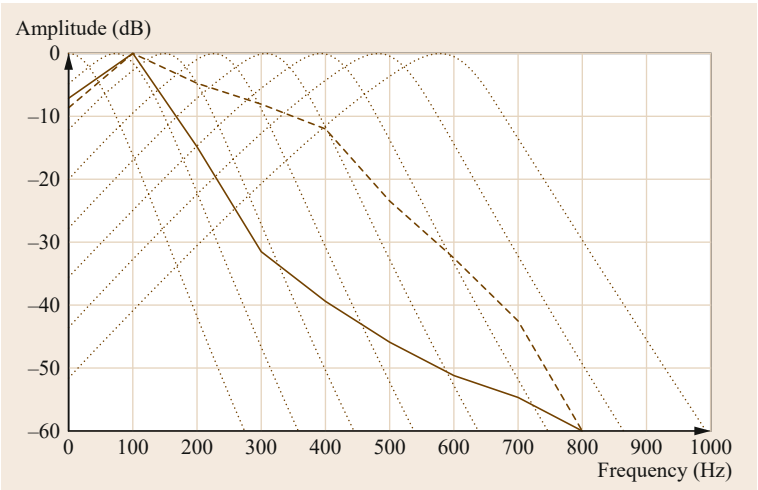


Fig. 33.2 Eight Bark filters (*dotted lines*) spaced at 0.8 Bark as applied to the analysis of a section of *Fight for your right to party* (N.Y.C.C. 1998); spectral energy at 0.1 s of the sound (*solid line*) and at 0.25 s (*dashed line*)

beats per measure marked by a (real or synthetic) bass drum sound, which can be gated (with a noise gate that opens only after a certain threshold level (dB) has been surpassed) so that the sound level rises steeply. The beats thereby become pulses as is shown in Fig. 33.1 from the CD sound file of *Fight for your right to party* in the version of N.Y.C.C., 1998, measure 1 after the intro.

As the graph demonstrates, the difference between minimum and maximum level is more than 30 dB, and the rise time from local minima to peaks is about 50 ms. Mean energy of the sound example (one measure, ≈ 1.8 s) is at 83.44 dB. Since the four beats are realized with a low bass drum sound, the energy can be expected to fall into only a few AF (or a few CBs respectively). Applying a set of gammatone filters whose center frequencies are spaced at 0.8 Bark (with the first filter centered at 0.8 Bark), and performing an analysis for the frequency range of 0–1000 Hz, the filter functions for eight filters (*dotted lines*) and the spectral energy distribution for windows of 100 ms calculated at 0.1 s (*solid line*) and 0.25 s (*dashed line*) of the time series representing the sound are shown in Fig. 33.2. Though the window size of 100 ms includes Gaussian weighting at the boundaries, it represents a reasonable integration time with respect to constants of temporal loudness integration, which have been reported to be of 80–100 ms [33.11, Chap. 11], [33.12, Chap. 8.5]. That is, loudness sensation for steady-state stimuli falling into one CB usually grows fast with stimulus duration up to ≈ 80–100 ms but only slightly for durations up to 200 ms (where the function almost saturates). From Fig. 33.2 it is obvious that particularly loud sounds from single instruments engage several adjacent AF.

Taking the AF paradigm, the energy falling into a certain filter band in a time interval t_i (ms) can be taken as a physical quantity suited to evoke an excita-

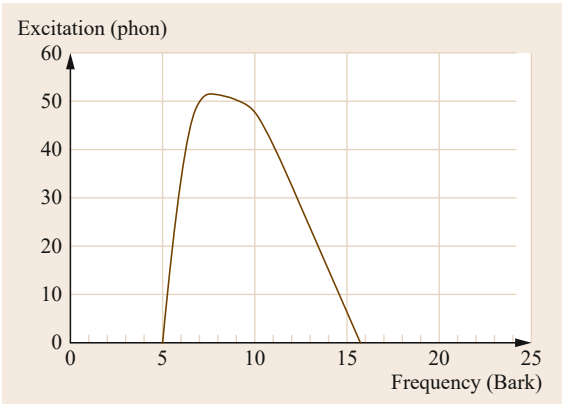


Fig. 33.3 Excitation pattern for three harmonic partials {800, 1000, 1200 Hz}; sound normalized to −24 dBfs peak level. Excitation level peaks at ≈ 50 phon

tion with a strength proportional to the the input signal. Such a view holds, in principle, if the signal is simple like a sine tone and low in input level ($L \leq 20$ dB SPL). Under such conditions, the excitation zone appears to be limited in spatial extension along the BM while transversal displacement of the BM is also small (see above). For higher levels, however, these conditions no longer apply since excitation clearly grows with level [33.14, p. 178, Fig. 7] so that even a sine tone of given frequency will bring about considerable spread of excitation at levels > 30 dB. For example, if several harmonics of a complex are used as input, even at moderate levels several AF will be engaged as is shown in Fig. 33.3 for partials nos. 4, 5, 6 of a virtual 200 Hz harmonic complex. The sound comprises three components {800, 1000, 1200 Hz} digitally normalized to −24 dBfs peak level. The three components are separated by 386 and 316 cent respectively, and should fall into separate CBs in the frequency range chosen. The resulting excitation zone, however, spans from $z = 6$ Bark centered at ≈ 700 Hz almost to $z = 16$ Bark at about 3 kHz.

Loud but undistorted sounds recorded from musical instruments are broadband signals in regard to CB/AF engagement as is shown in Fig. 33.4 for three sounds (cello, bell, crash cymbal) normalized to 0 dBfs peak level.

With increasing input level, displacement of BM segments increases in amplitude, which implies a spread of excitation along the BM. Such spread observed already at moderate levels, and massively so at high levels, means the number of inner hair cells (IHCs) activated along the BM increases with level. In particular for very high stimulus levels (120 dB) indications for multiple CP/BM resonances and corresponding neural activity have been reported [33.43]. For broadband signals such as music, energy typically falls into many AF (or CBs). In regard to perceiving the loudness of music at a certain time, it seems reasonable to assume integration of energy contributed by all the CBs that are activated within a certain period should evoke a sensation of loudness somehow proportional to the energy input and in particular to the excitation pattern corresponding to BM motion (displacement, velocity, acceleration). Taking up Fletcher's suggestion that the total loudness, N, of a sound is the integral or sum of specific loudnesses N_x contributed by each CB, one needs to measure the energy input and then find the excitation pattern needed to calculate the various N_x from which N is derived.

The model of loudness summation proposed by *Zwicker* and *Scharf* [33.29] (see also [33.44] and [33.12, Chap. 8.7]), comprises components that can be viewed as steps in a sequence of processing:

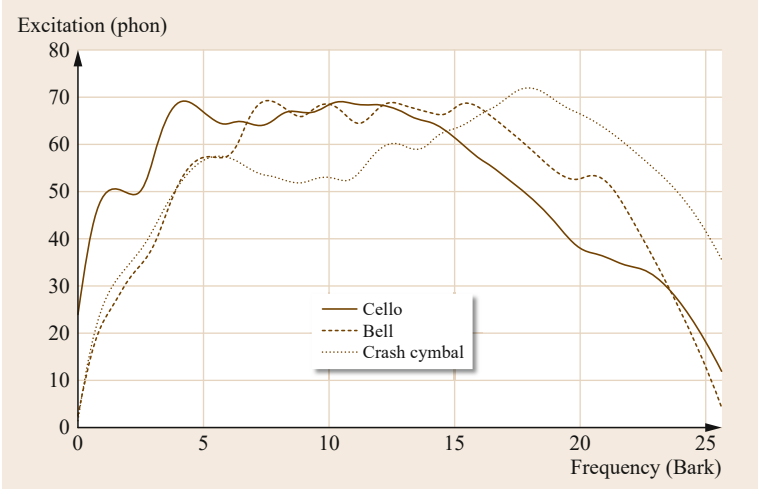


Fig. 33.4 BM excitation pattern for three loud sounds from cello (*solid line*), bell (*dashed line*), and crash cymbal (*dotted line*). Excitation peaks at ≈ 70 phon and involves all CBs

1. An input signal like a sine tone of given frequency and intensity (I, $re10^{-16}\,\mathrm{W/cm^2}$) enters the cochlea via the meatus and the middle ear ossicles (for characteristics and parameter values of the transduction, see Sect. 31.3).
2. The input signal causes a displacement on the BM at the location corresponding to the BM frequency map (Figs. 31.4–31.5).
3. The displacement then is transformed into an excitation, E.
4. From this excitation the specific loudness N' is calculated.
5. The total loudness N is summed in an integral along the Bark scale like

$$N = \int_0^{24\,\mathrm{Bark}} N' \mathrm{d}z \,.$$

The excitation level

$$L_\mathrm{E} = 10 \log \left(\frac{E}{E_0} \right) \mathrm{dB}$$

is calculated from the intensity I_G of the sound per band transformed into a critical band level,

$$L_\mathrm{G} = 10 \log \left(\frac{I}{I_0} \right) \mathrm{dB}$$

(for details, see [33.12, Chap. 6]). Assuming that the specific loudness per band approximately is

$$N' \sim \left(\frac{E}{E_0} \right)^\mathrm{k} ,$$

where E_0 is the level of excitation corresponding to the reference intensity ($I_0 = 10^{-12}\,\mathrm{W/m^2}$), the exponent needs to be determined. A value found fitting for noise bands is $k = 0.23$ [33.44, Chap. 15], [33.12, Chap. 8.7].

According to Zwicker's original model, the loudness (in sones) of a broadband sound can be calculated from the pattern of specific loudnesses if the sound is steady state over a period relevant for auditory sensation. Reasonable time constants are from 80 or 100 ms to 200 ms since the time constant is dependent on level as well as on frequency; at threshold and for tones at 500 Hz, 200 ms are more realistic while at 50 dB and 1000 Hz an integration time of 100 ms or less may apply (for empirical data obtained from tone pulses, see [33.45]). Loudness sensation for steady-state stimuli falling into one CB usually grows fast with stimulus duration up to $\approx$ 80–100 ms but only slightly for durations between 100–200 ms (where the function almost saturates). However, if stimuli are unmodulated both in bandwidth and temporal envelope, loudness adaptation to such steady-state sounds can occur if their duration exceeds a period from one to several seconds. Thereby, a reduction in loudness sensation takes place even though the amount of energy contained in the stimulus does not change. An explanation of such phenomena can be sought on the neural level where the firing rate of neurons constantly engaged might drop while the recovery time between firings grows at saturation levels.

Given that many environmental sounds are modulated rather than steady state, more recent models [33.46–48]) provide algorithms suited to cover various rates of amplitude modulation (AM) and other temporal effects such as tone bursts. These models differ in certain respects while the general design includes the ear canal transfer, an auditory filter bank and a temporal integration stage. In addition, a compressive nonlinearity has been introduced to simulate the behavior of the hearing system for higher SPL. Spectral processing (simulating AF/CB filtering) has been implemented either as short-term Fourier transform (STFT) (with several FFT processes running in parallel to cover different frequency bands, [33.46]), where the analysis window is time-shifted over the signal in small increments, or as Fourier time transformation (FTT [33.47]). Temporal integration is a special problem given the range of natural and technical signals in various environments. While a number of signals show slow fluctuations in level, many exhibit rapid ones, and some are markedly *peaky* with a more or less regular pulse structure (Fig. 33.1). Hence, the inner ear must be quick to deal with impulse-like sounds that have very short rise and decay times respectively. An obvious choice, therefore, is to provide at least two different integration constants, which in some models are accessed either in parallel or in a consecutive processing line [33.49, 50]. A low-pass filter accounts for the fact that the hearing system cannot follow temporal fluctuations in the signal beyond an upper limit frequency. The compressive nonlinearity, which relates SPL to loudness, is implemented by a weighting where a proper exponent for the power function must be chosen (common values are $0.23 < a < 0.6$). The model of *Sottek* [33.48] includes an autocorrelation analysis for the calculation of the specific loudness within each channel (to account for tonal and noise content in bandpass signals).

There have been several comparative tests [33.48, 50, 51] in order to evaluate the performance of various models relative to a range of different sounds, in particular time-varying technical sounds (most of them

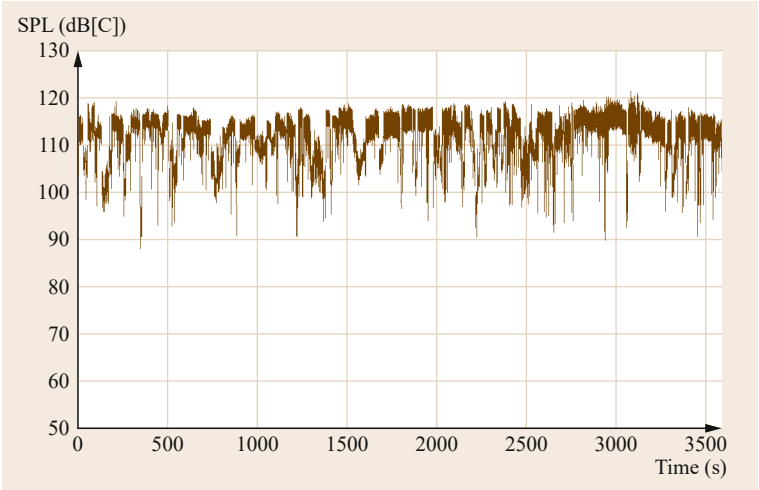


Fig. 33.5 SPL (dB[C]) over time, 1 h of hard trance music recorded in a disco (one channel)

from machines or weapons plus a snare drum). The results show that the models underestimated the loudness of certain sounds (relative to behavioral data for the same sounds taken as stimuli) while overestimating others. The deviations could be quite small in some cases (< 1.5 dB) but significant (> 10 dB) in others. However, one has to see that the behavioral data for subjective loudness usually scatter over a wide range (Sect. 30.1.4, *Loudness Scaling* ...), which poses a problem for defining reference levels as well as suited exponents for the nonlinearity. In addition to spectral and temporal loudness summation parameters, also binaural loudness must be considered in the design of models. Taking into account that binaural loudness across both ears is less than expected from linear summation, *Moore* and *Glasberg* [33.52] have modified their model assuming inhibition between outputs of both ears. The model parameters were chosen so that a signal presented diotically may appear 1.5 times as loud as the same signal presented monaurally. While previous loudness models suited to measure steady-state sound used longer integration constants, the concept of recent models adapted to time-varying sounds is to calculate short-time or even *instantaneous* loudness. With respect to sensation, in models based on FFT or FTT, window size (number of samples) can be chosen so as to yield meaningful integration constants. A software package like ArtemisS (Head Acoustics) offers a range of psychoacoustic parameters including various SPL and loudness measurements, which were used, for example, in examining loudness levels and sound characteristics in discotheques and music clubs [33.53, 54]. In one such venue where *hard trance* is the main musical genre played to the visitors, a recording of the sound did yield a SPL for one hour as displayed in Fig. 33.5. For the calculation of SPL over time (1 h = 3600 s), the integration constant *fast* (125 ms rise time, 125 ms decay) and spectral weighting according to the dB(C) filter function was applied. One can see that the average SPL is at about 112–115 dB(C), which means an average pressure of ≈ 8–11 Pa. From the same raw data recorded in this club, loudness N (sone) over time was calculated for one hour of music according to a standard method (FFT/ISO 532 B, now revised by ISO 532-1:2017). Loudness had some peaks far beyond 300 sone and reached 178.1 sone on average, corresponding to 114.76 phon. Exposure to such levels for hours can cause temporary or permanent hearing loss in subjects (see below).

A decisive aspect in algorithmic modeling of loudness seems to lie in the sequence of temporal and spectral processing stages, i. e., whether spectral analysis according to AF/CBs is performed before or after analysis of temporal envelope features. Most models carry out spectral analysis first and then look into the output of each CB filter for the envelope. It seems feasible, though, that in mammalian auditory processing temporal and spectral information is derived conjointly from BM displacement and velocity, and is coded in a parallel process rather than sequentially.

33.3 From Lab to Disco: Measurements and Perceptual Variability of Loudness

Most of the measurements on loudness sensation and perception have been carried out in lab situations where, in the tradition of psychophysics, quite often designs are employed that investigate variation in one sensory variable as a function of variation in a physical parameter, or combine a few input variables (e.g., frequency, SPL, duration) to measure their effect on sensation. In this respect, sensation of loudness as depending on the frequency and/or SPL as well as the duration of pure tones has been studied; likewise, loudness of noise bands as depending on bandwidth and/or level and/or modulation has been investigated (several of the chapters in [33.55] provide comprehensive overviews). Many of these experiments have led to understanding fundamental mechanisms of auditory processing such as masking and lateral suppression, and have also been directed to deriving basic scales of loudness (Sect. 30.1.4, *Loudness Scaling* ...). A critical aspect is that *ecological validity* might be limited as lab experiments for a number of reasons are often restricted to a small number of variables and also employ, in most cases, simple sound stimuli that are far from the complexity of environmental sounds. In this respect, there is a gap between the lab and real-life situations. However, fundamental research in psychoacoustics can provide a basis to start from for investigations that are more contextual as is necessary in regard to music perception. The study of loudness is a good case in point. While many of the lab experiments used steady SPL for sounds (pure tones, noise bands) to find quantitative relations between physical magnitudes (intensity, power) and sensations of loudness, music is intrinsically time-variant both in spectral energy distribution (Sect. 32.2.3) and in dynamics. In reality, many if not most of the tones played on natural instruments undergo some AM and FM (frequency modulation) either for physical reasons or as an effect of playing technique. Also, several tones often interact in dissonant chords or sonorities so that AM is observed in the time signal, which can cause a sensation of roughness as well as a sensation of fluctuant loudness if the modulation rate is low enough for the auditory system to follow the changes in level. The factor that influences loudness sensation apparently is not the modulation index but the difference between maximum and minimum level. For broadband noise, the level of a modulated signal (with modulation frequencies from 4 to 32 Hz) needs to be ≈ 1 dB below that of the unmodulated noise of corresponding bandwidth to be judged as equally loud [33.56]. If several or even many harmonic partials are phase locked in a sound (Fig. 31.6a), it appears louder than a complex comprising the same number of components yet in random phase [33.57]. Greater loudness of a harmonic complex with all partials locked in phase can be attributed to the absolute height of amplitude peaks as well as to the crest factor. For example, five partials from $f_1 = 100$ Hz locked in cosine phase (Fig. 30.10) yield a peak amplitude that is larger than that of a complex having the same frequency components and amplitudes but where phase angles of components are chosen at random. The crest factor apparently also plays a role in loudness sensation; noise signals were judged as equally loud as a standard when their sound level decreased while the crest factor increased [33.56], indicating that the *peakiness* of signals can add to loudness sensation. This is particularly relevant for electronic dance music and related genres where, typically, the beat is marked by sequences of strong pulses. As is evident from Fig. 33.1, the rise time of these pulses can be quite short so that the sound jumps to peak level with a steep pulse slope. Such steep pulses that reach high peak levels can be expected to cause rapid transfer of energy into the auditory system where sections of the BM and the structures it carries are accelerated accordingly. In discotheques, visitors are exposed to high sound levels of 110–125 dB(C) and to music which, typically, is based on such pulse sequences that are realized, for the most part, with sounds in the low frequency range. In effect there is massive concentration of pulsed energy in particular in low AF/CBs that can affect hearing temporarily or permanently [33.53, 54]. According to a hypothesis advanced by *Todd* [33.58, 59], the vestibular system becomes involved in hearing at very loud sounds, which can be explained in evolutionary terms (as a remnant from early vertebrate hearing systems) and might explain why young people like sound levels above the *rock 'n' roll threshold* of ≈ 110 dB(C). In a large sample of 423 grammar school pupils (age: 16–19 yr, city: Berlin) almost 72% declared that they preferred loud, very loud or even extremely loud levels in discotheques [33.60, 61]. Many of the pupils also asserted that they would listen to pop and rock music with earphones at rather loud levels. Though the intake of considerable quantities of energy into the auditory system during visits to discos or at live rock concerts as well as by using earphones (or earplugs) is dangerous with respect to hearing loss, subjects seem to respond differently to loud or even very loud sound. Apparently, musical preferences and other personal factors influence the range of SPLs that subjects can tolerate or even find appropriate for listening.

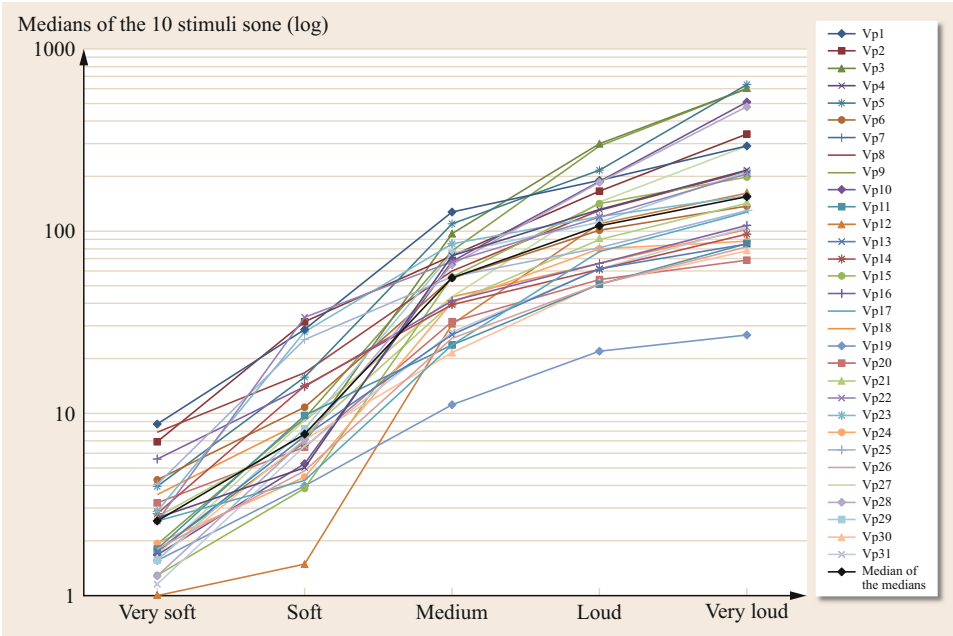


Fig. 33.6 Individual loudness functions for 31 subjects averaged over each subject's responses to ten stimuli (after [33.62], with permission from A. von Ruschkowski). The levels in sone (ordinate) are in a log scale to accommodate the wide range of levels chosen in loudness judgments. The five levels (abscissa) are *very soft*, *soft*, *medium*, *loud*, *very loud*

In an empirical study of loudness with musical stimuli covering various genres, *von Ruschkowski* [33.62] had subjects make loudness judgments by adjusting the level with a continuously variable level control according to five categories (very soft, soft, medium, loud, very loud). The signal was fed into two headsets (Sennheiser HDA 200) constructed to match the very narrow tolerances needed for audiometry, and capable of handling high input without distortion; one of the headsets was used by the 31 musically trained subjects volunteering in the experiment (all tested with audiometry for normal hearing), the other was attached to an artificial head (HSU III.2, Head Acoustics), which housed two calibrated measuring microphones (Microtech Gefell MK 250) from where the signal was continuously recorded in digital format at 24 bit/48 kHz. These data could be used to calculate SPL over time (with or without weightings) as well as loudness over time (with a choice of several methods and scales). In addition to level adjustments, four variables were included in the design (gender, age, musical preferences, current mental state and mood). The last variable was checked with a standardized psychological test (Baseler Befindlichkeits-Skala, BBS).

The results from this study [33.62] showed that responses of subjects were highly variable, resulting in broad dispersions of data and large standard deviations (SD). Though subjects in general realize a category scale from *very soft* to *very loud* as well as a monotonous loudness function, absolute values for the categories vary considerably among subjects (for instance, the range for *very loud* for 30 subjects out of 31 is from 70 sone to 600 sone; Fig. 33.6) and also with the individual stimuli. Averaged over the sample of subjects, the category *very loud* for the Adagio from Beethoven's piano sonata Nr. 14 (op. 27, no. 2, *Moonshine*) means about 100 sone (or a loudness level of $L_N = 106.44$ phon) while the same category for a thrash metal production (Slipknot, *People = shit*) means $\approx$ 227 sone ($L_N = 118.26$ phon). Besides level adjustment the variable gender did yield significant results (women prefer music at lower levels than men); further, the variable musical preferences showed a correlation with level settings.

The enormous variability in level adjustments can be partly attributed to stimulus parameters (type of music and composition, temporal and spectral structure, dynamics, production including sound effects, compression rate applied to dynamics, etc.) and partly to musical preferences and habits. Though sound pressure level (measured in mPa) certainly is the strongest factor in human loudness perception, there are evidently several personal and social factors involved as well when it comes to loudness evaluations of real music. The variability in loudness judgments, which is high when musical stimuli are presented to subjects who have different social and musical backgrounds, might indicate that loudness depends significantly on *context* (what it apparently does). However, variability in response magnitudes was observed already by Stevens and other psychophysicists in a host of lab experiments where elementary stimuli were used (Sect. 30.1.4, *Loudness*

Scaling ...). The great variability found in subjective responses to sound stimuli measured as loudness is clearly different from pitch where judgments based on sensations in general vary within much narrower limits. Though the dependence of loudness sensation on SPL or intensity is obvious, one would expect much less variability in empirical data if sound pressure and energy would condition loudness sensation in total. Rather, one needs to look into the spectral content and temporal structure of sound signals to identify features that can explain why sounds having equal energy are often perceived as different in regard to loudness. Besides sound signal characteristics, aspects discussed for intensity sensation and loudness also included the possibility of memory traces and of internal anchor points as well as decision processes mediating intensity discrimination and identification ([33.20] and a number of subsequent articles leading to [33.63]). One of the considerations was that internal *low* and *high* anchor points might narrow the dynamic range subjects actually relate to. It has in fact been observed [33.62] that some subjects, when asked to adjust sound levels so as to match five categories (from *very soft* to *very loud*, see above) do not use the full range one might expect (where *very soft* could mean ≈ 30 dB SPL and *very loud* perhaps ≈ 100 dB), but a much smaller range. Such behavior probably reflects individual experience since, for example, subjects practicing chamber music usually dislike high sound levels and may find a level of 70 or 80 dB *very loud* while subjects playing in a hard rock or metal band may regard 60 dB *very soft* and only 120 dB *very loud*. Subjective narrowing of the dynamic range hence should not be confused with the well-known bias according to which subjects tend to avoid extreme categories at both ends of rating scales. Rather, subjects seem to use loudness categories of different individual bandwidth (dB, sone) that are placed at different absolute levels (mPa, dB) probably serving as anchor points. In sum, at least musically experienced subjects distinguish between *very soft* and *soft*, or between *loud* and *very loud* sound levels, but the category centers and boundaries vary considerably along the dynamic range (0−120 dB), for individual ratings.

Another concept offered as an alternative to conventional models relating stimulus intensity to loudness was the physical correlate theory proposed by *Warren* [33.64, 65]. He pointed to the relation between perceived loudness and imagined distance of a sound source. In this respect, loudness judgments involve distance estimates since in a natural environment the loudness of a sound source can convey its distance relative to a perceiving subject that may react in his or her behavior accordingly. The physical correlate theory was proposed to explain doubling and halving of loudness (Sect. 30.1.4, *Loudness Scaling* ...) in terms of imagined distance judgments; it takes sound field characteristics into account since the proportion of direct sound to diffused (reverberant) sound changes with distance. *Warren*'s [33.64] finding that half-loudness for speech signals (−11 dB with headphones, −12 dB with loudspeakers) differed considerably from half-loudness for tone and steady-state noise signals (about −6 dB) can be taken as evidence that a single loudness scale (like the sone scale) might not be suited to represent various types of sound signals. Notwithstanding substantial evidence from both neural intensity coding studies and behavioral experiments, a full understanding of loudness sensation and perception is still an objective for research.

33.4 Summing up

This section briefly sums up some of the content as well as of the critical issues discussed in Chaps. 30–33 and concludes with a tentative model concerning relationships between pitch, timbre, and loudness.

This part of the handbook, covering fundamentals of psychophysics and psychoacoustics, opens with Chap. 30 introducing theoretical and methodological considerations in regard to sensation and perception as well as apperception. Reasons are given for an approach to sensation and perception of sound based on realism and naturalism seeking causal explanations of empirical observations. Psychophysical scaling is among the issues and problems treated in greater detail. For better understanding (and provision of some historical background for readers from areas such as musicology and music education), the development of several scale concepts has been outlined and illustrated by examples (in particular, the sone scale and the mel scale).

Subsequent sections on sensation and perception of pitch (Chap. 31) repeatedly refer to the principles of periodicity and spectral harmonicity, with special emphasis on the Wiener–Khintchine theorem. In line with findings from neurophysiological research of the past decades, pitch and also consonance are viewed as natural phenomena reflecting organization and functions of the auditory periphery and higher stations of the AuP as much as the lawful structure of sound. A basically naturalistic and empiricist perspective, which seems es-

sential for psychoacoustics, does by no means preclude regard for historical and sociocultural contexts (as was underpinned already by *von Helmholtz* [33.66]). Rather, it can be useful if not inevitable to take historical as well as ethnological and other cultural background information into account (as has been pursued in several sections of this part, for example, the sections on tone systems and scale formation, Sect. 31.8). The aim in this respect is to illustrate certain developments both in music and in research pertaining to musical acoustics and psychoacoustics or music perception. A rather critical view has been taken on some concepts of *categorical perception* based on ET12 *pitch class* constructs as applied to pitch perception and intonation practice in music. Also, some interpretations of tonal *fusion* found in the literature do not correctly represent the original concept of *Verschmelzung* as outlined by *Stumpf* [33.67–69]. Further, a general prevalence of musicians and listeners for the *stretched octave* (a commonality reiterated time and again in the literature) has been questioned on the basis of contrary evidence. In regard to so-called *auditory illusions*, the scale type constructed by *Shepard* [33.70, 71] using octave-spaced components circling under a fixed symmetric envelope (so that the sum of their amplitudes remains the same for all tones and scale steps) has been shown to realize a well-known experimental paradigm, namely keeping one parameter constant (in this case, tone height or brightness) while varying another (in this case, tone quality or *chroma*). Hence listeners perceive a circle of tone qualities presented with constant brightness. The construct becomes more of an illusion when the envelope is shifted contrary-wise to the movement of the octave-spaced components forming complex tones. Such constructs have been incorporated into works of contemporary music [33.72]. A discussion of the so-called *tritone paradox* (also based on Shepard tones, though on a version critically reduced in the number of components) has not been included into Chap. 31, for several reasons, among them being the controversial explanations the *tritone paradox* has found in regard to linguistic and other issues [33.73, 74]. Moreover, reanalyses of empirical data from earlier experiments that had led to postulating a *tritone paradox* and replications of experiments demonstrated that stimuli and conditions in previous experiments and their interpretations were not consistent or that alternative interpretations centered on sound features might be more apt [33.75, 76].

The sections on timbre and sound color (Chap. 32) perhaps differ from other accounts in that the interrelations with pitch are accentuated even though there are features distinctive of timbre that call for systematic analysis and description. The approach taken here is to discuss main concepts and issues, and to illustrate certain phenomena by sound analyses. Since timbre is a truly multifaceted area, any attempt to offer an exhaustive treatment of timbre within a single chapter would be all but impossible. In regard to music as performed and recorded, the overarching category of *sound*, which comprises aspects of sensation and perception along with factual evidence from room acoustics, studio technology, and music production, is briefly alluded to but not treated in much detail (for more background information, see [33.77–80]). Fundamentals of auditory stream segregation have been included in the sections on timbre; however, this area of research meanwhile has greatly expanded into modeling based on signal processing so that an in-depth survey of methodology and results would require an extra chapter.

The sections on intensity sensation and loudness perception (Chap. 33) were written with the goal of summing up at least part of the relevant observations and discussions while the number of publications on loudness over the past 70–80 yr is immense (comprehensive surveys are included in several chapters of [33.55]). However, a considerable portion of the literature is technical in content since studies on loudness include aspects of noise control and risks of hearing loss as well as the norms and standards of measurement. The relation of physical parameters and magnitudes to sensation apparently is even more complex for loudness than it is for the psychoacoustic domain of *pitch*. Though it is obvious that local sound pressure and sound intensity measurable in physical units are fundamental parameters for the excitation of parts in the auditory system (the mammalian ear, with some possible participation also of the vestibular system), measurement of sensation actually seems quite demanding in particular for real-life broadband sound such as music. Scales for loudness level L_N (phon) and loudness N (sone) derived from elementary stimuli in the lab probably will not cover more complex (in terms of spectral composition and temporal modulation) sound material. In addition, it seems that some of the norms of measurement even of sound level are inadequate given the actual temporal and spectral structure of musical sound. For example, it is astounding to see that an outdated weighting function such as dB(A) is still applied even in many measurements of high-level sounds where, in a techno disco or at a festival of heavy metal music, huge amounts of spectral energy is found concentrated in low frequency bands ($\approx$ 50–200 Hz). Applying dB(A) weighting to techno sound or heavy metal (or to other genres of highly amplified rock and pop) in fact means high-pass filtering where a significant portion of the energy contained in low frequency bands is suppressed. Hence, such a measurement hides the real distribution of spectral energy

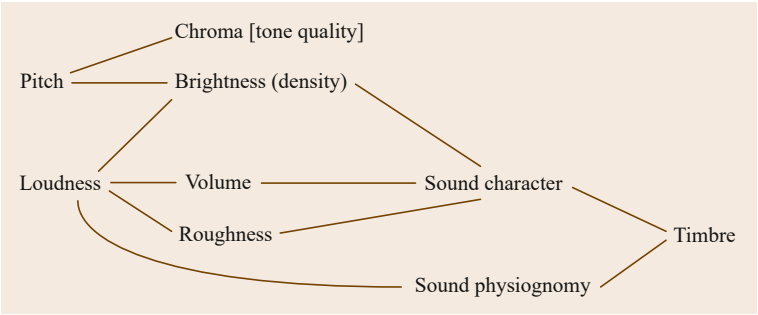


Fig. 33.7 Interrelations of pitch, loudness and timbre (after [33.81, 82])

radiated from a public address system and is unsuited for an assessment of possible risks of hearing loss. Among the temporal patterns of modern rock and pop music genres, the use of pulsed sounds is a feature that is clearly of relevance to sensation. Pulse-like sounds as are available from synthesizers as well as by *shaping* the envelope of natural sounds (by means of noise gates, filters and compressors) seem of particular interest in regard to energy transfer and BM excitation characteristics. However, in particular pulse-like sound structures are still difficult to asses with respect to effects they have on hearing since one needs to determine suitable integration constants as well as parameters covering the *peakiness* in sounds [33.83].

Finally, though research efforts in special areas such as pitch perception and loudness sensation will be continued, interrelations between pitch, timbre, and loudness should be studied as well since these are relevant for sensation and perception of music as performed and recorded. A scheme outlined already by *Hesse* [33.81] and *Thies* [33.82] from theoretical and empirical studies connects pitch, loudness, and timbre and sums up a number of attributes under the categories of *sound character* and *sound physiognomy* (Fig. 33.7, adapted from *Thies* [33.82, p. 21]).

Pitch in this scheme is largely a function of the period (f_0) of the complex envelope while *sound character* depends on the microstructure of a sound wave. *Sound physiognomy* [33.81, p. 139 f.] means the subjective experience of the sound profile in terms of the onset and the transients of a sound and the temporal envelope in general. Thus, perception of timbre can be explained as resulting from information pertaining to the interaction of the temporal and the spectral envelope, which must be understood as enclosing dynamic energy distributions that vary considerably with time in their structure. In binaural listening (e.g., to music performed in a concert hall or live music venue), spectral energy distribution and temporal envelopes can be quite different at both ears. While input is analyzed in both channels of the AuP, there are several links connecting left and right channel (Fig. 31.2), in particular SOC and inferior colliculus (IC) so that information is integrated for processing pitch, timbre, and loudness at higher stations (IC, corpus geniculatum mediale (CGM), cortical areas).

References

33.1 E. Shaw, M. Vaillancourt: Transformation of sound pressure level from the free field to the eardrum in numerical form, J. Acoust. Soc. Am. **78**, 1120–1123 (1985)

33.2 D. Keefe, J. Bulen, K. Hoberg Arehart, E. Burns: Ear-canal impedance and reflection coefficient in human infants and adults, J. Acoust. Soc. Am. **94**, 2617–2638 (1993)

33.3 S. Voss, J. Allen: Measurement of acoustic impedance and reflectance in the human ear canal, J. Acoust. Soc. Am. **95**, 372–384 (1994)

33.4 R. Aibara, J. Welch, S. Puria, R. Goode: Human middle-ear transfer function and cochlear input impedance, Hearing Res. **152**, 100–109 (2001)

33.5 J. Rosowski: The effects of external and middle-ear filtering on auditory threshold and noise-induced hearing loss, J. Acoust. Soc. Am. **90**, 124–135 (1991)

33.6 H. Fletcher, W. Munson: Loudness, its definition, measurement and calculation, J. Acoust. Soc. Am. **5**, 82–108 (1933)

33.7 W. Rhode, A. Recio: Study of mechanical motions in the basal region of the chinchilla cochlea, J. Acoust. Soc. Am. **107**, 3317–3332 (2000)

33.8 I. Russell, K. Nielsen: The location of the cochlear amplifier: Spatial representation of a single tone on the guinea pig basilar membrane, Proc. Nat. Acad. Sci. **94**, 2660–2664 (1997)

33.9 M. Ruggero, N. Rich, A. Recio, S. Narayan, L. Robles: Basilar-membrane responses to tones at the base of the chinchilla cochlea, J. Acoust. Soc. Am. **101**, 2151–2163 (1997)

33.10 C. Shera: Intensity-invariance of fine time structure in basilar-membrane click responses: Implications for cochlear mechanics, J. Acoust. Soc. Am. **110**,

332–348 (2001)
33.11 S. Gelfand: *Hearing. An Introduction to Psychological and Physiological Acoustics*, 4th edn. (Dekker, New York 2004)
33.12 E. Zwicker, H. Fastl: *Psychoacoustics, Facts and Models*, 2nd edn. (Springer, Berlin 1999)
33.13 G. Yates, I. Winter, D. Robertson: Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range, Hearing Res. **45**, 203–219 (1990)
33.14 B. Moore: Frequency analysis and masking. In: *Hearing*, ed. by B. Moore (Academic, San Diego 1995) pp. 161–205
33.15 E. Relkin, J. Doucet: Is loudness simply proportional to the auditory nerve spike count?, J. Acoust. Soc. Am. **101**, 2735–2740 (1997)
33.16 J. Doucet, E. Relkin: Neural contributions to the peristimulus compound action potential: Implications for measuring the growth of the auditory nerve spike count as a function of stimulus intensity, J. Acoust. Soc. Am. **101**, 2720–2734 (1997)
33.17 J.O. Pickles: *Introduction to the Physiology of Hearing*, 3rd edn. (Emerald, Binkley 2008)
33.18 Chr Plack, R. Carlyon: Loudness perception and intensity coding. In: *Hearing*, ed. by B. Moore (Academic, San Diego 1995) pp. 123–160
33.19 St Uppenkamp, M. Röhl: Human auditory neuroimaging of intensity and loudness, Hearing Res. **307**, 65–73 (2014)
33.20 N. Durlach, L. Braida: Intensity perception. I. Preliminary theory of intensity resolution, J. Acoust. Soc. Am. **46**, 372–383 (1969)
33.21 R. Schlauch: Loudness. In: *Ecological Psychoacoustics*, ed. by J. Neuhoff (Elsevier, San Diego 2004) pp. 317–345
33.22 L. Marks: Binaural summation of the loudness of pure tones, J. Acoust. Soc. Am. **64**, 107–113 (1978)
33.23 V. Sivonen, W. Ellermeier: Binaural loudness. In: *Loudness*, ed. by M. Florentine (Springer, New York 2011) pp. 169–197
33.24 J. Marozeau, M. Epstein, M. Florentine, B. Daley: A test of the binaural equal-loudness-ratio hypothesis for tones, J. Acoust. Soc. Am. **120**, 3870–3877 (2006)
33.25 M. Florentine, M. Epstein: To honour Stevens and to repeal his law. In: *Fechner Day 2006. Proc. 22nd Annu. Meet. Int. Soc. Psychophys.*, ed. by D. Kornbrot, R. Msetfi, A. MacRae (Univ. of Hertfordshire Press, Hatfield 2006) pp. 37–41
33.26 H. Fletcher: Loudness, masking and their relation to the hearing process and the problem of noise measurement, J. Acoust. Soc. Am. **9**, 275–293 (1938)
33.27 H. Bauch: Die Bedeutung der Frequenzgruppe für die Lautheit von Tönen, Acustica **6**, 40–45 (1956)
33.28 E. Zwicker, G. Flottorp, S. Stevens: Critical bands with loudness summation, J. Acoust. Soc. Am. **29**, 548–557 (1957)
33.29 E. Zwicker, B. Scharf: A Model of loudness summation, Psych. Rev. **72**, 3–26 (1965)
33.30 W. Hartmann: *Signals, Sound and Sensation* (Springer, New York 1998)
33.31 E. Zwicker, E. Terhardt: Analytical expressions for critical-band rate and critical bandwidth as a function of frequency, J. Acoust. Soc. Am. **68**, 1523–1525 (1980)
33.32 B. Moore, B. Glasberg: Suggested formulae for calculating auditory-filter bandwidths and excitation patterns, J. Acoust. Soc. Am. **74**, 750–753 (1983)
33.33 H. Fletcher: Auditory patterns, Rev. Mod. Phys. **12**, 47–66 (1940)
33.34 A. Schneider, V. Tsatsishvili: Perception of intervals at very low frequencies: Some experimental findings. In: *Systematic Musicology: Empirical and Theoretical Studies*, ed. by A. Schneider, A. von Ruschkowski (Lang, Frankfurt 2011) pp. 99–125
33.35 R. Meddis, M. Hewitt: Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification, J. Acoust. Soc. Am. **89**, 2866–2882 (1991)
33.36 R. Meddis, M. Hewitt: Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity, J. Acoust. Soc. Am. **89**, 2883–2894 (1991)
33.37 R. Patterson, K. Robinson, J. Holdsworth, D. McKeown, C. Zhang, M. Allerhand: Complex sounds and auditory images, Adv. Biosci. **83**, 429–443 (1992)
33.38 R. Patterson, M. Allerhand, C. Giguère: Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform, J. Acoust. Soc. Am. **98**, 1890–1894 (1995)
33.39 E. Terhardt: *Akustische Kommunikation* (Springer, Berlin 1998)
33.40 B. Glasberg, B. Moore: Derivation of auditory filter shapes from notched noise data, Hearing Res. **47**, 103–113 (1990)
33.41 B. Moore: Basic psychophysics of human spectral processing. In: *Auditory Spectral Processing*, Intern. Rev. Neurobiol., Vol. 70, ed. by M. Malmierca, D. Irvine (Elsevier, Amsterdam 2005) pp. 49–86
33.42 M. Slaney: An efficient implementation of the Patterson–Holdsworth auditory filter bank, Apple Comput. Tech. Report **35** (1993)
33.43 T. Lin, J. Guinan: Auditory nerve fiber responses to high-level clicks: Interference patterns indicate that excitation is due to the combination of multiple drives, J. Acoust. Soc. Am. **107**, 2615–2630 (2000)
33.44 E. Zwicker: *Psychoakustik* (Springer, Berlin 1982)
33.45 T. Poulsen: Loudness of tone pulses in a free field, J. Acoust. Soc. Am. **69**, 1786–1790 (1981)
33.46 B. Glasberg, B. Moore: A model of loudness applicable to time-varying sounds, J. Audio Eng. Soc. **50**, 331–342 (2002)
33.47 J. Chalupper, H. Fastl: Dynamic loudness model (DLM) for normal and hearing-impaired listeners, Acustica **88**, 378–386 (2002)
33.48 R. Sottek: A hearing model approach to time-varying loudness, Acustica **102**, 725–744 (2016)
33.49 J. Hots, J. Rennies, J. Verhey: Modelling temporal integration of loudness, Acustica **100**, 184–187 (2014)
33.50 J. Rennies, M. Wächtler, J. Hots, J. Verhey: Spectro-temporal characteristics affecting the loudness of technical sounds: Data and model predictions,

Acustica **101**, 114–1156 (2015)
33.51 J. Rennies, J. Verhey, H. Fastl: Comparison of loudness models for time-varying sounds, Acustica **96**, 383–396 (2010)
33.52 B. Moore, B. Glasberg: Modeling binaural loudness, J. Acoust. Soc. Am. **121**, 1604–1612 (2007)
33.53 A. Schneider, A. von Ruschkowski: Techno, decibels, and politics: An empirical study of modern dance music productions, sound pressure levels, and loudness perception. In: *Systematic Musicology: Empirical and Theoretical Studies*, ed. by A. Schneider, A. von Ruschkowski (Lang, Frankfurt/M. 2011) pp. 13–62
33.54 A. von Ruschkowski, A. Schneider: Schallstruktur und potentielle Risiken für das Gehör: Eine empirische Studie in einer Hamburger Diskothek, Z. Audiol. **51**, 115–121 (2012)
33.55 M. Florentine (Ed.): *Loudness* (Springer, New York 2011)
33.56 G. Grimm, V. Hohmann, J. Verhey: Loudness of fluctuating sounds, Acta Acust. united with Acust. **88**, 359–368 (2002)
33.57 H. Gockel, B. Moore, R. Patterson: Influence of component phase on the loudness of complex tones, Acustica **88**, 369–377 (2002)
33.58 N. Todd, F. Cody: Vestibular responses to loud dance music: A physiological basis of the "rock'n'roll threshold"?, J. Acoust. Soc. Am. **107**, 496–500 (2000)
33.59 N. Todd: Evidence for a behavioural significance of saccular arousal sensitivity in humans, J. Acoust. Soc. Am. **110**, 380–390 (2001)
33.60 W. Babisch, B. Bohn: *Schallpegel in Diskotheken und bei Musikveranstaltungen. Part II* (Umweltbundesamt, Berlin 2000)
33.61 W. Babisch, B. Bohn: *Schallpegel in Diskotheken und bei Musikveranstaltungen. Part III* (Umweltbundesamt, Berlin 2000)
33.62 A. von Ruschkowski: *Lautheit von Musik. Eine empirische Untersuchung zum Einfluss von Organismusvariablen auf die Lautstärkewahrnehmung von Musik*, Ph.D. Thesis (Univ. of Hamburg, Systematic Musicology, Hamburg 2013), http://ediss.sub.uni-hamburg.de/volltexte/2014/6576/
33.63 L. Braida, J. Lim, J. Berliner, N. Durlach, W. Rabinowitz, S. Purks: Intensity perception. XIII. Perceptual anchor model of context coding, J. Acoust. Soc. Am. **76**, 722–731 (1984)
33.64 R.M. Warren: Subjective loudness and its physical correlate, Acustica **37**, 334–346 (1977)
33.65 R. Warren: *Auditory Perception. An Analysis and Synthesis*, 3rd edn. (Cambridge Univ. Press, Cambridge 2008)
33.66 H. von Helmholtz: *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (Vieweg, Braunschweig 1863), 3rd edn. 1870, 6th edn. 1913
33.67 C. Stumpf: *Tonpsychologie*, Vol. 1 (Barth, Leipzig 1883)
33.68 C. Stumpf: *Tonpsychologie*, Vol. 2 (Barth, Leipzig 1890)
33.69 C. Stumpf: *Die Sprachlaute* (Springer, Berlin 1926)
33.70 R. Shepard: Circularity in judgements of relative pitch, J. Acoust. Soc. Am. **36**, 2346–2353 (1964)
33.71 R. Shepard: Demonstrations of circular components of pitch, J. Audio Eng. Soc. **31**, 641–649 (1983)
33.72 J.-C. Risset: Computer, synthesis, perception, paradoxes, Hamburger Jahrb. Musikwiss. **11**, 245–258 (1991)
33.73 B. Repp: The tritone paradox and the pitch range of the speaking voice: A dubious connection, Music Percept. **12**, 227–255 (1994)
33.74 D. Deutsch: The tritone paradox and the pitch range of the speaking voice: Reply to Repp, Music Percept. **12**, 257–263 (1994)
33.75 B. Repp: Spectral envelope and context effects in the tritone paradox, Perception **26**, 645–665 (1997)
33.76 C. Friedrich: *Die Ambivalenz der Tonhöhenwahrnehmung des Tritonus. Eine empirische Studie basierend auf der Zweikomponenten-Theorie der Tonhöhe*, MA Thesis (Univ. Hamburg, Hamburg 2006)
33.77 D. Howard, J. Angus: *Acoustics and Psychoacoustics*, 2nd edn. (Focal, Oxford 2001)
33.78 A. Schneider: Klanganalyse als Methodik der Popularmusikforschung, Hamburger Jahrb. Musikwiss. **19**, 107–129 (2002)
33.79 A. Schneider: Komposition und Produktion von "U-Musik" unter dem Einfluss technischer Medien. In: *Handbuch Musik und Medien*, ed. by H. Schramm (UVK Verlagsgesellschaft, Konstanz 2009) pp. 495–530
33.80 D. Huber, R. Runstein: *Modern Recording Techniques*, 8th edn. (Focal, Oxford 2013)
33.81 H.P. Hesse: *Die Wahrnehmung von Tonhöhe und Klangfarbe als Probleme der Hörtheorie* (Gerig, Köln 1972)
33.82 W. Thies: *Grundlagen einer Typologie der Klänge* (Wagner, Hamburg 1982)
33.83 G. Goley, W. Song, J. Kim: Kurtosis-corrected sound pressure level as a noise metric for risk assessment of occupational noises, J. Acoust. Soc. Am. **129**, 1475–1481 (2011)

Part E

Music Em

Part E Music Embodiment

Ed. by Marc Leman

One of the main goals of systematic musicology is to understand the effects of musical (acoustical) structure on human behavior. Throughout the 20th century, this understanding was approached by considering music perception and its effect on behavior. That approach was strongly influenced by developments in the Gestalt psychology of the 1920s, the information psychology and cybernetics of the 1960s, and especially the empirical and computational approaches of the cognitive sciences since the 1980s. However, thanks to the advent of technologies that facilitated the tracking of human movement (with video cameras and sensors) it became possible to focus on new areas of music interaction as well, such as music performance and musical gesturing (while listening). These new areas of study started to emerge in the late 1990s (despite some historical antecedents in the 1920s) and they culminated in a new approach that is nowadays known as *embodied music cognition*.

Embodied music cognition puts emphasis on the human body as a contributing factor in listening, dancing and playing music. The former focus on music perception has therefore been replaced by a focus on music-related action. Understanding the effect of music on human behavior is approached from an embodied perspective. Music perception is also now reconsidered from this perspective. The rationale is that the perception of music is influenced by the way in which music is interacted with, and these interactions influence how musical structures are perceived. But these interactions depend on bodily constraints, such as physical, biomechanical, and biological possibilities and restrictions of the human body.

Embodied music cognition is in fact a natural outcome of the cognitive research paradigm that characterized the systematic musicology of the late 20th century. Its emphasis on action, and on the principles that support action (such as motor control) has broadened the concept of cognition, and this broad perspective can be considered as a major characteristic of the new approach. The rapid acceptance of this new approach by the music research community is probably due to the explanatory and predictive power of the embodiment concept; to its ability to cope with music perception as well with music performance; to its potential to generate new types of technological applications that involve both action and perception in domains such as music and wellbeing, music engineering, and neuroscience of music; and last but not least, to its influence in ethnomusicology, where studies of social interaction draw upon research results from embodied interactions with music. Embodied music cognition has rapidly evolved in the direction of a new embodied music interaction paradigm. In the near future, it is likely that the term *cognition* will disappear and be replaced by the term *interaction*. Essentially, it means that the effects of music on human behavior are based on interactive processes that involve perception *and* action.

The chapters of this section reflect the work in progress in this branch of systematic musicology. Despite its rapid acceptance by the community, embodied music cognition is still a very young research paradigm. Its epistemological and methodological foundations have barely been addressed. Its basic concept of embodiment needs further refinement in view of new findings. Its support by empirical studies requires further substantiation. Its technological basis requires a constant update with the rapid ongoing developments of technology in our society. The chapters introduce basic concepts and basic methods, and they focus on different application areas of interactions, such as music making and wellbeing.

Chapter 34 (*What Is Embodied Music Cognition*, by *Marc Leman*, *Luc Nijs*, *Pieter-Jan Maes*, *Edith Van Dyck*) explains what embodied music cognition is about. The chapter gives an overview of the major concepts behind this research paradigm of systematic musicology, including its basic ontology and epistemology, and the architecture of embodiment (involving concepts such as prediction, emergency, enactment and expression). This basis is then followed by an overview of some analytical and empirical studies, which illustrate contributions of the embodied music cognition approach to the understanding of expressive gestures, synchronization and entrainment, and the effects of action on perception.

Chapter 35 (*Sonic Object Cognition*, by *Rolf-Inge Godoy*) goes deeper into the nature of the musical material that forms the basis of human interaction with music. This material is described in terms of sonic objects. Starting from the work of Pierre Schaeffer and others, the chapter considers the nature of sonic objects in relation to human interaction behavior. A major breakthrough in our understanding of sonic objects was realized when sonic objects were associated with body motions, shape cognition and actions. This new way of understanding sonic objects offers new possibilities for sound synthesis control, and new possibilities for interaction.

Chapter 36 (*Investigating Embodied Music Cognition for Health and Wellbeing*, by *Micheline Lesaffre*) provides a framework that aims to validate musicological know-how in applications for health and wellbeing. The chapter defines the field of health and wellbeing and ongoing work in music therapy. It then explains that

the unique contribution of the embodied music cognition approach is related to user-driven and technology-driven research in which measurement and analysis of embodied interaction with music offers a path to evidence-based validation.

Chapter 37 (*A Conceptual Framework for Music-Based Interaction Systems*, by *Pieter-Jan Maes*, *Luc Nijs*, and *Marc Leman*) looks at the role of technology in music interaction. It is claimed that interactive systems can reinforce interactions with music, such that the overall experience with music is intensified, and the social, cognitive, affective, and motor skills are made stronger. The chapter introduces an overall model for music interaction based on an embodied music cognition framework that also includes the benefits of reward processing and motivation, and social interaction. Examples are given of an educational technology for learning to play a music instrument, and a technology that facilitates the synchronization of human movement to music.

Chapter 38 (*Methods for Studying Music-Related Body Motion*, by *Alexander Refsum Jensenius*) presents an overview of methods for music-related motion description. The focus is on qualitative and quantitative description methods that prepare the ground for further analysis of human-music interaction behaviour, such as methods for describing qualitative motion, or methods for describing motion features such as quantity and centroid. A distinction is made between camera-based systems and sensor-based systems. All technologies have strengths and weaknesses and the right choice depends on the research question and analysis methods.

34. What Is Embodied Music Cognition?

Marc Leman, Pieter-Jan Maes, Luc Nijs, Edith Van Dyck

Over the past decade, embodied music cognition has become an influential paradigm in music research. The paradigm holds that music cognition is strongly determined by corporeally mediated interactions with music. They determine the way in which music can be conceived in terms of goals, directions, targets, values, and reward. The chapter gives an overview of the ontological and epistemological foundations, and it introduces the core concepts that define the character of the paradigm. This is followed by an overview of some analytical and empirical studies, which illustrate contributions of the embodied music cognition approach to major topics in musical expression, timing, and prediction processing. The chapter gives a viewpoint on a music research paradigm that is in full development, both in view of the in-depth refinement of its foundations, as well as the broadening of its scope and applications.

In this chapter, we focus on the quintessence of the embodied music cognition paradigm; namely, its claim that bodily involvement is crucial in human interaction with music, and therefore, also in our understanding of that interaction. The embodied viewpoint holds that bodily involvement shapes the way we perceive, feel, experience, and comprehend music. Embodiment determines, to a large extent, why sound is experienced as music, which has a rewarding nature and provokes personal interest. But what does bodily involvement, and embodiment, really mean? What are the main assumptions, the findings, the perspectives? Are these assumptions so different from the classical viewpoint on music cognition? And what are the consequences for our understanding of music, or for future music research?

Critics could claim that nothing in the paradigm of (classical) music cognition ever claimed that perceiving music does not involve movement [34.1]. Indeed, several descriptive studies show, for example, that the expression of music is reflected in body responses (locomotion, arousal), or that expression is present in gestures that support music playing. However, the descriptive approach may fail short in adequately justifying a theory of embodied musical meaning formation. What is needed, therefore, are *more direct proofs* of the influence of embodiment on perception and on meaning formation, such as proof or evidence that music perception is determined by states of embodiment: movement states, emotional states and so on. Do such studies currently exist? Do the studies only navigate and reach beyond showing that music induces movements? Moreover, what does this imply?

What follows is an exposition of a viewpoint on the foundations of embodied music cognition, which focuses on the above questions. In the first part, we define the theoretical framework of embodied music cognition. We start with a specification of the ontolog-

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_34

ical and epistemological foundations. Then we delve deeper into some of the core concepts that define the character of the approach. In the second part, we discuss analytical and empirical studies (mainly sourced from our own laboratory) that illustrate the contribution of embodied research to topics in musical expression, timing and, especially, prediction. In the third part, we give a general appreciation of the paradigm and identify general trends and perspective for future research.

34.1 Ontological and Epistemological Foundations

The paradigm of embodied music cognition can be seen as an extension of an earlier paradigm, in which cognition was mainly considered from the viewpoint of perception. Focusing on *cognition in perception*, this paradigm treated the ability to perceive music in relation to memory, learning, and predictive processes. Music and musical processing were thereby dealt with in terms of structural features, or patterns, while emotions and movements were seen as outcomes of that kind of prediction-based pattern processing [34.1–4]. For example, the viewpoint was that a mismatch between particular perceptions and predictions would cause tensions that find their way towards bodily expression, emotion, and, ultimately, reward. Needless to say, perhaps, is that the cognitive approach to music perception has been formulated with reference to an impressive theoretical framework and viewpoint for understanding how music cognition works [34.5]. Importantly, this *cognition in perception* approach focused on predictive processes in perception, irrespective of possible bodily influences on that prediction. Recent scientific evidence, however, suggests that bodily involvement in perception cannot be ignored. For example, body movements may determine the perception of meter [34.6, 7], and bodily states may be central determinants of the musical interaction experience [34.8, 9]. Evidently, these and similar findings have serious consequences for the theoretical underpinning of music cognition research. After all, they indicate the importance of interactive processes in musical signification. Therefore, the embodied music cognition paradigm extends the *cognition in perception* approach, and replaces it with a *cognition in interaction* approach. This approach builds, to a certain extent, on achievements in previous music cognition research, but at the same time it reformulates the epistemological assumptions of music cognition. It is justifiable to say that it is traditional in its methodology, but radically different in its epistemology.

34.1.1 Assumptions About the World

Over the past decade, the concept of embodiment has become influential in music research [34.10, 11]. Embodied music cognition means that the cognitive processing of music (such as learning, memory use, prediction, etc.) is based on corporeally mediated interactions with music. Another way to approach this is to see it in terms of patterns and states, and the predictive models surrounding the connections between patterns and states [34.12–17]. Patterns are the elements that can be observed, while states are the conditions that lead to those patterns (Fig. 34.1). Predictive models of the human brain are specialized in making assumptions about those conditions, given the observed patterns. The mechanism works in fact in dual direction; namely, action and perception. First of all, humans learn that actions cause particular changes in the environment. Actions are supported by predictive models that associate motor commands for actions that have perceived sensations of the outcomes of those actions. Secondly, and parallel with the first, humans learn that particular changes in the environment can be understood in terms of actions, probably not their own action, but maybe the actions of others, or of other things. Perceptions are supported by predictive models that associate observed patterns as action outcomes in dual direction; namely, from actions to action outcomes and observed patterns, and from observed patterns to assumed conditions and actions that cause these patterns. A complication is that the owned states (such as intentional states, arousal states, energetic states, and so on) may influence the construction of these predictive models. The owned states are known through proprioceptive observations. The model of the owned state makes an assumption, therefore, about the true condition of the body, and

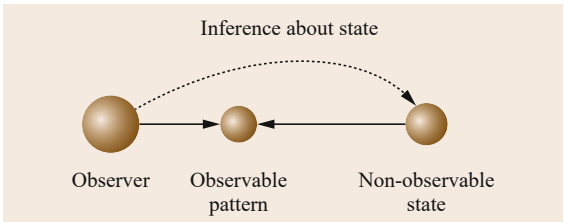


Fig. 34.1 States are conceived as causal conditions of being. Assumptions about states are based on patterns that can be observed

this assumption, or predictive model, may interfere with predictive models about the environmental conditions. In essence, the key point is that predictive models are not based on a direct line of information that informs a model about the true state of the world, nor even does it define the true state of the proper body. The true state remains unknown, and needs to be inferred based on information that its mediators provide. Stated differently, the human brain doesn't perceive the state of the world directly, but it constructs models of the state of the world using proprioceptive and exteroceptive corporeal mediators that provide the information for the models on the basis of interactions with that world.

What we call *music* can in fact be seen as the result of such a reconstruction. During the act of listening, for example, we construct models of a condition (or state) that could have generated these sonic patterns. For example, in listening to a simple tonal piece, the chord progression appears as being driven by a compelling tonal force that explains why and how the chord sequence progresses the way it does. Clearly, this assumption of force is the result of a predictive model that emerges from, at least, (i) the available sonic patterns, and their physical-acoustical arrangement, (ii) the auditory mediator and its innate way of rearranging acoustical information, and (iii) previous encounters with music that provide knowledge about the repertoire. The tonal piece, thus, rests upon a predictive model that assumes a tonal goal-directed force as the condition behind the observed patterns. Embodied music cognition draws attention to the fact that the predictive model is largely the result of the constraints of the mediators and the listener's own corporeal states [34.10]. In short, the embodied paradigm assumes that music perception is the outcome of an interaction of the perceiver's states with states of the musical environment. The construction of a predictive model of music depends on corporeal mediators and on corporeal internal states, which determine how the sonic patterns will be perceived in terms of an interactive functionality.

34.1.2 Assumptions About Our Knowledge of the World

Embodied music cognition thereby offers a way of understanding the above ontology in relation to an epistemology of interaction processing. A basic idea is that the knowledge about the world, or music in our case, becomes simple when it is related to interaction. Complex music patterns that are linked with interactions (synchronized movements, for example) acquire roles and functions that are easy to understand because the complexity of the pattern, as such, is reduced to an interaction issue that is basically dealt with at the intentional level. The embodied viewpoint holds that this process is strongly influenced by knowledge and skills (i. e., the classical music cognition viewpoint), as well as by the nature of bodily mediators (such as the auditory system and biomechanics of bodily effectors) and states of being (arousal levels, and other bodily states such as fatigue, fitness, and so on) [34.10]. The mediators impose their proper constraints by the way in which music can interact. They determine the access to the true state of the music. In addition, mediators influence the construction of predictive models about the observed musical patterns. Besides that, they offer a way of simplifying the complexity of the patterns, by helping to reduce them to an action-oriented perspective on patterns. Mediators thereby constrain the observed musical pattern in view of the proper intentional, energetic and affective states of the person involved in the interaction, but they constrain these musical patterns in such a way that they can be used for interaction. Accordingly, extremely complex musical patterns (containing dense spectral information) can be linked with gestures, or body movements that form a unit of an action, such that they appear as objects of action targets. That simplifies the concept of the patterns. Gestures thereby support the construction of an intentional state of interaction [34.18, 19]. What we claim here is that once that level of interaction is achieved the complexity of the pattern as such disappears and the pattern becomes a component of an interaction that is driven by action goals. To sum up, interaction with music is based on a coupled system of person–environment states, or a coupled system of person–music states. This system produces patterns that can be exchanged, observed, and processed, either by the person, or perhaps even by others. This allows people to generate predictive models about the patterns and about the person–environment state interaction. The predictive models reduce the complexity of the involved patterns by linking them with actions and intentions that drive the interaction. Patterns take on interaction roles. Cognition is itself an aspect of the condition for having a person–environment interaction. It is an aspect of the state of being, juxtaposed with aspects such as physical effort and emotion.

Note that the above ontology and epistemology also applies to scientific work, in particular the research in music cognition. Empirical studies thereby aim at unraveling the causality of the interaction of the person–environment states, using observations about these states as points of departure to build prediction hypotheses. The way in which a person acquires knowledge about music runs parallel with the way in which a scientist acquires knowledge about that person's acquisition of knowledge about music. Methods

and approach are different but the ontological and epistemological basis is similar.

34.1.3 A New Paradigm for Music Research?

Although the interest in embodied music cognition could be traced back to historical work on music and gymnastics [34.20, 21], and empathic moving along with music as the foundation for aesthetic experience [34.22, 23], it is only up until quite recently that embodied music cognition was explicitly proposed as a paradigm for music research [34.10], and there are several reasons for that. One is that new technologies became available that have enabled the recording of body movements. Other reasons refer to the classical debates about scientific paradigms, discussions about the restricted focus of *cognition in perception*, about anomalies in empirical studies, and ways of viewing things [34.24]. A common example is tonality perception, where the classical cognitive paradigm is said to neglect the effects of sensory processing, and where evidence is gathered about incorrect conclusions drawn from empirical facts; mainly because the world is seen from the viewpoint of a different paradigm [34.25]. Embodied music cognition can be seen as an attempt to further develop the classical cognitive paradigm. As mentioned, the empirical approach for inquiry is fundamentally the same but the main extension concerns the role of corporeal mediators and states during ongoing musical interactions, including perception. Seen in that context, the embodied music cognition paradigm has expanded its research field from *cognition in perception* to *cognition in interaction*.

Embodied music cognition research is currently characterized by two trends; namely an in-depth exploration of embodiment and a broadening of the concept of embodiment. The first trend explores in greater detail the nature of embodiment. Several studies have, indeed, introduced challenging new ideas about bodily articulations and support for expressive gesturing in relation to music [34.26–29]. These ideas currently profit from more detailed viewpoints on the sensorimotor principles and direct proof of embodied influences on perception [34.30–32]. The second trend is to broaden the perspective of embodied music cognition and to link it to research areas such as musical affect research [34.33], wellbeing [34.34], healing [34.35], music engineering [34.36], and brain studies [34.37]. This trend suggests that the power of embodiment can be put to work in a broader context, embracing sister disciplines such as music engineering, music psychology and music and brain science. Embodiment is embedded here into an overall motivational approach.

34.2 The Architecture of Embodied Music Cognition

A cognitive theory has a focus on processes that handle the anticipation, goal-satisfaction, and emergent forms (traditionally called gestalt formation) of patterns, to be understood here as musical patterns and movement patterns – in short as materiality that moves and where its movement constitutes a pattern. Patterns can be learned, and based on this acquired knowledge one can predict and estimate what happens in an interaction. The role of cognition, therefore, is to construct predictive models (through learning) and apply these models in ongoing interactions with the environment. Admittedly, this is not a new idea, as it has been a major research idea of the music cognition paradigm for decades.

34.2.1 Prediction

Prediction focuses on patterns that are expected to occur. The occurrence of those patterns would then confirm the prediction, or otherwise [34.2, 4, 5]. According to the cognitive theory of music, anticipation would typically be built up by an exposure to previous musical patterns, which is called *the context*. The patterns that define this context have inherent constraints (due to acoustical structures) that are combined with inherent constraints of human auditory physiology (the innate disposition to process sounds) and with habituations to the musical repertoire (due to learning). The combination of these constraints sets up a particular context for the prediction and the subsequent processing.

The work on tonality has been paradigmatic for this approach. Tonal sequences have a strong anticipatory character. The classical theory thereby assumes that the outcome of a prediction (whether it matches with the perceived patterns or not) may be linked with affect processing, such as relaxation, tension, and ultimately, with emotion and arousal (e.g., [34.1, 38]), and a reward (or pleasure) system that reinforces the satisfaction mechanism [34.39, 40]. Moreover, prediction can be described in terms of a pattern-processing model that functions independently from the physical carrier (the brain). The typical processing components that frame this type of cognitive processing are working memory, long-term memory,

schema construction, and schema automation. These components define a *cognitive architecture* that, together with a reward system, allows us to understand interaction with music as meaningful and motivating [34.41].

The viewpoint of embodied music cognition adds to this approach, but it stresses the bodily involvement in predictive processing. In fact, it states that bodily involvement influences the anticipation of perceptual events, in addition to perceptual events themselves. The viewpoint implies that anticipation needs to be conceived in terms of schemes that combine perception-related issues with sensorimotor-related issues. For example, embodied music cognition would stress the fact that predictions of patterns rely on innate predictive mechanisms, such as auditory mechanisms that draw upon the inherent sonic arrangements of pitches, and the temporal integration of these arrangements in short-term memory [34.42–44]. As convincingly shown in [34.25], this already forms a powerful basis for predicting the underlying tonal sense of a musical piece. Of course, it may not predict the specific musical arrangements of a particular piece. After all, the innate mechanisms do not *know* the repertoire. Learning the repertoire is a typical capability of habitual cognitive processing. Other examples can be given in relation to rhythm prediction. There we see that the predictions can be based on particular biomechanical properties of the bodily effectors, such as the propensity for 2 Hz resonances [34.45].

The embodied music cognition approach assumes that predictive schemes tap into the kinaesthetic, tactile, and haptic sensing particularities of the body. Together with the biomechanic constraints of the corporeal effectors of movement (such as legs, arms), the state of arousal, fatigue, and energy characterizes a system that filters the back and forth interaction with the environment. Accordingly, it is assumed that this system has a large impact on the prediction machinery and the prediction outcomes. Anticipation is therefore understood as the expected outcome of bodily mediated perceptions and actions, rather than the expected outcome of some type of direct line from environment to brain. The implication of this insight is far reaching, as suggested in the previous section. For example, it implies that music can no longer be seen as consisting of note-symbol arrangements, because these arrangements are lacking the inherent constraints of real musical patterns. In addition, it implies that the perceived musical structure is itself an outcome of the internal state interacting with the musical patterns. Consider the anticipation of the beat in samba music based on movement states, for instance. This is known to be difficult for Western listeners because samba has an ambiguous binary-ternary structure. *Naveda* and *Leman* [34.46] however, suggest that

> *perception of samba may be movement-based in the sense that through self-movement (of the dancer in response to music) musical patterns become rhythmically disambiguated.*

Through movement states, the perception of binary structure can be emphasized, which facilitates the prediction (see also [34.7]). Or consider the interaction with music through dancing. Based on the induced emotional state, our perception of music changes, and this perception is reflected in our dance expression [34.8, 47]. These studies show that a perception-only approach is no longer defensible. Music interaction is based on action–perception coupling, integrated within the predictive machinery, and above all, tightly integrated with the corporeal mediator constraints.

34.2.2 Emergent Pattern Building

A core aspect of the embodiment theory concerns the reconception of what musical structures are. Music cognition research thereby draws upon emergence as a basic principle of musical structure. Emergence implies processes that use structures and relationships among patterns as ingredients for new types of patterns. Emergence is a formative principle based on patterns at a lower level generating other patterns at a higher level, and this principle is fundamental to the understanding of music. For example, the Western music repertoire with its typical tonal-harmonic system of anticipation and goal-satisfaction can be seen as the result of an emergent process that is based on auditory dispositions (the way the auditory physiology works), environmental dispositions (the type of instruments used, and the fact that these instruments have a harmonic overtone structure), and previous repertoires (acquired through learning). Obviously, there are many additional factors that may mediate the formation of the repertoire, such as beliefs (e.g., the belief that music is a divine art that should display stability, or the other view, that it should display the vibrations and variability of a divine being) that foster the selection and molding of instruments and structures until they lead to a repertoire that, in turn, influences the way in which people habituate towards particular phrasings.

Embodiment draws upon the idea that this emergence is largely influenced by interactions. The focus here is on the added value of interaction in processes that leads to emergent patterns. As these patterns are often hierarchically structured, with lower-level patterns being ingredients of higher-level patterns, perception of these emergent patterns can be guided by move-

ments or actions. The role of actions is related to the attenuation (or filtering) of particular characteristics of emergent patterns, the use of action as a facilitator of particular interaction patterns (such as timing), the disambiguation of complex patterns through movement, and possibilities for motor retraining [34.30]. Emergence is a crucial aspect of the reductive process discussed above, through which complex patterns can be simplified and related to interaction roles.

34.2.3 Enactment

Enactment can be defined as the instance of acting something out. This concept embraces the idea that our corporeal involvement with music is based on acting along with the music, such as during synchronizing footsteps with music, or during the alignment of arm and hand movements along with the musical expressive flow, like in dancing, conducting, and music playing. Enactment can be seen as a musical intentionality induction mechanism. That means that it is a mechanism that turns musical patterns, as they appear to our senses, into goal-directed sonic moving forms, with value and reward. Patterns thus become endowed, and are simplified, by considering them as properties of the states built on an assumption that our predictive machinery makes about these patterns. And this assumption is largely based on the enactment, that is, on processes that ensure the synchronization and alignment of actions with those patterns. Note that enactment involves the engagement of the human action repertoire (to carry out the enactment). This action repertoire is associated with representations about the expected outcomes of these actions. Thus, when actions are carried out, they automatically become associated with representations of action outcomes (also called intentions).

Enactment may be seen as a *method* for realizing an intentional outcome. Basically this means that corporeal interactions determine prediction models for the ongoing and future interactions with music, so that music can be conceived in terms of goals, directions, targets, values, and reward. Embodiment thus provides a basis for updating the predictions of what is anticipated during the interaction with music, and this is a core element of the intentionality induction process; the way in which intentionality with music gets established; the way in which music is conceived in terms of goals, values, and reward.

34.2.4 Expression

Finally, some mention should be made of expression (the act of expressing), or expressiveness (the property of being expressive). Expression is a core feature of musical communication, and one of the key research domains of the embodied music cognition approach in today's research environment. Expression research is a challenge because it is not understood well in cognitive science. However, it is worth mentioning that a good part of the older work in continental musicology was especially focused on musical expression [34.48–52]. Interestingly, in this tradition, musical expression and expressiveness are often approached from the viewpoint of gesture. For example, *Schaeffer* [34.50] discusses sound objects and their association with the actions that might have produced these sounds. Much of the former work was embedded within a phenomenological tradition (e.g., [34.22, 53]), in which concepts such as embodiment, or motility (the capability of moving) were explored from a philosophical and music-analytical point of view. For example, according to *Broeckx* [34.52], music is expressive because music appears to humans as a material organism possessing properties and behaviors. The attribution of expressive properties to music is based on prereflective sensorimotor associations of sound properties with tactile, visual, and muscular and vascular sensations (e.g., flatness, brightness, rolling, tingling) leading to associations with feelings and moods. Moreover, it is inferred that the movement of sound properties engages a kinaesthetic process that results in associations of body movement and gestures.

This musicological tradition can, and rather interestingly, be linked with the current embodied viewpoint on music cognition. The latter thereby provides a refinement of the ontological, epistemological, and methodological framework for understanding musical expression in performance and perception. The main difference is that the current focus of study has an empirical orientation, whereas the earlier focus of study took a music-analytical philosophical orientation. The study of expression, when seen from the viewpoint of embodied music cognition, can be linked with action theories and brain theories that assume (on the basis of empirical evidence) the existence of predictive processing states related to action–perception couplings. Expressive gestures are thereby assumed to tap into innate motor expressive responses that have been integrated with predictive processing states that control expressive habits and learned expressive behaviors. The call for action science [34.15], for example, has a huge advantage because specific hypotheses about expression and aesthetics can be tested within a much larger relevant empirical framework.

34.3 Empirical Evidence for Embodied Music Cognition

We now turn to some of the empirical work that relates to the embodied music cognition paradigm. Our goal is to focus on studies that reveal evidence for embodied music cognition. However, rather than aiming at an exhaustive overview, we confine ourselves here to three core aspects of the research; namely understanding expressive gestures, understanding synchronization and entrainment, and understanding the effects of action on perception [34.54, 55].

34.3.1 Understanding Expressive Gestures

Although gesture research has a long tradition in musicology (see above) and in research on interactive music systems [34.56], it is only recently that attempts have been made to better understand gesture in relation to music [34.11, 18, 19]. The role of gesture can be understood in terms of encoding and decoding of musical expression. Encoding of expressive qualities in music happens during music performance. The gestures that support the encoding are called *expression-supporting* gestures. Decoding of expressive qualities in music happens during listening. The gestures that support the decoding are called *expression-responding* gestures. In both cases it is assumed that expression can somehow be transferred from gesture to music, of from music to gesture. This transfer implies a *mirroring* from qualities of sound patterns to qualities of movement patterns. However, rather than speaking about mirrors, it is more convenient to speak about action opportunities. Gesturing can be seen as an encoding and decoding of expressive affordances, or opportunities for action, to be put into the sound, or to be taken from the sound [34.57, 58]. This viewpoint implies that gestures are supported by actions, and that mirroring is an aspect of the gestural alignment with music. Mirroring indeed suggests a one-to-one relationship between music and gesture, but it may involve counterpoint, and an emphasis of the contrast between music and gesture. Affordances also imply that action opportunities have a goal-directed or intentional character. This means that the associated movements, or gestures, are embedded in a predictive machinery. So far, several studies have been conducted that aimed to find a better understanding of the role of gesturing in musical activities. Most studies are explorative and seek to get to grips with the complexity of the matter. This is indeed a new field in which our understanding of how gestures work in relation to expression, intentionality, and timing is still in its infancy. In what follows, we give a brief overview of some explorations in this domain.

Elementary Gestures, Articulation and Coarticulation

First consider the concept of elementary gesture, which we developed in the context of a study that addressed the playing of the guqin (the old Chinese zither) [34.59–61]. Elementary gestures can be understood as small action units, or articulations, such as moving one finger from one position on a string to another. In the context of wind instruments, an elementary gesture would be the production of one tone. These elementary gestures are needed for sound production but they also contain expressive features. For example, the movement of the finger on the string can be realized through different shades of velocity: speeding up, or slowing down, or having a straight velocity [34.59]. In wind instruments, articulation is defined by a particular tonguing technique, often denoted as *tu*, *du*, *fu*, or *ku* depending on the strength of the attack and the position of the tongue. For example, in the *ku* attack, the tongue moves backwards, rather than forwards, as in *tu*. The interesting issue here is that elementary gestures point to very basic actions that seem to be controlled by a motor impulse (see also [34.62]). Once the impulse is given, the further deployment of the action is a sort of automatism. More complex playing gestures can be understood as concatenations of an alphabet of such short actions, or movement primitives.

Note that the next stage, in which two tones are connected in coarticulation, is immediately more complex [34.63]. It involves the connection of movement primitives, often also in combination with coordinated actions, such as finger movement in the left hand, combined with string plucking in the right hand in the guqin, or a technique in combination with fingering technique in wind instruments. Articulation and coarticulation provide the expressive ingredients of larger expression arcs that define the tension of sound production during a longer term (a few seconds).

Expression Arcs and Musical Intentions

Following on, we consider expression arcs. Research shows that expression arcs are typically supported bodily gestures that accompany intentional states when realizing expressive outcomes. To facilitate that aim, performers seem to rely on a repertoire of expressive-supporting gestures (see also, e.g., [34.64, 65]). For example, in [34.66], it was possible to observe a consistent relationship between a clarinettist's musical expressive-structural analysis (indicating target notes, supporting notes, crescendo, decrescendo) and the clarinettist's gestures that support the playing of these musical pas-

sages according to the planned expressions. The study shows that intentional actions are meant to structure music according to the expressive acting. It means that expressions in music are related to particular targets that are planned ahead and it demonstrates that the musical realization is supported by bodily gestures. Expressive arcs may require accurate timing. *Maes* et al. [34.32] suggest that the accurate timing needed to carry out expressive arcs is based on embodiment. In this study, limb movements were either legato or staccato movements of the arm and hand that controlled the bow of a cello. Legato implies continuous limb movements with no pause in between the onsets of tones, while staccato implies event-based movements with a pause (a nonmovement state) in between the onsets of tones. The results show that legato playing can compensate for disrupting effects of cognitive load on timing production and expression, while staccato playing is affected by these effects, especially at slower tempi. This is probably related to the timer mechanisms involved. The continuous movements allow the outsourcing of timing to a mediator (called embodied timing), while staccato implies nonmovement duration that requires cognitive timing. Given a supplementary cognitive task in addition to playing, the timing of staccato is more affected because it taps into the same cognitive resources as the supplementary cognitive task. In contrast, the timing of legato is less affected because the timing is outsourced to a corporeal mediator that, due to its proper movement constraints and the time needed to realize the movements, provides a clock mechanism that is highly functional. The study therefore supports the idea that a mediator mechanism can provide a timer for expressive performance. Overall, this study is one of the first to show why expressive-supporting gestures are relevant in music performance. The answer the study provides is that these gestures allow the cognitive outsourcing of timing. Attention can then be devoted to other crucial aspects of the expression.

Expressive Gestural Shapes in Relation to Musical Timing

Thirdly, consider gestural shapes in relation to timing. This is of particular relevance in the context of dancing. The type of gestures considered here are generally repetitive and constrained within an overall timing framework imposed by the musical meter. The repetitive character of the gestures suggests that the movement choreography, or movement sequencing, is based on a shape (called the basic shape, or basic gesture), that reoccurs over defined time periods. The overall timing framework imposed by the music suggests that elements of the basic shape are linked with timing. In that context, we investigated two different methods of studying this link [34.28, 29, 46]. One method has a focus on the basic shape and aims at extracting this shape from the movement trajectories. A basic gesture could be defined as the average of several movement trajectories over one single period of the movement. However, due to inherent uncertainties in the neuromotor control, there may be quite some variability in the way in which these movements are timed and shaped. Due to this variability it is not possible to define the basic gesture as the average of repetitive movement trajectories. That would lead to nonrealistic irregular shapes. The best way to obtain this is by using periodicity transformations [34.46], or by building probabilistic models that allow a smooth reconstruction of the movement trajectories [34.67]. Such a model can then be linked with the timing imposed by an external musical source.

The other method has a focus on timing and aims at understanding the connection between shaping and timing by projecting features of the musical timing onto those shapes. This allows for an analysis in terms of point clouds, and spatiotemporal regions showing how the shaping is connected with the timing. Interestingly, it is also possible to plot musical features onto these basic gestures. For example, it is possible to extract the time of the beat and plot a mark of the beat on the gesture at the time that corresponds with the beat. One method is to plot these features onto the basic gesture, while another involves plotting these features onto the entire periodic trajectory. The difference is that the basic gesture provides an average of that trajectory. When musical features are plotted on the entire trajectory, one typically obtains point clouds with musical features. Accordingly, this can be done for different metrical periods (half beat, double beat, and so on). As such, one obtains a spatiotemporal plot of the musical meter onto the gesture. Such graphs show interesting differences between, for example, the Charleston and samba dances, and especially between novices and experts. When applied to different dancers, we see an interesting variability in how humans realize bodily expression in relation to music. Moreover, when dancing, it is likely that different body parts capture different periods related to the musical beat. For example, horizontal sway could occur on two beats in period, while the arms might register one beat in the period, and hands at half the beat period. As such, the human body can reflect the different frequencies present in the musical structure [34.27, 68]. The positions of the beat points may reflect *reference points*, or landmarks, that play a prominent role in the predictive control of the dance. The study on samba dance, for example, provided evidence that dancing may facilitate and disambiguate the perception of music. This is assumed to be based

on predictive mechanisms that involve the suppression of filtering of particular features that can be predicted. Moreover, the activity, as such, accords with the idea that decoding musical expressiveness is a sense-giving activity.

34.3.2 Understanding Synchronization and Entrainment

Research on sensorimotor synchronization is firmly rooted in traditions of studies that focus on the timing of finger tapping in relation to metronome ticks [34.69]. The research has recently been expanded to dancing and music, where similar issues of timing exist [34.70, 71]. Sensorimotor synchronization is deemed to be highly relevant for embodied music cognition. One of the basic ideas here is that the synchronization of movement and music is driven by predictive processes that drive error-correction mechanisms. Errors are perceived deviances between expected action outcomes and the perceived external musical stimulus. There appears to be an internal mechanism that aims at correcting these deviances so that action outcomes optimally match with the musical stimulus. For example, a particular posture in dancing should be reached at a particular timing cue. This synchronization is governed by a dynamic (called entrainment) that aims at minimizing the error between what is anticipated and what is realized.

Note that music–movement entrainment is quite a complicated timing phenomenon, involving constraints of the corporeal (and even technological) mediators in the entrainment setup [34.31, 72]. One of these constraints is related to resonance. Starting with a tapping study, *Van Noorden* and *Moelants* [34.73] found a resonance frequency (or embodied eigenfrequency) around 2 Hz. Later on, this *eigen-frequency* phenomenon was observed in several studies on walking, such as in [34.45, 74]. An interesting result has been obtained in a study that provides direct evidence for a vigor-entrainment effect [34.74]. In this study, subjects were instructed to walk in synchrony with musical excerpts that had the same tempo but different musical expression. The speed of walking is determined by two parameters; namely, the pace (step frequency) and the stride length, and the pace is fixed due to the synchronization with the music. Therefore, the only parameter that could influence the speed is the stride length, which is an effect of the muscle strength. The study showed that some music has an activating effect, stimulating subjects to walk faster, while other music has a relaxing effect, stimulating subjects to walk slower, while metronome ticks (and many other types of music) have no effect. Interestingly, the adaptation of muscle strength to music depends on the expressive characteristics of the music. These can be defined in terms of acoustical features, showing that binary and ternary arrangements in music play an important role.

Social entrainment, or the adaptation of human movement to other humans while listening and producing music, has been studied in several contexts. *Van Dyck* et al. [34.70] asked subjects to dance to music. They found a group effect for activity count and tempo entrainment when people danced in groups of five. [34.75] is based on an experiment with more than 400 children who, in groups of four, had to continue the example of an avatar that tapped along with the beat of the music. Several tempi were tested, showing that children have a 2 Hz resonance in their propensity for tapping. The results suggest that the perception and production of the pulse is related to the control of the human body, cognitive development (age), and the ability to control timing either by counting or by using outsourced embodied timer mediators. Moreover, the analysis shows a social entrainment effect in the sense that the social group, in addition to the musical stimulus, also entrains children and provokes their excitement so that they start to invent variations on their tapping rhythm. Synchronization and entrainment is a field of study that provides strong evidence in favor of the embodied music cognition paradigm.

34.3.3 Understanding Effects of Actions on Music Perception

Finally, it is of interest to consider embodiment from the viewpoint of the coupling of action and perception. Research thereby focuses on how action influences perception, as well as how perception, thanks to its close association with action, can influence action. The latter may be of relevance, for example, in motor rehabilitation where music is used to retrain maladapted motor habits. Studies that explicitly address action–perception couplings are highly relevant for the embodiment theory because they somehow provide direct proof of the role of embodiment in music perception. A crucial concept in our understanding of action–perception couplings is that of internal sensorimotor predictive models; in short, internal models. There is accumulating evidence that the sensory (e.g., auditory, visual, etc.) predictions made by these models influence the online processing of sensory information, such as music, emerging from the external environment [34.8, 76–78]. The particular influence of these predictions on the perception of music differs depending on the nature of the auditory information (when merely listening to music), or audio-visual information (when watching and listening to a music performance, for instance). *Maes* et al. [34.30] provide an overview of the recent studies

that consider the effects of actions on music perception. One of the effects of internal models is that perception becomes attenuated when the incoming sensory information matches the sensory prediction, while perception is facilitated when sensory input is incoherent with the sensory prediction. Also, when sensory input is ambiguous in nature, body movements and the related sensory predictions enable it to disambiguate incoming sensory information. Additionally, it has been shown that when moving to music, structural and affective properties inherent to the body's movements become attributed to the music. *Maes* et al. [34.30] refer to the concepts of *selective attention*, and *cue selection/identification* to explain how body movements performed to music may direct attention towards specific cues in the music. It is argued that the mechanisms described by these concepts can steer people's perception of structural features in the music (e.g., melody, beat, pitch, etc.) in a specific direction (as shown in [34.6, 7]). *Maes* and *Leman* [34.8] showed that expressive body movements can condition children's perception of musical expressiveness. In the experiment, children were conditioned with a happy or a sad choreography in response to music that had an expressively ambiguous character. Afterwards, the children's perception of musical expressiveness in terms of valence and arousal was assessed. The results suggest that the expressive qualities of the movements they learned to associate with the music had a significant impact on how children perceived musical expressiveness.

Of particular interest are studies that address the effect of emotions on music-driven movements. In [34.79], the effect of two basic emotions, happiness and sadness, on dance movement was examined. Participants were induced to feel emotional states of either happiness or sadness and then danced intuitively to an emotionally neutral piece of music, composed specifically for the experiment. The results revealed that when induced with happiness, participants moved faster, with more acceleration, and made more expanded and more impulsive movements than in the condition with sadness induced. In [34.47], it was investigated whether the induced emotion could be successfully decoded from movements (without music as extra stimulus). Observers watched a set of silent videos showing depersonalized avatars of dancers moving to an emotionally neutral musical stimulus after emotions of either sadness or happiness had been induced. Results revealed that observers were able to identify the emotional state of the dancers with a high degree of accuracy.

34.4 Embodiment and Dynamic Cognition

Based on the above selective overview, we can distinguish two important trends that mark current embodied music cognition research. The first trend is to consider musical activity as the outcome of a network of mutually interdependent sensory, motor, and affective processes [34.54]. These processes define the overall state of a person's activity involved with music. Predictive processing plays an important role in the control of these processes. In that context, perception and action are no longer seen as separate processes, but as mutually reinforcing processes that subsume predictive processing in connection with bodily states. However, the road to fully understanding the underlying mechanisms, in particular concerning the origin (innate/learned) and development (e.g., association learning, contingency, contiguity, etc.) of these associations is long. A major challenge concerns the understanding of expression in relation to this network of mutually interdependent processes. Expression can be conceived as a facilitator of the interaction among these processes but more work is needed to fully comprehend how this works. The second trend is that music cognition is being defined in terms of interactive dynamics, in which concepts such as emergence, affordance, and dispositions define states (e.g., homeostasis) and processes (e.g., error-correction mechanisms) of the dynamics. Emergence and affordance, for example, appear as interaction outcomes and interaction opportunities. They offer a new way of viewing music as the outcome of a person–environment interaction, in which particular constraints in the environment and in the mediators are controlled and adjusted in functions of particular outcomes. Accordingly, music perception should be understood as a phenomenon that emerges from the dynamic interaction of many intertwined processes evolving over time. Embodied music cognition theory draws upon a dynamical theory of music cognition in which the interaction between the various sensorimotor, affective, and cognitive systems (attention, memory, metaknowledge, etc.) and the external environment are of crucial importance.

34.5 Contributions to a Paradigm Shift in Systematic Musicology

We believe that embodied music cognition research has become an important paradigm for music research. The role of bodily involvement is no longer questioned and the empirical methodology has been expanded to quite a broad domain of musical activities. The strength of the paradigm is that it builds on a solid empirical and modeling framework. The weakness of the paradigm is that it may need further elaboration to meet historical and cultural contextualization. In recent years, the paradigm of embodied music cognition has been embraced and filtered into the following directions. Firstly, in the methodology, studies have contributed to an expansion of the existing methods. New observation devices (e.g., all kinds of sensors and camera systems) and analysis methods (e.g., multivariate, and functional data analyses) have been refined and introduced. These methods have been combined with new approaches in audio analysis, which draw upon massive feature extraction and subsequent machine learning methods [34.80]. In addition, subjective methodologies (e.g., questionnaires that probe flow or presence) have been fine-tuned and linked with performer-inspired analysis methods [34.66] and with educational technologies [34.81]. In general, the methods that are inter- and cross-disciplinary have greatly improved compared to the state-of-the-art music research of a decade ago. New user-oriented approaches have been developed in the direction of person-centered approaches in a cross-disciplinary context of collaboration with other research groups and institutions (e.g., health care institutions) [34.82]. Secondly, the ontological assumptions and the associated epistemology have been refined by expanding the traditional viewpoints (e.g., [34.83, 84]) in the direction of an enactive approach. Last but not least, the paradigm received much attention thanks to the idea that practical tools (based on interaction) contribute much to the development of the new paradigm [34.81, 85–91]. This approach may be rather new in the field of musicology, although it draws upon a certain tradition that was associated with music cognition research [34.56]. Tools offer a means by which intervention can be made into the action–perception cycle. By contributing to the theoretical development therefore, they coerce the researcher into focusing on the principles that lie behind those tools. Success in developing concrete tools can serve as a future measuring instrument used to explore the power of the paradigm of embodied music cognition. In that sense, the issue of embodied music cognition versus nonembodied music cognition is not only a matter of epistemology (and trying to discover the realities about how music interaction works) but is also a matter of practical application. Recent work clearly indicates that the embodied approach to music cognition may lead to useful knowledge about music interaction, in fields such as music education, music recreation, and even physical rehabilitation.

The paradigm of embodied music cognition has stimulated developments in cultural contextualization studies. Several recent studies (e.g., [34.92–96]) draw upon concepts and results from embodied music cognition, such as gesture analysis, analysis of entrainment in concrete musical performances, the role of music in relation to empathy, and several issues in action–perception couplings. On the one hand, these studies point to issues that may require further in-depth laboratory research. On the other hand, finding a good trade-off between laboratory findings and a profound cultural study is still a challenge. In any case, this mutual influence of cultural musicology and systematic musicology is very promising for both disciplines. More profound cultural contextualization, as well as historical contextualization, certainly injects added value into the paradigm of embodied music cognition.

34.6 Conclusion

In this chapter we provided an overall viewpoint on embodied music cognition, addressing questions related to the nature, assumptions, findings, and perspectives of the paradigm. We thereby addressed the ontological and epistemological assumptions and provided a specification of the cognitive architecture in relation to embodiment. The result is a perspective that greatly capitalizes on the framework of the (classical) music cognition research approach. However, as argued, the embodied approach is much more oriented towards pragmatism, dynamics, and interaction. The focus on *cognition in perception* has been replaced by a focus on *cognition in interaction*, of which perception research is a part. Based on an overview of recent studies, many of which are studies sourced from our own laboratory, we aimed at providing research topics and research outcomes that illustrate the development of the approach. Along that way, we identified the trend towards an in-depth exploration of the concept of embodiment in terms of sensorimotor processing and concrete proof

of the necessity for the embodied viewpoint, and the trend towards a broadening of the concept of embodiment in the direction of sister disciplines. These trends fit quite well with the accumulating evidence in favor of the embodied approach. This partners well with the exploration of novel methods for observation and analysis, adopting a pragmatic approach to music interaction, and exposing a high level of interest in the development of tools that make the theoretical views applicable in domains such as music education, sports and recreation, and rehabilitation and wellbeing. Overall, this overview provides a strong incentive for further research, especially in the direction of expressive interactions with music [34.97, 98].

References

34.1 R. Jackendoff, F. Lerdahl: The capacity for music: What is it, and what's special about it?, Cognition **100**(1), 33–72 (2006)

34.2 L. Meyer: *Emotion and Meaning in Music* (Univ. of Chicago Press, Chicago 1956)

34.3 M. Jones, M. Boltz: Dynamic attending and responses to time, Psychol. Rev. **96**(3), 459–491 (1989)

34.4 D. Huron: *Sweet Anticipation: Music and the Psychology of Expectation* (MIT Press, Cambridge 2006)

34.5 H. Honing: *Musical Cognition: A Science of Listening* (Transaction Publishers, Piscataway 2011)

34.6 J. Phillips-Silver, L. Trainor: Feeling the beat: Movement influences infant rhythm perception, Science **308**(5727), 1430 (2005)

34.7 J. Phillips-Silver, L. Trainor: Hearing what the body feels: Auditory encoding and rhythmic movement, Cognition **105**(3), 533–546 (2007)

34.8 P.-J. Maes, M. Leman: The influence of body movements on children's perception of music with an ambiguous expressive character, PLoS One **8**(1), e54682 (2013)

34.9 T. Fritz, S. Hardikar, M. Demoucron, M. Niessen, M. Demey, O. Giot, Y. Li, J.-D. Haynes, A. Villringer, M. Leman: Musical agency reduces perceived exertion during strenuous physical performence, Proc. Natl. Acad. Sci. **110**(44), 17784–17789 (2013)

34.10 M. Leman: *Embodied Music Cognition and Mediation Technology* (MIT Press, Cambridge 2007)

34.11 R. Godøy, M. Leman (Eds.): *Musical Gestures: Sound, Movement, and Meaning* (Routledge, New York 2010)

34.12 T. Metzinger: *Being No One: The Self-Model Theory of Subjectivity* (MIT Press, Cambridge 2003)

34.13 J. Stewart, O. Gapenne, E.A. Di Paolo (Eds.): *Enaction: Toward a New Paradigm for Cognitive Science* (Cambridge Univ. Press, Cambridge 2010)

34.14 A. Clark: Whatever next? Predictive brains, situated agents, and the future of cognitive science, Behav. Brain Sci. **36**(3), 181–204 (2013)

34.15 W. Prinz, M. Beisert, A. Herwig (Eds.): *Action Science: Foundations of an Emerging Discipline* (MIT Press, Cambridge 2013)

34.16 A. Engel, A. Maye, M. Kurthen, P. König: Where's the action? The pragmatic turn in cognitive science, Trends Cogn. Sci. **17**(5), 202–209 (2013)

34.17 F. Picard, K. Friston: Predictions, perception, and a sense of self, Neurology **83**(12), 1112–1118 (2014)

34.18 A. Gritten, E. King (Eds.): *New Perspectives on Music and Gesture* (Ashgate, London 2011)

34.19 A. Gritten, E. King (Eds.): *Music and Gesture* (Ashgate, London 2006)

34.20 A. Truslit: *Gestaltung und Bewegung in der Musik* (CF Vieweg, Berlin–Lichterfelde 1938)

34.21 E. Jaques-Dalcroze: *Rhythm, Music and Education* (Putnam's Sons, New York 1921)

34.22 T. Lipps: *Grundlegung der Ästhetik* (Leopold Voss, Leipzig 1903)

34.23 G. Becking, N. Nettheim: *How Musical Rhythm Reveals Human Attitudes* (Peter Lang, Frankfurt am Main 2011)

34.24 T. Kuhn: *The Structure of Scientific Revolutions* (Univ. of Chicago Press, Chicago 2012)

34.25 E. Bigand, C. Delbé, B. Poulin-Charronnat, M. Leman, B. Tillmann: Empirical evidence for musical syntax processing? Computer simulations reveal the contribution of auditory short-term memory, Front. Syst. Neurosci. **8**(94), 1–27 (2014)

34.26 B. Sievers, L. Polansky, M. Casey, T. Wheatley: Music and movement share a dynamic structure that supports universal expressions of emotion, Proc. Natl. Acad. Sci. **110**(1), 70–75 (2013)

34.27 P. Toiviainen, G. Luck, M.R. Thompson: Embodied meter: Hierarchical eigenmodes in music-induced movement, Music Percept. **28**(1), 59–70 (2010)

34.28 M. Leman, L. Naveda: Basic gestures as spatiotemporal reference frames for repetitive dance/music patterns in samba and charleston, Music Percept. **28**(1), 71–91 (2010)

34.29 L. Naveda, M. Leman: The spatiotemporal representation of dance and music gestures using topological gesture analysis (TGA), Music Percept. **28**(1), 93–111 (2010)

34.30 P.-J. Maes, M. Leman, C. Palmer, M. Wanderley: Action-based effects on music perception, Front. Psychol. **4**(1008), 1–14 (2014)

34.31 B. Moens, M. Leman: Alignment strategies for the entrainment of music and movement rhythms, Ann. New York Acad. Sci. **1337**(1), 86–93 (2015)

34.32 P.-J. Maes, M. Wanderley, C. Palmer: The role of working memory in the temporal control of discrete and continuous movements, Exp. Brain Res. **233**(1), 263–273 (2015)

34.33 T. Cochrane, B. Fantini, K. Scherer (Eds.): *The Emotional Power of Music: Multidisciplinary Perspectives on Musical Arousal, Expression, and Social Control* (Oxford Univ. Press, Oxford 2013)

34.34 R. MacDonald, G. Kreutz, L. Mitchell: *Music, Health, and Wellbeing* (Oxford Univ. Press, Oxford 2012)

34.35 B. Koen, J. Lloyd, G. Barz, K. Brummel-Smith (Eds.): *The Oxford Handbook of Medical Ethnomusicology* (Oxford Univ. Press, Oxford 2011)

34.36 A. Kirke, E. Miranda (Eds.): *Guide to Computing for Expressive Music Performance* (Springer, Heidelberg, Berlin 2013)

34.37 M. Arbib (Ed.): *Language, Music, and the Brain: A Mysterious Relationship* (MIT Press, Cambridge 2013)

34.38 S. Koelsch: Brain correlates of music-evoked emotions, Nat. Rev. Neurosci. **15**(3), 170–180 (2014)

34.39 R. Zatorre, V. Salimpoor: From perception to pleasure: Music and its neural substrates, Proc. Natl. Acad. Sci. **110**(Suppl. 2), 10430–10437 (2013)

34.40 V. Salimpoor, I. van den Bosch, N. Kovacevic, A. McIntosh, A. Dagher, R. Zatorre: Interactions between the nucleus accumbens and auditory cortices predict music reward value, Science **340**(6129), 216–219 (2013)

34.41 M. Leman, L. Nijs: Music cognition and technology – Enhanced learning for music playing. In: *The Routledge Companion to Music, Technology and Education*, ed. by A. King, A. Ruthmann, E. Himonides (Routledge, New York 2015)

34.42 M. Leman: An auditory model of the role of short-term memory in probe-tone ratings, Music Percept. **17**(4), 481–509 (2000)

34.43 E. Large, F. Almonte: Neurodynamics, tonality, and the auditory brainstem response, Ann. New York Acad. Sci. **1252**(1), E1–E7 (2012)

34.44 G. Bidelman, J. Grall: Functional organization for musical consonance and tonal pitch hierarchy in human auditory cortex, NeuroImage **101**, 204–214 (2014)

34.45 F. Styns, L. Van Noorden, D. Moelants, M. Leman: Walking on music, Hum. Mov. Sci. **26**(5), 769–785 (2007)

34.46 L. Naveda, M. Leman: A cross-modal heuristic for periodic pattern analysis of samba music and dance, J. New Music Res. **38**(3), 255–283 (2009)

34.47 E. Van Dyck, P. Vansteenkiste, M. Lenoir, M. Lesaffre, M. Leman: Recognizing induced emotions of happiness and sadness from dance movement, PLoS One **9**(2), e89773 (2014)

34.48 E. Kurth: *Musikpsychologie* (Krompholz, Bern 1947)

34.49 A. Wellek: *Musikpsychologie und Musikästhetik: Grundriss der systematischen Musikwissenschaft* (Bouvier, Bonn 1982)

34.50 P. Schaeffer: *Traité des Objets Musicaux: Essai Interdisciplines* (Seuil, Paris 1977)

34.51 W. Coker: *Music & Meaning: A Theoretical Introduction to Musical Aesthetics* (Free Press, New York 1972)

34.52 J. Broeckx: *Muziek, Ratio en Affect: Over de Wisselwerking van Rationeel Denken en Affectief Beleven bij Voortbrengst en Ontvangst van Muziek* (Metropolis, Antwerp 1981)

34.53 M. Merleau-Ponty: *Phénoménologie de la Perception* (Gallimard, Paris 1969)

34.54 M. Leman, P.-J. Maes: Music perception and embodied music cognition. In: *The Routledge Handbook of Embodied Cognition*, ed. by L. Shapiro (Routledge, New York 2014) pp. 81–89

34.55 M. Leman, P.-J. Maes: The role of embodiment in the perception of music, Empir. Musicol. Rev. **9**(3/4), 236–246 (2014)

34.56 A. Camurri, G. Volpe, G. de Poli, M. Leman: Communicating expressiveness and affect in multimodal interactive systems, IEEE Multimed. **12**(1), 43–53 (2005)

34.57 R. Godøy: Gestural affordances of musical sound. In: *Musical Gestures: Sound, Movement, and Meaning*, ed. by R. Godøy, M. Leman (Routledge, New York 2010) pp. 103–125

34.58 J. Krueger: Affordances and the musically extended mind, Front. Psychol. **4**(1003), 1–13 (2013)

34.59 L. Henbing, M. Leman: A gesture-based typology of sliding-tones in guqin music, J. New Music Res. **36**(2), 61–82 (2007)

34.60 H. Penttinen, J. Pakarinen, V. Valimaki, M. Laurson, H. Li, M. Leman: Model-based sound synthesis of the guqin, J. Acoust. Soc. Am. **120**(6), 4052–4063 (2006)

34.61 H. Penttinen, J. Pakarinen, V. Vlimki, M. Laurson, M. Kuuskankare, H. Li, M. Leman: Aspects on physical modeling of a chinese string instrument – The guqin. In: *Proc. 9th Int. Congr. Acoust* (2007) pp. 2–7

34.62 R. Godøy: Quantal elements in musical experience. In: *Sound–Perception–Performance, Current Research in Systematic Musicology*, ed. by R. Bader (Springer, Berlin, Heidelberg 2013) pp. 113–128

34.63 R. Godøy: Understanding coarticulation in music. In: *Sound, Music, and Motion*, ed. by M. Aramaki, O. Derrien, R. Kronland-Martinet, S. Ystad (Springer, Berlin, Heidelberg 2013)

34.64 C. Palmer: Music performance, Annu. Rev. Psychol. **48**(1), 115–138 (1997)

34.65 B. Vines, C. Krumhansl, M. Wanderley, D. Levitin: Cross-modal interactions in the perception of musical performance, Cognition **101**(1), 80–113 (2006)

34.66 F. Desmet, L. Nijs, M. Demey, M. Lesaffre, J.-P. Martens, M. Leman: Assessing a clarinet player's performer gestures in relation to locally intended musical targets, J. New Music Res. **41**(1), 31–48 (2012)

34.67 D. Amelynck: *The Analysis of Bodily Gestures in Response to Music: Methods for Embodied Music Cognition Based on Machine Learning*, Ph. D. Thesis (Ghent University, Ghent 2014)

34.68 B. Burger, M. Thompson, G. Luck, S. Saarikallio, P. Toiviainen: Hunting for the beat in the body: On period and phase locking in music-induced movement, Front. Hum. Neurosci. **8**(903), 1–16 (2014)

34.69 B. Repp, Y.-H. Su: Sensorimotor synchronization: A review of recent research (2006–2012), Psychon. Bull. Rev. **20**(3), 403–452 (2013)

34.70 E. Van Dyck, D. Moelants, M. Demey, A. Deweppe, P. Coussement, M. Leman: The impact of the bass drum on human dance movement, Music Percept. **30**(4), 349–359 (2013)

34.71 B. Burger, M.R. Thompson, G. Luck, S. Saarikallio, P. Toiviainen: Influences of rhythm- and timbre-related musical features on characteristics of music-induced movement, Front. Psychol. **4**(183), 1–10

(2013)
34.72 B. Moens, C. Muller, L. van Noorden, M. Franěk, B. Celie, J. Boone, J. Bourgois, M. Leman: Encouraging spontaneous synchronisation with D-Jogger, an adaptive music player that aligns movement and music, PLoS One **9**(12), 40 (2014)
34.73 L. Van Noorden, D. Moelants: Resonance in the perception of musical pulse, J. New Music Res. **28**(1), 43–66 (1999)
34.74 M. Leman, D. Moelants, M. Varewyck, F. Styns, L. van Noorden, J.-P. Martens: Activating and relaxing music entrains the speed of beat synchronized walking, PloS One **8**(7), e67932 (2013)
34.75 L. Van Noorden, L. De Bruyn, R. Van Noorden, M. Leman: Embodied social synchronization in children's musical development. In: *The Routledge Companion to Embodied Music Interaction*, ed. by M. Lesaffre, P.-J. Maes, M. Leman (Routledge, New York 2017) pp. 195–204
34.76 S. Schütz-Bosbach, W. Prinz: Perceptual resonance: Action-induced modulation of perception, Trends Cogn. Sci. **11**(8), 349–355 (2007)
34.77 J.K. Witt: Action's effect on perception, Curr. Dir. Psychol. Sci. **20**(3), 201–206 (2011)
34.78 V. Halász, R. Cunnington: Unconscious effects of action on perception, Brain Sci. **2**(2), 130–146 (2012)
34.79 E. Van Dyck, P.-J. Maes, J. Hargreaves, M. Lesaffre, M. Leman: Expressing induced emotions through free dance movement, J. Nonverbal Behav. **37**(3), 175–190 (2013)
34.80 M. Varewyck, J.-P. Martens, M. Leman: Musical meter classification with beat synchronous acoustic features, DFT-based metrical features and support vector machines, J. New Music Res. **42**(3), 267–282 (2013)
34.81 L. Nijs, M. Leman: Interactive technologies in the instrumental music classroom: A longitudinal study with the music paint machine, Comput. Educat. **73**(2014), 40–59 (2014)
34.82 M. Lesaffre, L. Nijs, M. Leman: Interacting with music mediation technology for hearing impaired – First tests with normal hearing subjects. In: *Proc. 2009 Eur. Soc. Cogn. Sci. Music Conf. (ESCOM)* (2009)
34.83 O. Elschek: *Die Musikforschung der Gegenwart* (Stiglmayr, Wien-Föhrenau 1992)
34.84 M. Leman, A. Schneider: Origin and nature of cognitive and systematic musicology: An introduction. In: *Music, Gestalt, and Computing-Studies in Cognitive and Systematic Musicology*, ed. by M. Leman (Springer, Berlin, Heidelberg 1997) pp. 13–29
34.85 P.-J. Maes, D. Amelynck, M. Lesaffre, D. Arvind, M. Leman: The "Conducting Master": An interactive, real-time gesture monitoring system based on spatiotemporal motion templates, Int. J. Hum.-Comput. Interact. **29**(7), 471–487 (2013)
34.86 P.-J. Maes, D. Amelynck, M. Leman: Dance-the-Music: An educational platform for the modeling, recognition and audiovisual monitoring of dance steps using spatiotemporal motion templates, EURASIP J. Adv. Signal Process. **2012**(35), 1–16 (2012)
34.87 L. Nijs, P. Coussement, B. Moens, D. Amelynck, M. Lesaffre, M. Leman: Interacting with the music paint machine: Relating the constructs of flow experience and presence, Interact. Comput. **24**(4), 237–250 (2012)
34.88 P.-J. Maes, M. Leman, K. Kochman, M. Lesaffre, M. Demey: The "One-Person-Choir": A multidisciplinary approach to the development of an embodied human–computer interface, Comput. Music J. **35**(2), 22–35 (2011)
34.89 B. Moens, L. van Noorden, M. Leman: D-Jogger: Syncing music with walking. In: *Proc. SMC Conf. 2010, Barcelona (Universidad Pompeu Fabra)* (2010) pp. 451–456
34.90 M. Leman, M. Demey, M. Lesaffre, L. van Noorden, D. Moelants: Concepts, technology and assessment of the social music game "Sync-in-Team". In: *Proc. 2009 Int. Conf. Comput. Sci. Eng.*, Vol. 4, ed. by J. Calder (IEEE Computer Society, Vancouver 2009) pp. 837–842
34.91 M. Demey, C. Müller, M. Leman: DanSync: A platform to study entrainment and joint-action during spontaneous dance in the context of a social music game. In: *INTETAIN 2013*, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Vol. 124, ed. by M. Mancas, N. d'Alessandro, X. Siebert, B. Gosselin, C. Valderrama, T. Dutoit (Springer, Mons 2013) pp. 124–135
34.92 M. Clayton, B. Dueck, L. Leante (Eds.): *Experience and Meaning in Music Performance* (Oxford Univ. Press, Oxford 2013)
34.93 N. Moran: Music, bodies and relationships: An ethnographic contribution to embodied cognition studies, Psychol. Music **41**(1), 5–17 (2013)
34.94 F. Bonini-Baraldi: *Tsiganes, Musique et Empathie* (Editions de la maison des sciences de l'homme, Paris 2013)
34.95 M. Rahaim: *Musicking Bodies: Gesture and Voice in Hindustani Music* (Wesleyan Univ. Press, Middletown 2012)
34.96 E. Wilf: *School for Cool: The Academic Jazz Program and the Paradox of Institutionalized Creativity* (Univ. of Chicago Press, Chicago 2014)
34.97 M. Leman: *The Expressive Moment: How Interaction (with Music) Shapes Human Empowerment* (MIT Press, Cambridge 2016)
34.98 M. Lesaffre, P.-J. Maes, M. Leman (Eds.): *The Routledge Companion to Embodied Music Interaction* (Routledge, New York 2017)

35. Sonic Object Cognition

Rolf Inge Godøy

We evidently have features at different timescales in music, ranging from the sub-millisecond timescale of single vibrations to the timescale of a couple of hundred milliseconds, manifesting perceptually salient features such as pitch, loudness, timbre, and various transients. At the larger timescales of several hundred milliseconds, we have features such as the overall dynamic and timbral envelopes of sonic events, and at slightly larger timescales, also of various rhythmic, textural, melodic, and harmonic patterns. And at still larger timescales, we have phrases, sections, and whole works of music, often lasting several minutes, and in some cases, even hours.

Features at these different timescales all contribute to our experience of music, however the focus in the present chapter is on the salient features of what has been called *sonic objects*, meaning on holistically perceived chunks of musical sound in the very approximately 0.5—5 s duration range. A number of known constraints in the production and perception of musical sound as well as in human behavior and perception in general, seem to converge in designating this timescale as crucial for our experience of music.

The aim of this chapter is then to try to understand how sequentially unfolding and ephemeral sound and sound-related body motion can somehow be transformed in our minds to sonic objects.

35.1 Object Focus

The aim of this chapter is to present theories and tools for research on what we call *sonic objects*. The term *sonic object* can be defined as a fragment of musical sound, typically in the approximately 0.5—5 s duration range, a fragment perceived holistically as a coherent and somehow meaningful unit. A sonic object is arguably the most basic unit in musical experience, capable of making a rich set of perceptually salient sonic and multimodal features present in our minds.

A sonic object may encompass a single tone or chord, a short phrase of several tones and/or chords in succession, a single sound event (e.g., of hitting a tamtam, of slamming a door, of breaking a bottle), or a more composite but still holistically perceived sound event (e.g., a rapid glissando on a harp or on a washboard, a burst of marbles rolling out onto the floor, a whirl of dry leaves in the wind). The duration limits of a sonic object are determined at one end by the minimal duration necessary to perceive salient features and at the other end by a maximal duration for perceiving the object as a singular and coherent entity, i. e., as not readily divisible into smaller parts.

The term *object* here serves to emphasize the perception of a sound fragment as a coherent entity,

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_35

as something present in our minds in an instant, in a *now-point*, to borrow the expression from *Edmund Husserl* [35.1–3], although sound is something that unfolds in time. How sequentially unfolding sound is transformed into somehow more solid auditory objects in our minds seems still to be largely enigmatic [35.4], however, the point of departure in this chapter is that such *flux-to-solid* transformations evidently do take place in auditory perception and that we may indeed denote salient auditory features by way of object-related concepts and metaphors. Focusing on the *cognition* of sonic objects means trying to provide tools for exploring perceptually salient features of sonic objects in music-related research and in practical contexts for composers, musicians and producers; in short, for anyone working with musical sound.

The idea of sonic objects is generally ascribed to the seminal work of *Pierre Schaeffer* et al. in the 1950s and 1960s and as emerging from practical work in electroacoustic composition of the so-called *musique concrète* [35.5–7]. Before the advent of tape recorders, composers would record sound fragments as loops on phonograph discs, enabling the mixing of sounds by lowering and raising the pickup arm, effectively turning any sound fragment on and off in a mix. Such a loop was called *closed groove* (*sillon fermé*) and when Schaeffer et al. listened to such loops innumerable times, they discovered that their perception of these sound fragments changed, that they tended to shift their attention towards more internal and subtle features of the sound itself, away from the original and more everyday and anecdotal significations of the sound fragments. They called this shifting of attention *reduced listening* (*écoute réduite*), signifying a shift toward perceptually salient sonic features, a shift of focus that eventually lead to a very extensive theory of sonic objects in *Schaeffer*'s monumental *Traité des objets musicaux* [35.5] and related publications [35.6, 7].

This origin is remarkable in that an extensive music theory grew out of practical composition work, and notably out of a radically new way of sound-based composition, very different from traditional note-based Western composition practice. This departure from Western music tradition drove the effort to make a more general and universally applicable music theory, based on a seemingly naïve questioning of the subjective listening experience, a kind of Socratic method of trying to distinguish the various perceptually salient features in auditory experience. This approach permeates the entire project of developing a new music theory and may be termed a *top-down* approach where the point of departure is the overall shapes of any sound, what we could call its *envelopes*, with regards to loudness, pitch features and timbral features and then successively differentiating more and more subfeatures of these main features (e.g., the distribution in time and spectrum of the energy, the amplitude and rate, regularity versus irregularity, of changes in these distributions, etc.) and then, sub-subfeatures of these, etc., progressively exploring more and more details in the sonic objects. It was only at a later stage that *Schaeffer* et al. realized the affinity of this method with classical phenomenological approaches to perception (referring to Husserl and Merleau-Ponty) of taking the subjective mental images of sound as point of departure for investigations, leading *Schaeffer* to remark that they were indeed *doing phenomenology without realizing it* [35.5, p. 262].

The next step of this research program was then seen as establishing *correlations* between the subjective features and the acoustic substrates of the sound, a long-term endeavor that needed to take into account the many nonlinear relationships between the physical signals and the subjective percepts, relationships that were characterized by *anamorphosis*, or *warping*. It should be remembered that psychoacoustics as we know it today was quite different in the 1950s and 1960s when Schaeffer et al. developed their theories and that research in this area was dominated by a more *scientistic* attitude as *Michel Chion* has put it [35.7, p. 30], of regarding human auditory perception as flawed, unreliable and often distorting the *real* features of sound. The attitude of Schaeffer et al. was quite different in naming subjective perception of sound the most important tool for research, as the point of departure for extensive and systematic explorations of musical sound, and only at a later stage going on to map out the correlations between subjective perception and the acoustic signals.

The term *cognition* in the title of this chapter implies a focus on the perceptual output of any sound-producing process, be that in sound synthesis, effects processing, composition, improvisation, or any kind of instrumental or vocal performance. The idea of sonic object cognition is also linked with the method of *analysis-by-synthesis*, meaning exploring something by active creation (and/or imagining) of incrementally different variants in view of finding out what the perceptually salient features (or ingredients) of any sonic object are. Lucidly presented by *J.-C. Risset* as a strategy for exploring timbral features in digital sound synthesis [35.8], this was actually also a strategy of Schaeffer et al. in exploring the contribution of different feature dimensions to the overall subjective impressions of sonic objects (for an instructive example of this, listen to [35.6, CD 2, tracks 90–95]). In other words, the idea of sonic object cognition here denotes a two-stage process of first trying to differentiate (and give names to) what seem to be subjectively salient features of

a sonic object and then produce, or imagine, a number of variants of this sonic object where these features are somehow incrementally varied, so as to enable systematic testing of the effect of these features in subjective experience.

In short, sonic object cognition encompasses the capacity to think analytically and practically about fragments of sound in musical contexts and should aspire to combine insights from classical work (e.g., from composition and orchestration) with recent findings (e.g., from musical acoustics, music technology and music perception areas). With such a multitude of elements converging on sonic object cognition, we shall in this chapter first elaborate more on what is meant by the expression *sonic object*, try to demarcate what it is and what it is not, with some considerations of ontology: sonic objects are evidently associated with our experiences of sound, but sound is ephemeral, manifest here and now before vanishing, yet we fortunately usually have some memory trace of what we just heard. Are sonic objects then just as much mental images and in that case, what is the content of such mental images? One answer that has emerged from our own and others' research is that images of sonic objects are closely associated with images of body motion, so we shall then take a look at what is called *motor theory*, as well as associated issues of timescales and duration thresholds in musical experience. This leads to questions of how sonic objects emerge from musical experience by *chunking*, as well as how they are generated and classified. On this background, the remaining sections of this chapter shall focus on features of sonic objects as well as on how we may explore such features by the above-mentioned analysis-by-synthesis strategy, before some reflections on prospects and challenges of research on sonic object cognition.

35.2 Ontologies

Sonic objects, defined as fragments of musical sound in the approximately 0.5–5 s duration range, will in most cases have multiple significations and multiple features, i. e., be several things at once. To enhance our knowledge of sonic objects, as well as to avoid misunderstandings, we need to clarify *what-is-what*, or clarify the *ontologies*, of sonic objects. As a first step, we may take a look at Schaeffer's analysis of listening, an analysis that is the basis for the subsequent ontological differentiations of sonic objects. Briefly stated, Schaeffer distinguished four components in listening (see [35.9, pp. 129–133], for a more extensive presentation):

1. *Listen* (*écouter*), which denotes the basic capacity for discerning different sounds in our environment.
2. *Hear* (*ouïr*), i. e., the basic physiological capacity for sensing sound.
3. *Hark* (*entendre*), meaning the intentional focus on some sound(s).
4. *Understand* (*comprendre*), denoting the transition from basic listening to understanding the significations of the sounds.

In everyday situations, these four components may variably be active or not, e.g., if I hear the squeaking of a door when I am expecting a visitor, I would understand this sound as indicating the arrival of my visitor and probably not be so much interested in the sonic features of the squeak. But being sensitive to such squeaking noises, I could direct my attention towards the squeak, making a note that I should lubricate the hinges to get rid of the squeak, or I could even be intrigued by the brass instrument-like timbral features of the squeak, its pitch and dynamical envelope; in short, start to focus on the squeak as a sonic object. Shifting my attention away from the contextual and/or causal signification of the squeak to its sonic features, is typically an act of the *reduced listening* mentioned earlier.

Furthermore, that we, in the case of electroacoustic music as well as in several other listening situations, only hear the sound and don't see the source of the sound, was by Schaeffer called *acousmatic listening*, alluding to the account of Pythagoras allegedly hiding behind a screen when teaching so that his pupils should not be distract by seeing him. The principle of acousmatic experience signifies a general divide between the production and the perception of musical sound, encouraging us to focus on the actually perceived features and to disregard whatever generative scheme is behind the sounds that we hear.

This divide between production and perception could also be seen as a critique of some prominent 20th century Western music that advocated generative schemes (e.g., dodecaphony, integral serialism, various algorithmic composition schemes and also more recent sonification schemes). It could be generalized to a model consisting of a *control space* and a *morphology space* as was suggested by *morphodynamical* theory [35.9–11]. The control space is concerned with the control variables input to any generative process, e.g., the oscillator frequencies, amplitudes, envelope

shapes and modulation index in a frequency modulation (FM) synthesis model, and the morphology space is concerned with the perceptual output of any model, e.g., the timbral output of the FM synthesis model. As anyone familiar with FM synthesis probably has experienced, there may sometimes be a linear, seemingly coherent relationship between the input variables and the perceived output features, however, there may also often be a nonlinear relationship between incremental changes in control input values and perceived changes in output features, e.g., that small input value changes may result in disproportionately large and seemingly uncorrelated changes in output features.

Such discrepancies between changes in control input and perceived output may be a general risk when transferring features from one domain to another: it is not obvious that we readily perceive the relationships between any two domains, e.g., when sonifying visual or numerical information [35.12]. Such discrepancies of input and output are then a matter of mistaken ontological identity, encountered in various Western 20th century compositional schemes when an ordering scheme from one domain is uncritically transferred to another domain [35.9], and may be generalized as mistaken *mappings*. Being careful in making perceptually pertinent mappings is one of the main issues of interactive sonic design and of new instruments of musical expression [35.13], and such mappings should be based on careful analysis of the salient perceptual feature dimensions involved in sonic objects, not on more abstract numerical relationships.

Furthermore, given the acousmatic principle of disregarding the source of any sonic object, the next stage becomes an ontological analysis of the sonic object as such. First, we can list what Schaeffer indicated that the sonic object is *not* [35.7, pp. 34–35]:

- The sonic object is not the sounding body.
- The sonic object is not the physical signal.
- The sonic object is not a fragment of a recording.
- The sonic object is not a symbol notated in a score.
- The sonic object is not a state of the mind.

Furthermore, Schaeffer made an extensive differentiation of salient features that may be present in parallel in sonic objects (more on this later), noting that our intentional focus may wander from one feature or set of features to another feature or set of features, thus making it difficult or even futile to try to pin down exactly what a sonic object might be at any moment. *Schaeffer*'s conclusion after various ontological considerations and differentiations was that the sonic object is an *intentional unit*, meaning that it has several aspects, both sequentially and in parallel, that are kept together in our minds by our active mental focus [35.5, p. 263].

In the decades after the publication of Schaeffer's *Traité* in 1966, various psychoacoustic and neurocognitive research has focused on what we could broadly call *auditory object perception*. Work on so-called *auditory scene analysis* [35.14] has demonstrated a number of low-level signal features that contribute to our perceptual judgments of sonic objects, such as spectral coherence and qualitative discontinuities, documenting the effects of various gestalt principles, as well as the interaction of these low-level features with more high-level schema-based factors in our ability to discern sonic objects in listening [35.15]. Other recent neurophysiological research seems to confirm the basic gestalt-related principle of so-called *exclusive allocation* [35.16], suggesting similarities with visual domain object criteria, something that is also pursued in mapping out the different brain processes assumed to be involved here, as well as suggesting that there are cross-modal elements at work in auditory object perception [35.4]. This last point is of particular interest to our subsequent sections of multimodal features of sonic objects, in particular body motion and haptic features that clearly help us to grasp the ephemeral features of sound as more solid objects in our minds.

This research on auditory objects may be summarized as concerning the question of *coherence* in auditory perception and as also suggesting that there are both similarities and links between the different sense modalities at work here. Although there are still many unanswered questions, one candidate for such sonic object coherence is the motor schemas involved in auditory perception, an idea that has been much in focus in the so-called *motor theory of perception* and which will be an important element in the present chapter.

35.3 Motor Theory

The gist of the motor theory of perception is that we mentally (and sometimes even overtly) simulate body motion associated with whatever it is that we are perceiving, e.g., that when we hear ferocious drumming we mentally simulate energetic hand motion, or when we hear soft string music, we mentally simulate slow protracted bowing.

In our context of sonic object cognition and Schaeffer's research strategy of disregarding sources and ordinary significance, it is also clear that Schaeffer's

strategy does not exclude a motor theory approach to how we perceive and classify sonic objects [35.17]. The point of motor theory is that sensations of body motion and postures are fundamental and ubiquitous in human cognition, hence also is the basis for sonic object perception, meaning that sonic objects may be perceived in terms of body motion and body posture schemas and not only as *pure sound*.

There are several links between sound and motion in music and we have in the past decades seen a growing number of research projects and publications in this area (see Chap. 38 this volume, and [35.18] for overviews). We have in the past decades also seen a general trend in the direction of understanding body motion as essential in most areas of human cognition [35.19], such as in social contexts [35.20] and also in abstract thought [35.21]. Motor cognition may be seen as integral to several modalities, e.g., in vision with active simulated tracing of what we are seeing [35.22], or in hearing with active simulation of the sound-producing vocal apparatus motion as suggested by the motor theory in linguistics [35.23, 24]. The motor theory perspective has been progressively better supported with the arrival of so-called noninvasive brain observation methods [35.25–27]. Also various behavioral studies have clearly demonstrated the strong links between body motion and sound perception, e.g., the effects of seeing lip motion on the perceived sound in the so-called *McGurk effect* [35.28].

What is essential here is the overall energy and motion trajectory images associated with sonic objects, meaning that such images are generic and transferable from one sound source to another as long as the overall envelopes are reasonably similar. This is the deeper and more general understanding of the reduced listening: departing from the everyday and anecdotal source and signification (e.g., crashing car, breaking twig, marble tumbling down a staircase, etc.) to a more generic energy-trajectory envelope [35.17].

The basis for mental presence of the sonic object could be called *motor-mimetic cognition* [35.29–31]. The idea of motor-mimetic cognition is that any feature of sound, and of musical experience in general, can be traced as a shape. This is a strong version of motor theory and one that can be used here to explore features of sonic objects. Specifically, we see this motor-mimetic behavior in sound imitating activities, such as in onomatopoeias and in *beatboxing*, *scat singing*, but also clearly evident in so-called *air instrument performance*, where people with little or no musical training seem to be able to simulate sound-producing body motion [35.32] and in so-called *sound-tracing*, where people are asked to trace the body motion shape they believe best reflect salient features of sonic objects they are hearing [35.33].

We have various kinds of music-specific body motion, all of which may variably contribute to the motor images of sonic objects; body motion that may be categorized as follows [35.18]:

- *Sound-producing* body motion. This includes various kinds of *excitatory motion* such as hitting, stroking, bowing and blowing, as well as *modulatory motion* that modifies the sound, such as the left hand motion on string instruments, the closing and opening of mutes on brass instruments, etc. In addition, various kinds of so-called *ancillary motion* belong in this main category: motion that is not strictly sound-producing but still necessary in order to avoid fatigue or strain injury, or to help shaping the musical expression and also *communicative motion* in relation to fellow musicians or the audience.
- *Sound-accompanying* body motion. This includes dancing, walking, gesticulating, tracing, etc., often reflecting sound-producing body motion, but often more vaguely the energy and affect suggested by the musical sound.

Also, there are a number of more global, i. e., nonlocal, music-related body motion sensations, such as calm–agitated, fast–slow, smooth–rugged, regular–irregular, etc., where there are also often correlations with affective features. Hence, we should next consider some elements of timescales and duration thresholds in relation to sonic objects.

35.4 Timescales and Duration Thresholds

It is well known that we have a timescale range from approximately 20–20 000 events per second for perceiving pitch as well as stationary timbral and loudness features of sounds [35.34]. In our context of sonic objects, the sub-20 Hz region is of particular interest, as this is where we have the nuances and fluctuations of pitch, loudness and timbre, including microtextural elements such as trills and tremolos, as well as various slower elements such as envelopes of pitch, timbre and loudness of tones, and sensations of motion (gait) and rhythmical and melodic patterns.

Also in the sub-20 Hz region we have feature duration thresholds, meaning how long an excerpt we have to hear in order to get some impression of salient

features. It has been suggested that we may perceive salient features of sound in fragments as short as 350 ms [35.35], but for other musical features such as rhythmical, melodic and harmonic patterns, we have to hear longer fragments, i. e., approaching the upper limit for sonic object duration of approximately 5 s. The question then becomes that of trying to see what is *inherent* in the sound as necessary minimum duration for the recognition of salient features and what is more a matter of the peculiarities of our perceptual and cognitive processing. Hence, we should try to determine duration thresholds from two points of view:

1. That which captures most salient information, i. e., neither too short nor too long, e.g., sufficiently long for perceiving salient basic dynamic, pitch-related and timbral features, as well as more high-level features of rhythm, texture, melody, harmony, style, genre, sense of gait and affect, but not so long as to seem being redundantly repetitious.
2. That which is intrinsically in accordance with our cognitive constraints and/or predispositions. This includes what are typical human action durations, optimal durations in motor control, durations of short-term memory stores and of attention spans and timescales for awareness.

Both points of view seem to be relevant and they sometimes seem to converge with respect to duration, perhaps by adjustment of salient information timescales in human communication with ecological attention timescales, as suggested by *Ernst Pöppel* [35.36]. What seems to be clear is that we have rather different qualitative features at different timescales, and for this reason we have in our own research on sonic objects and music-related body motion applied the following main timescale categories:

- *Micro timescale*, typically in the sub-0.5 s range and including continuous and quasistationary features such as pitch, dynamics and stationary timbral features. At this timescale, we may also have some fast fluctuations in the sound such as rapid tremolos and trills, what we shall with Schaeffer call *grain*.
- *Meso timescale*, typically in the 0.5–5 s range, encompassing whole sonic objects. This timescale includes most salient musical features such as rhythmical, textural, melodic, timbral, etc. patterns, readily enabling identification of genre, style, sense of gait and affect.
- *Macro timescale*, typically in the above-5 s range, including whole phrases, sections and movements and usually consisting of concatenations of sonic objects of the meso timescale. The macro timescale is also that of narratives or dramaturgy, however, it seems not to be so well researched in terms of actual perceptual features.

One crucial feature of the meso timescale is that sequences of elements (tones, sounds) at this timescale give rise to sonic objects with qualitative new features not present at smaller timescales, e.g., a melodic contour is present at the meso timescale as a sonic object because it is *present as one extended and coherent object with a shape and not primarily as a sequence of individual tones*. Similarly, body motion at this meso timescale is also readily perceived and conceived as forming coherent motion objects with shapes, and details of such object formation are found in the topic of chunking.

35.5 Chunking

We may think of music as a continuous stream of auditory, visual, proprioceptive, haptic, etc. sensations, yet it seems that we make sense of such streams by segmenting them into somehow meaningful units, into what we call *chunks*. This process of chunking is the basis for the perception of sonic objects, hence one of the main topics in this chapter. Chunking here means not just a segmentation of a continuous stream, but also a *transformation* of otherwise sequential events or entities into qualitative new and larger units, making the formerly single events/entities totally fuse, or *disappear*, into the new chunk unit in the sense of belonging exclusively to this new unit by the so-called *exclusive allocation* principle of gestalt theory [35.14].

There are two main approaches to chunking in music, *endogenous* and *exogenous* [35.37]:

- *Endogenous chunking* is based on qualitative discontinuities in the signal such as that between sound and silence, between different pitches, timbres or levels of loudness or various repeated patterns, e.g., metrical, melodic or harmonic. Yet this may not always work, either because of insufficient discontinuities, e.g., protracted and unchanging or slowly changing sounds, or because of competing discontinuities, e.g., in a series of staccato tones alternating between sound and silence, necessitating the projection of endogenous chunking schemas onto the sounds.

- *Exogenous chunking* is based on the projection of schemas onto what we perceive, schemas based on what we have acquired in previous experience and in particular, various motor schemas acquired from experiences of sound and motion production.

In cases of exogenous chunking, i. e., when applying schemas from previous experience onto sensory input, we can see that such schemas are often constraint-based in their origin. We should thus consider some constraints at work in traditional (i. e., pre-electronic) means for sound production. This includes various features of the instrument and room acoustics (reverberation) that may shape our experience of chunks, e.g., by incomplete damping of sounds resulting in a smearing that glues otherwise distinct sonic events together in qualitative new and larger chunks. But we also have constraint-based chunking originating in our organism, both of attention [35.36], and related to motor control [35.38, 39]. Specifically, we have the following chunk-inducing constraints in musical performance:

- *Phase transition*: There are thresholds of grouping dependent on duration and rate of events [35.40], typically causing singular events to fuse into higher-level units with increase in event rate (e.g., singular impulses fuse into a tremolo with acceleration). Conversely, units may split into lower-level units with decrease in event rate (e.g., a tremolo split into a series of impulses with deceleration).
- *Coarticulation*: The contextual fusion of otherwise distinct motion and sound by the effector (finger, hand, arm, vocal apparatus) always being in a context of having just made some sound and is about to make another sound in the immediate future and that there is a corresponding contextual smearing of the resultant sound because of incomplete damping between the sound events [35.41].
- *Goal points*: That human motion control is goal directed [35.42], typically resulting in so-called *key-postures* at salient moments in time, i. e., at downbeats and other accented points in the music, resulting in chunks being centered on these goal points [35.43].

It seems that various research converges in documenting motion chunking in human behavior in general and rhythmic gestalts in particular [35.39]. As noted by *Klapp* et al. in connection with polyrhythmic patterns [35.44, p. 318]:

> *The limitation to only one motor Gestalt may be analogous to limits that arise with visual patterns such as the Necker cube. That figure can be perceived in only one of its configurations at any given instant. In either configuration, however, all of the lines of the cube are perceived simultaneously as one pattern. Thus, the Gestalt is not restricted in terms of the number of lines that can be perceived. Instead, the limit is that only one organization can be activated. Similarly, the limit in concurrent motor actions is assumed not to lie in the number of muscles that can be controlled, but, instead, the limit is that only one action pattern can be active.*

In other words, it seems that even rather complex patterns of motion may be conceived and perceived as a chunk in motor control.

In summary, we may conclude that sonic objects represent the convergence of chunking factors from the signal (constraints and qualitative discontinuities in both sound and motion), the sense of motion (internal sense of effort and proprioception, i. e., action gestalts, goal points, coarticulation, phase transitions, etc.) and attention constraints; hence, the convergence of both the exogenous and the endogenous factors of chunking in musical experience.

35.6 Sound Generation

In the beginning of the *musique concrète*, the sound fragments used were typically so-called *found objects*, meaning ready-made fragments of sound from the environment, (human, animal, mechanical, etc., but also from more conventional instrumental or vocal sources), hardly processed beyond the initial cutting before the subsequent concatenations into musical compositions. As for the boundaries (i. e., start and stop points) of these fragments, Schaeffer noted that these could be either *natural* in the sense of occurring at some qualitative discontinuity in the sound, e.g., between sound and silence, or they could be *artificial* in the sense of being cut out of a context. In the latter case, the effect of the cutting would in turn become an integral part of the sonic object, e.g., as a steep attack on an otherwise smooth, sustained sound.

Although the acousmatic attitude and the strategy of reduced listening encourages us to focus on intrinsic perceptual features of sound, we know that such fea-

tures are often the result of peculiarities of the sound source. We should then also consider some aspects of sound generation in general, including more traditional instrumental and vocal sound generation as well as sound synthesis and processing, which contribute to our perception of sonic objects. As suggested by auditory scene analysis research [35.14], we tend to perceive sounds as emanating from a source because of spectrotemporal coherence, i. e., that the spectral components tend to fuse into one sonic object because of low-level gestalt principles such as synchrony of onsets and motion of partials. Yet also, there are cases when schemas from past experience come into play in identifying sound sources, i. e., where there is great spectral variation over the entire pitch range and different playing modes of an instrument (e.g., the full range of a piano), yet where the instrument is usually perceived as *the same* in spite of such significant differences.

Also, we should remember that motor theory suggests that production schemas are projected onto what we hear, largely independent of the source and it may be argued that this also applies to acousmatic listening [35.17]. The point is that motor theory in general applies to any perceptual feature by its relation to body motion kinematics, haptics, proprioception and effort, something that was in fact suggested also by Schaeffer in the three basic dynamic envelopes of sonic objects, regardless of source:

1. *Sustained*: Continuous transfer of energy from the body to the instrument such as in bowing or blowing, resulting in a more or less stable and continuous sound.
2. *Impulsive* A spike or sudden peak of effort such as in hitting or kicking, resulting in a sudden attack in the sound followed by a longer or shorter decay.
3. *Iterative*: A rapid back and forth motion as in a tremolo or a trill, resulting in fast ripple-like features in the sound.

These main categories of sound-producing motion and the corresponding resultant sounds are subject to phase transitions dependent on rate and duration of the events: shortening a sustained motion turns it into a impulsive motion, and conversely, lengthening an impulsive motion turns it into a sustained motion, and slowing down an iterative motion turns it into a series of impulsive motions, etc.

Additionally, there are also body postures projected onto the sound by way of past knowledge of sound-producing motion, e.g., narrow positioning of hands versus spread positioning of hands on a keyboard, or vocal tract shapes of open, closed, pointed, etc. The motor theory perspective suggests that in general, there is a projection of assumed body motion onto whatever it is that we are hearing and that this in particular goes for overall sense of effort, including sense of velocity, acceleration and jerk in the music; in short, that music has what could be called rich gestural affordances [35.45], and that all these sensations of body motion evoked by the sound contributes to our perception and subsequent *objectification* of the music.

In summary: Schemas of sound production may reflect basic body motion schemas, regardless of the sound source in question, meaning that sonic object cognition is largely determined by body motion shapes, in turn based on a number of constraints. In the case of very long and stationary or slowly evolving sounds, the energy envelopes may tend to be perceived as *non-human* in that there are no rests in effort and also that the duration may be above the usual thresholds of attention, c.f. *Pöppel*'s suggestions of the perceptual present and the tendency to shift attention in experiences lasting significantly longer than approximately 3 s [35.36].

35.7 Constraints and Idioms

Needless to say, the scope of musical expression is vast, if not infinite, but there are also a number of constraints in musical sound production. The physics of musical instruments and the ergonomics of sound production together make up a number of constraints on music making, constraints that in turn also influence our perceptions of sonic objects.

The basic scheme of sound production is that of energy transfer from the human body to the instrument by way of hitting, stroking, plucking, bowing, blowing, etc. and the response of the musical instrument to any such excitatory body motion. The energy dissipation of an instrument is then a shaping factor of sonic objects, providing its overall envelopes of pitch, timbre and loudness, as well as a number of transients and fluctuations in the course of the sonic object. Additionally, the room acoustics, e.g., the reverberation and other resonant features, contribute to the features of the sonic object. The patterns of excitation and subsequent energy dissipation are highly characteristic of different instruments and people seem to possess quite extensive knowledge about these patterns [35.46].

The combined physical and ergonomic constraints of any instrument typically result in what is commonly

called *idioms*. Idioms are sound-producing motions, on any instrument or the human voice, that are considered particularly well-sounding and also comfortable or easy to perform, hence based on successful combinations of instrument physics and ergonomics. Idioms may variably so contribute to images of sonic objects and are of particular importance in the more composite sonic objects we find in orchestration (more on this later).

The most basic constraint of music performance is that *all sonic events are embedded in action trajectories* and that there is no immediate transition from one sound-producing event to another, i. e., that all effector motion (by fingers, hands, arms, lips, tongue, feet) takes time and that all effector motion (regardless of speed) makes for continuous trajectories. In sum, we have the following constraints of sound-producing body motion contributing to sonic object cognition:

- *Effort constraints*, i. e., there are limits to endurance and musicians need to take rests when performing to avoid fatigue and strain injury.
- *Speed constraints*, related to the previous constraint, but partly also a matter of motor control.
- *Phase transitions*, i. e., the transitions based on rate, speed and duration of basic motion-sound categories mentioned earlier.
- *Coarticulation*, based on the fact that human motion cannot be instantaneous and that there always is a transition from one effector position to another effector position, hence there always are both so-called *spillover effects* and *anticipatory effects* in performance, leading to a contextual smearing of both motion and sound.
- *Goal points*, meaning that performance motion is organized hierarchically and that these hierarchies are also (variably so) reflected in the shaping of the sonic objects.

In summary, there are a number of constraints on traditional sound production in music, constraints that can be summarized as concerning envelopes of energy and as forming schemas that also seem to extend to electronic music.

35.8 Sound Synthesis

Although available technologies for sound synthesis and sound processing present great possibilities for generating a large variety of sonic objects, they also present substantial challenges of how to control the parameters involved in this generation (see Chap. 38 for more on this). The main point in our context is then how to access perceptually salient sonic object features in sound synthesis and processing. (The division between *synthesis* and *processing* may not always be so clear cut, e.g., convolving two signals may be a case of cross-synthesis or source-filter synthesis, but may also be a case of effects processing, e.g., adding reverb to a sound or simulating various other room acoustic coloring of sound.)

This is initially much dependent on the type of synthesis model being used. The main classification into so-called *signal models* and *physical models* may be useful here: physical models shall in principle incorporate some element of energy input (excitation) and energy dissipation (resonance), albeit often in a highly simplified form. On the other hand, the signal models are inherently more abstract in the sense that there is really no simulation of real-world excitation and resonance except for what we may put into the control parameters, e.g., by designating envelopes that approximate expected behavior of real-world sonic events such as of plucking strings, hitting membranes, or stroking metal plates.

Compared with traditional musical instruments and the human voice, there are then no transfers of energy, no motion constraints and and no idioms, involved in sound synthesis (although for the expert listener there may be detectable peculiarities of the synthesis models being used, in a sense resembling *idioms*, e.g., of the often-heard FM and granular synthesis models). This does not preclude that we may project various schemas from past musical experiences onto new kinds of sonic objects (so-called *anthropomorphic projection*), or that the composition is organized more in accordance with traditional score arrangements of sounds (as is sometimes the case in multitrack music production). What seems to be an inherent tendency in our perception is then that of grouping sounds into events and sources as if they were originating from real-world sources and that we project various motor schemas onto sonic objects that have really no body motion element in their origins.

However, the desire to introduce more nuances and expressiveness into the domain of sound synthesis and processing has in the past couple of decades lead to much effort in developing new means for real-time control, including new interfaces of various kinds (e.g., New Interfaces for Musical Expression (NIME) conferences), as well as extending traditional musical instruments with new control possibilities. The latter development has the advantage of exploiting musicians'

already acquired skills, as well as to explore the sonic territory stretching from that of traditional instruments into new sounds.

Besides the ergonomics, one main issue here is that of mapping, i. e., of how body motion is assigned to the control parameters of whatever sound synthesis or processing model is being used. Different schemes of mapping have been explored [35.13], however one crucial question here is actually that of trying to determine what are the perceptually salient feature dimensions of sonic objects and how these may be correlated with body motion features.

35.9 Feature Taxonomy

Schaeffer's work suggests that any sound whatsoever, natural or synthesized, may be considered a sonic object provided it fulfills the criteria of having a suitable duration and capable of being perceived as a holistic and coherent entity. The challenge then becomes that of developing some kind of conceptual apparatus that enables us to diagnose salient perceptual features of sonic objects, i. e., to differentiate these features and compare their relative variable values (e.g., degree of dynamic fluctuation, degree of stability of perceived pitch, etc.).

The project of developing a feature taxonomy becomes a multistage process of firstly discerning and naming some feature that we believe is salient in our subjective image of a sonic object, and secondly to distinguish further what may be subfeatures of this main feature and sub-subfeatures of these features, etc., as far as we see useful for our analytic and/or practical purpose at any time. This means applying a number of metaphorical labels and try to define what they refer to in the sonic object (e.g., the fast dynamic fluctuation as the *grain* of the sonic object). The next step will then be to see these metaphor-labeled features as dimensions in a multidimensional feature space, where each dimension may have a minimum and a maximum value (e.g., between minimum and maximum amplitude of the grain fluctuations) and where any sonic object may be positioned in such a multidimensional feature space.

Such a scheme is very general, applicable to any sonic object and in no way limited to our traditional Western music theory. It is not constrained by a symbolic notation system and all features may be taken into consideration. In this sense, Schaeffer's feature taxonomy is concerned with *concrete* and not with *abstract*, features [35.7, p. 39] and [35.9, p. 216]. There are of course categories at work here, e.g., for pitch and for timbre, but the point is that the feature space of sonic objects is much more extensive than that afforded by traditional Western music theory and its associated categories. As we know, e.g., Western categories of pitch are not universal and other musical cultures (e.g., in India) have more nuanced concepts for pitch. In this sense, the symbolic versus subsymbolic divide is bypassed by Schaeffer's taxonomy and we could add also the symbolic versus suprasymbolic divide as well, in the sense that several fused tone events (or sound events in the case of nonpitched sounds) may very well form a new and larger sonic object with its own overall salient features. In other words, this may include features beyond those of mainstream analysis (including music information retrieval (MIR) and its so-called chroma vectors (i. e., pitch extraction) and other pattern retrieval schemes), features that so far are not well captured by symbolic queries but which are highly significant in musical experiences, e.g., phrase shapes and other meso level emergent features.

In the decades since the publication of Schaeffer's *Traité* in 1966, there has been significant progress in detecting and representing a large number of previously unnamed features in psychoacoustics [35.47, p. 326]. Also software development, e.g., like the *MIRtoolbox* and the *Timbre Toolbox* and *Praat*, has made it much easier now to work directly with sonic object features such as spectral shapes, spectral centroid, spectral flux, formantic shapes, etc.

However, as argued earlier, sonic object cognition will not only include acoustic features, but also corresponding motion features, i. e., trajectories, contours, postures, etc.; that is, various perceptually salient features that may be correlated with acoustic features, as was the long-term goal of Schaeffer's research program.

The main principle of Schaeffer's feature taxonomy is that any feature whatsoever may be given a shape-related metaphor labeling. This gives us a very general and systematic conceptual apparatus for dealing with features in a lucid manner, as well as bridging the gap towards quantitative data of sound and motion.

35.10 Shape Cognition

Sound, as well as associated sensations of body motion, are ephemeral, yet may evidently also result in more solid mental images in our memory. That we now have the technologies for capturing, replaying and making various representations of the ephemeral is indeed one of the most significant developments of our times, as it was, albeit to a much more modest extent, in the late 1940s with the beginning of the *musique concrète*. However, these possibilities of freezing sound and motion also raises new questions of what kind of knowledge we may represent by these various technology-based schemes, and what are the fundamental knowledge goals of our research, questions that point to *shape cognition* as a very basic epistemological and also quite pragmatic paradigm for representing sonic object features.

The term *shape cognition* designates our capacity for thinking and representing sound and motion features visually, as geometric objects in two or more dimensions and as distinct from symbolic representations. As for the symbolic representations, ranging from common-practice Western notation to more recent schemes such as MIDI (musical instrument digital interface) and MusicXML (music extensible markup language), one major challenge is to overcome the discrete bias that prevents us from having good notions of sonic objects, i. e., we need representations that highlight the actual continuous sonic unfolding and make the individual Western pitches and durations fused into coarticulated, coherent sonic objects, as well as to represent the many significant nuances of musical sound. We have in the course of recent decades seen some interesting attempts to more closely represent the holistic features of sonic objects as shapes:

- Enhanced spectral representations of actual musical sound, from the pioneering work of *Robert Cogan* several decades ago [35.48] to the presently readily available software, e.g., the earlier mentioned toolboxes.
- Animations of various kinds that represent the unfolding of sonic objects, be that based on spectral images, on MIDI data, or conventional notation.
- Subjective sketches (also with a signal basis e.g., in the *Acousmographe* software) and various graphical scores [35.49].
- Various body-motion-based shape renderings of sonic objects, including the mentioned *sound tracings* [35.33] and *air instrument performances* [35.32], as well as possibilities for extracting shape information from music-related video and motion capture material of musical performances and other music-related body motion [35.50].

The overall rationale here is that any feature may in principle be traced as a shape, hence making even the most ephemeral sensation *solidly* present for scrutiny and further differentiation. Morphodynamical theory has given us an extensive and well-reflected basis for shape cognition [35.9–11, 51]. In the words of *René Thom* [35.51, p. 6]:

> [...] *the first objective is to characterize a phenomenon as shape, as a spatial shape. To understand means first of all to geometrise.*

In sum, shape cognition offers a top-down approach to sonic object design in that we start with the overall shapes for loudness, pitch- and timbre-related features, something that we following Schaeffer may call the *typology* of sonic objects, then going on to subfeatures and sub-subfeatures of these shapes and into the more internal features, what we may call the *morphology* of the sonic object.

However, shape cognition and shape representation is a long-term project where we have substantial challenges of developing schemes for graphics, animations, feature choice, etc. But for now, whenever confronted with some feature(s) of sonic objects we want to study, we can start by drawing them subjectively, with pencil on paper or just mentally and then successively formalize such sketches into graphs and animations.

35.11 Typology and Morphology of Sonic Objects

A good point of departure for more systematic shape cognition of sonic objects could be the typological and morphological classification scheme of Schaeffer. The *typology* can be summarized as denoting the overall shape, or envelope, of any sonic object, with regards to its dynamic (loudness), and timbre- and pitch-related content. The *morphology* is on the other hand primarily concerned with that which goes on within these overall shapes. Schaeffer's typology and morphology scheme implements a kind of *form-content* distinction, with the advantage of being a general, flexible, top-down differentiation scheme applicable to any sonic object, regardless of origin.

The typology can be seen as a first sorting of sonic objects in that the perceptually most prominent features are taken into account. The typology starts with the three general dynamic envelope categories we have presented earlier and that correspond clearly with body motion categories:

- *Sustained*
- *Impulsive*
- *Iterative*.

This overall dynamic envelope classification is followed by a similarly coarse classification with regards to pitch:

- *Tonic stable*, meaning having a clearly perceivable pitch and that the pitch is relatively stable during the course of the sonic object.
- *Tonic varying*, meaning having a clearly perceivable pitch, but that this changes during the course of the sonic object, e.g., has an upward glissando.
- *Nontonic*, meaning inharmonic or noise-based sounds without any clearly perceivable pitch.

This results in a 3×3 typological matrix, a first and rather coarse but still useful classification of sonic objects. It should also be mentioned that prior to this typology, some other criteria of selection have been applied, i. e., first of all of length and of variability, criteria that are summarized in the notion of the *suitable object* [35.17].

Moving on to the more *internal* features of the sonic object, these are classified according to the principles of the morphology of Schaeffer's theory, here with two main dimensions of the morphology:

- *Gait*: Slower fluctuations in the sound, e.g., as made by a series of slower repeated tone onsets on an instrument, by slower opening–closing of mutes on brass instruments, by slow up and down glissandi on string instruments – in short, any motion in the sound slower than that which is typically perceived as tremolos or trills.
- *Grain*: Fast fluctuations in the sound.

All kinds of grain fluctuations within the sonic objects can be differentiated with our own terms:

- Tremolo, active shake, e.g., as on a violin with rapid back-and-forth bow motion or with *flatterzunge* on a brass instrument.
- Tremolo, passive response, e.g., washboard, maracas, or the grainy sound of a deep double bass or deep bassoon tone.
- Vibrato and/or trills, possible on several instruments.
- Spectral flux by active modulation made e.g., by rapid opening–closing of mutes on brass instruments.
- Spectral flux as passive response to excitation on various instruments, e.g., by scraping a tam-tam with a metal rod.

Note that different terminologies are possible and that the main point is thinking about objects holistically and qualifying their feature dimensions top-down, from the overall to the microlevel. This also goes for Schaeffer's concept of *mass*, i. e., spectral content of sonic objects, and this may be related to psychoacoustic elements, such as the following:

- Spectral shape of any kind, both quasistationary and changing in the course of the sonic object
- Spectral centroid, or the *focus* of the mass
- Correlating body posture shapes with spectral shapes, both of the vocal apparatus and other effectors.

There are a number of categories, subcategories, sub-subcategories, etc., here, but the important thing to keep in mind is that this is all in view of being a practical tool for sonic diagnosis, i. e., it should be considered a *questionnaire* and not a *balance sheet* [35.7, pp. 92–93].

35.12 Singular, Composed, Composite and Concatenated Objects

Although one of the key features of a sonic object is that it should be perceived and conceived holistically as a unit, it is also possible to consider variants of sonic objects not only along feature dimensions (as in the preceding section on the typology and morphology), but also by adding and/or expanding sonic objects. For this reason, it could be useful to make distinctions between singular and more complex sonic objects, distinctions that are in fact often encountered in orchestration and multitrack music production:

- Sonic objects may be *singular*, i. e., one basic and *simple* sound, e.g., of a tone from a piano, a flute, or a plucked string. As we know, such sonic events may in themselves contain several components, typically attack transients in the beginning of the sound, various fluctuations in the course of the sound and in the decay parts.
- Singular sonic objects may also be expanded by superposition, typically by adding other sonic objects to the attack point, resulting in what Schaeffer called *composed*, sonic objects. This is often found in both instrumental and electroacoustic composition, e.g., as in the example from Lutoslawski's *Jeux Venitiens* in Fig. 35.1a, where there is a loud percussive attack object added to the soft attack of the sustained string instrument chord.
- Sonic objects may be extended in the temporal direction, i. e., temporally smeared by adding prefixes and/or suffixes, what Schaeffer called *composite* sonic objects, frequently encountered as various ornaments and other figures in different kinds of music and also associated with coarticulation, i. e., the fusion of rapid tone events and sound-producing motion into higher-level sonic objects [35.41]. In Fig. 35.1b, we see an example of this from Messian's *Regard de la Vierge* where a rush of demisemiquavers and the final quavers fuse together to one single composite sonic object.
- Sonic objects may of course be *concatenated* to form longer stretches of music, as is evident in various kinds of collage-type composition, not only in electroacoustic music and disc jockey scratching music, but just as much in more traditional instrumental music (e.g., Schnitke, Maxwell-Davies, etc., or it may be more homogeneous, e.g., as in works by Messiaen, Stravinsky, etc.). This is what happens at what we above defined as the macro timescale and this is also an area that is not well researched in the field of music perception, but it would not be unreasonable to assume that there are new emergent contextual effects at the macro timescale not present at the meso and micro timescales.

In general, questions about what happens in long stretches of music are not addressed by classical sonic object research. From the cognitive science research referred to above, we could guess that attention is not constant, that it may tend to fluctuate, similar to gestalt flips when looking at the Necker cube, as suggested by *Ernst Pöppel* [35.36]. What does seem to be both quite clear and also feasible to document in sound- and body motion-based research, are the overall global features of longer stretches as done by MIR software (e.g., global indicators of loudness, harmonicity, roughness, spectral flux, etc.) and in body motion software (quantity of motion, mean square jerk, etc.), features that probably are quite good indicators of overall subjective sensations of effort and affect in musical experience.

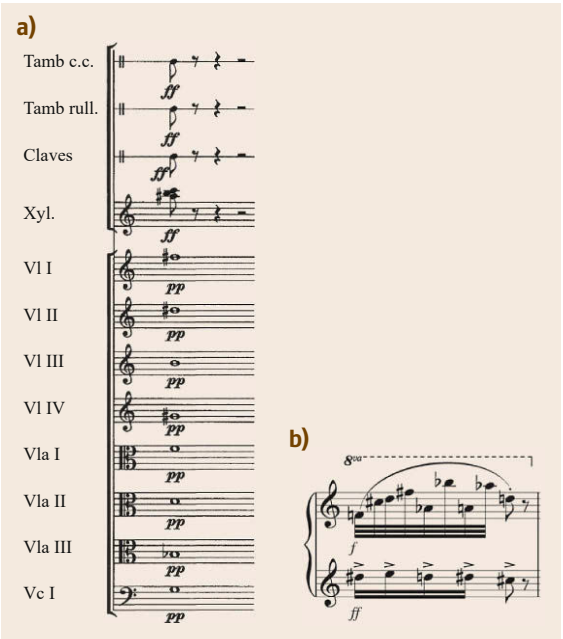


Fig. 35.1 **(a)** A *composed* sonic object from the first movement of Lutoslawski's *Jeux Venitiennes* consisting of a sustained and soft string chord with an added loud percussion attack. **(b)** A composite sonic object from Messian's *Regard de la Vierge* where there is a coarticulated prefix with a rush of demisemiquavers towards the final quavers (reproduced with permission from [35.31])

35.13 Textures, Hierarchies, Roles and Translations

The term *texture* can here be defined as the distribution of sound components within any sonic object. All sonic objects have internal textures, but notably, these textures may vary between totally stationary sounds (e.g., perfectly harmonic sounds) and complex, inharmonic, noise-dominated, or *chaotic* sounds, with most sonic objects found somewhere between these extremes. This is a wider definition of texture than usually encountered in more traditional music analysis, however, the traditional textural categories of homophony, polyphony, heterophony and monody, may be included in this definition as subcategories. This wide definition of texture, besides being in line with the morphological principles of Schaeffer, also reflects the theories and aesthetics of *Xenakis* of various statistical distributions of sound grains [35.52], and also reflects the textural focus in other late-20th century compositions e.g., by Ligeti and Lutoslawski.

The point with the wide definition of texture is to accommodate textural features of any sonic object and also to see that textures may sometimes be explicitly made (e.g., as in a tremolo on a piano) or may sometimes be emergent features of a singular sound, e.g., as in the case of the tremolo-like grain sound of a deep double bass tone. This view of texture is again very much in line with the acousmatic principle of Schaeffer's theory of disregarding the source and focusing on the perceived sonic object features.

Texture encompasses patterns in the sub-20 Hz region and may in speed range from very rapid tremolos and trills (i. e., close to the approximately 20 Hz continuous sound threshold) of *grain*, to the very slow, tentatively down to 30 BPM [35.53], and even slower of *gait*, ultimately also to completely stationary sounds.

In addition to the mentioned categories of grain and gait, there are various subcategories and dimensions, mainly specifying various kinds of spectral distributions and patterns of change. We could think of texture as situated in a multidimensional feature space with each dimension extending from some minimum to some maximum value, e.g., as in the following:

- Dense–spread
- Thick–thin
- Synchronous onsets–asynchronous onsets
- Short tones–long tones
- Many sustained tones–few or no sustained tones
- Wet–dry
- Little or no melodic movement–much melodic movement
- Small intervals in melodic lines–large intervals in melodic lines etc.

A division between foreground melody and background accompaniment is needless to say a very common texture in Western music. Added to this is the phenomenon of simultaneous different speeds in many musical textures, i. e., that the foreground is moving faster than the background, however, there are also deviations from this ordering. What is necessary then is to make an analysis of what could be called *role hierarchies* in sonic objects (and in longer musical excerpts as well). Such an analysis may typically reveal:

- Foreground, i. e., that which is the focus of attention at any moment.
- Background, i. e., that which is more in the periphery of attention.
- Role groups, meaning that complex sonic textures may often be composed of a smaller number of main roles (e.g., foreground + background), but that these main roles in turn are divided into subroles, as in the example in Fig. 35.2 below.
- In some cases, also totally fused, e.g., *Ligeti's Atmosphères* with the entire ensemble dedicated to one

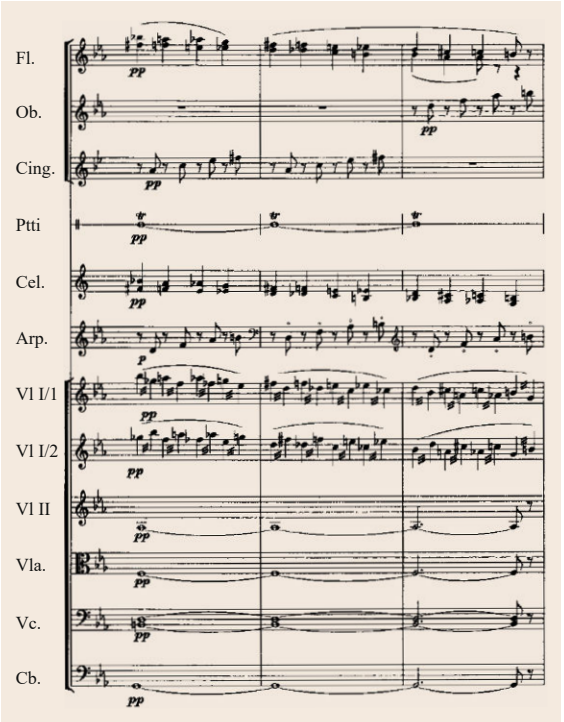


Fig. 35.2 An excerpt from Nikolai Rimsky-Korsakov's *Golden Cockerel Suite* containing different textural roles. See main text for details

singular sonic object, where the role of each instrument is more like that of being a partial in a complex spectrum.

Roles differentiated into subroles, sub-subroles, etc., are in cases of what may be considered successful orchestration, matched with suitable instrumental idioms. As an example, consider the fragment from the opening of Nikolai Rimsky-Korsakov's *Golden Cockerel Suite* in Fig. 35.2. Here we can observe a clear role organization:

- The foreground is made up by the English horn, transferred to the oboe in the third measure, doubled in the harp, the harp changing to harmonics in the second measure. This foreground may be regarded as a series of *impulsive* sounds and at that, *composite* with the attacks in the English horn doubled with the plucked attacks in the harp.
- Against this ascending foreground, we have a descending chromatic line in the flute, celesta and violin I and II, a line of basically sustained sounds that are in fact also a series of *composite* sonic objects with the sustained flute sounds as the core, with the celesta as attacks coloring and with the tremolos in the violins adding a *grain* texture to this.
- The *pp* in the strings enhances the white noise nature of string sound, making the *pp* tremolo on the cymbal, also with a soft white noise cloud, melt into this hushed background of the strings. Together, strings and cymbal clearly have the role of sustained sound, equivalent to adding a *wet* reverb signal to the sound.

In summary, we have in this excerpt three main roles and these main roles have in turn subroles, each of which matches the instrumental idioms well and, on the other hand, results in a rich and colorful orchestral sound as a whole.

To further study textural elements in music, we should also consider transfers of a musical idea/intention from one setting to another, as is done in orchestration and arrangements and which may be collectively called *musical translation*. But what is a musical idea and what is the gist of that which is translated? Besides a practical matter of making and evaluating different arrangements or orchestration versions of tunes or other musical compositions, translating sonic objects from one setting to another is also an encouragement to make us reflect on textural features and role functions in music, including on the role of idioms in music. As we know, it is usually more difficult to translate highly idiomatic expressions from one language to another and attempts at making a more literary translation, as close as possible word-by-word to the original, will often lead to quite strange, or even ridiculous, results. A translation that renders the more general meaning will usually be considered more useful as well as more true to the intentions of the original expression. Similar considerations apply in music, meaning that a task of translation could proceed as follows:

1. An analysis of textural elements
2. Assignments of textural elements to roles, including voice leading
3. Assignments of roles to instruments according to optimal match with idioms
4. Combinations of roles into textures, also taking optimal acoustic (spectral) distribution into account.

In sum, sonic object cognition may enhance our capacity to think analytically about orchestration or instrumentation from a combined perceptual-generative (including motor perspective) point of view, combining the best from the classics, e.g., the main principles of Rimskij-Korsakoff of idiom use, of spectral distribution, fusion, voice-leading, etc., with present schemes for sonic object feature analysis and synthesis, as well as more systematic research on similarity in music.

35.14 Analysis-by-Synthesis

Analysis-by-synthesis can be defined as the systematic exploration of features at all timescales, micro, meso, and macro, by producing and perceptually evaluating a series of sonic objects with incrementally different features. With a foundation in *J.-C. Risset*'s ideas [35.8], we may expand the idea of analysis-by-synthesis to include any activity where people are somehow engaged in producing a number of variant versions of a sonic object and evaluating these variant sonic objects in view of finding the most appropriate, suitable, pleasing, well-sounding, etc. sonic objects for the occasion or task at hand. This will then mean that most musical practice, rehearsal, studio production and composition work has some element of analysis-by-synthesis in that there is a trial-and-error, incremental tweaking of parameters going on. Analysis-by-synthesis is then de facto a familiar phenomenon in musical practice, however, this may have eluded much of music theory and analysis.

The strategy of analysis-by-synthesis applied to sonic objects is then a process with the following steps:

1. Determining some feature dimension(s)
2. Incrementally varying the value(s) of this feature's dimensions
3. Evaluating the subjective perceptual results of these variations.

It may turn out that incremental value changes are perceived as just that, as a gradual change in some hue, but it may also turn out that an incremental change suddenly results in a very distinct qualitative change. Examples of this are encountered in sound synthesis, e.g., the bowed-plucked transition with a gradual decrease in the attack time duration, or a significant change in timbre with FM synthesis.

In summary, the working strategy here is to explore categories, categorical thresholds, intercategorical and intracategorical variation of sonic objects through such analysis-by-synthesis, thus progressively building up our knowledge of sonic objects.

35.15 Summary

The focus on sonic objects, both in the works of Pierre Schaeffer and various later research [35.46, 54], is motivated by the belief that the sonic object is a highly significant element of musical experience, that it has a privileged status in our understanding of music and as a tool for music creation as well. The point is that in spite of significant advances in musical acoustics, psychoacoustics, music cognition, music technology and music information retrieval during the last couple of decades, there are still many unnamed salient features in musical sound and we definitely still need the top-down approach as was presented by Schaeffer et al. half a century ago. In other words, the concept of *object* has great potential as the core of now fast growing insights from a number of music-related research areas.

Needless to say, there are serious challenges of representation, however there is work going on that progressively gives us better signal-based representations, i. e., more selective and focused visual representations that enable in-depth, systematic explorations of sonic object features as shapes. In particular, various graphical and animation representations of combined sonic and body motion features are promising in offering us insights into how rich aesthetic sensations emerge from the ephemeral.

References

35.1 E. Husserl: *On the Phenomenology of the Consciousness of Internal Time, 1893–1917* (Kluwer, Doredrecht 1991), English translation by John Barnett Brough

35.2 R.I. Godøy: Thinking now-points in music-related movement. In: *Concepts, Experiments and Fieldwork. Studies in Systematic Musicology and Ethnomusicology*, ed. by R. Bader, C. Neuhaus, U. Morgenstern (Peter Lang, Frankfurt 2010) pp. 245–260

35.3 R.I. Godøy: Sound-action awareness in music. In: *Music and Consciousness*, ed. by D. Clarke, E. Clarke (Oxford Univ. Press, Oxford 2011) pp. 231–243

35.4 T.D. Griffiths, J.D. Warren: What is an auditory object?, Nat. Rev. Neurosci. **5**(11), 887–892 (2004)

35.5 P. Schaeffer: *Traité des Objets Musicaux* (Éditions du Seuil, Paris 1966)

35.6 P. Schaeffer: *Solfège de l'objet Sonore* (INA/GRM, Paris 1998), with sound examples by Reibel, G. and Ferreyra, B. 1967

35.7 M. Chion: *Guide des Objets Sonores* (INA/GRM Buchet/Chastel, Paris 1983)

35.8 J.-C. Risset: Timbre analysis by synthesis: Representations, imitations and variants for musical composition. In: *Representations of Musical Signals*, ed. by G. De Poli, A. Piccialli, C. Roads (MIT Press, Cambridge 1991) pp. 7–43

35.9 R.I. Godøy: *Formalization and Epistemology* (Scandinavian Univ. Press, Oslo 1997)

35.10 J. Petitot: *Morphogenèse du Sens I* (Presses Universitaires de France, Paris 1985)

35.11 J. Petitot: *Forme in Encyclopædia Universalis* (Encyclopædia Universalis, Paris 1990)

35.12 A.R. Jensenius, R.I. Godøy: Sonifying the shape of human body motion using motiongrams, Empir. Musicol. Rev. **8**(2), 73–83 (2013)

35.13 A. Hunt, M. Wanderley, M. Paradis: The importance of parameter mapping in electronic instrument design, J. New Music Res. **32**(4), 429–440 (2003)

35.14 A. Bregman: *Auditory Scene Analysis* (MIT Press, Cambridge 1990)

35.15 J.K. Bizley, Y.E. Cohen: The what, where and how of auditory-object perception, Nat. Rev. Neurosci. **14**, 693–707 (2013)

35.16 I. Winkler, T.L. van Zuijen, E. Sussman, J. Horváth, R. Näätänen: Object representation in the human auditory system, Eur. J. Neurosci. **24**(2), 625–634

(2006)
35.17 R.I. Godøy: Gestural-sonorous objects: Embodied extensions of Schaeffer's conceptual apparatus, Organ. Sound **11**(2), 149–157 (2006)
35.18 R.I. Godøy, M. Leman: *Musical Gestures: Sound, Movement and Meaning* (Routledge, New York 2010)
35.19 V. Gallese, T. Metzinger: Motor ontology: The representational reality of goals, actions and selves, Philos. Psychol. **16**(3), 365–388 (2003)
35.20 M. Wilson, G. Knoblich: The case for motor involvement in perceiving conspecifics, Psychol. Bull. **131**(3), 460–473 (2005)
35.21 V. Gallese, G. Lakoff: The brain's concepts: The role of the sensory-motor system in conceptual knowledge, Cogn. Neuropsychol. **22**(3/4), 455–479 (2005)
35.22 A. Berthoz: *Le Sense du Mouvement* (Odile Jacob, Paris 1997)
35.23 A.M. Liberman, I.G. Mattingly: The motor theory of speech perception revised, Cognition **21**, 1–36 (1985)
35.24 B. Galantucci, C.A. Fowler, M.T. Turvey: The motor theory of speech perception reviewed, Psychon. Bull. Rev. **13**(3), 361–377 (2006)
35.25 J. Haueisen, T.R. Knösche: Involuntary motor activity in pianists evoked by music perception, J. Cogn. Neurosci. **13**(6), 786–792 (2001)
35.26 E. Kohler, C. Keysers, M.A. Umiltà, L. Fogassi, V. Gallese, G. Rizzolatti: Hearing sounds, understanding actions: Action representation in mirror neurons, Science **297**, 846–848 (2002)
35.27 M. Bangert, E.O. Altenmüller: Mapping perception to action in piano practice: A longitudinal DC-EEG study, BMC Neuroscience **4**, 26 (2003)
35.28 H. McGurk, J. MacDonald: Hearing lips and seeing voices, Nature **264**, 746–748 (1976)
35.29 R.I. Godøy: Imagined action, excitation and resonance. In: *Musical Imagery*, ed. by R.I. Godøy, H. Jorgensen (Swets and Zeitlinger, Lisse 2001) pp. 237–250
35.30 R.I. Godøy: Motor-mimetic music cognition, Leonardo **36**(4), 317–319 (2003)
35.31 R.I. Godøy: Images of sonic objects, Organ. Sound **15**(1), 54–62 (2010)
35.32 R.I. Godøy, E. Haga, A.R. Jensenius: Playing air instruments: Mimicry of sound-producing gestures by novices and experts. In: *GW2005, LNAI 3881*, ed. by S. Gibet, N. Courty, J.-F. Kamp (Springer, Berlin, Heidelberg 2006) pp. 256–267
35.33 K. Nymoen, R.I. Godøy, A.R. Jensenius, J. Torresen: Analyzing correspondence between sound objects and body motion, ACM Trans. Appl. Percept. **10**(2), 9 (2013)
35.34 B.C.J. Moore: *Hearing* (Academic, San Diego 1995)
35.35 R. Gjerdingen, D. Perrott: Scanning the dial: The rapid recognition of music genres, J. New Music Res. **37**(2), 93–100 (2008)
35.36 E. Pöppel: A hierarchical model of time perception, Trends Cogn. Sci. **1**(2), 56–61 (1997)
35.37 R.I. Godøy: Reflections on chunking in music. In: *Systematic and Comparative Musicology: Concepts, Methods, Findings*, ed. by A. Schneider (Peter Lang, Frankfurt 2008) pp. 117–132
35.38 N. Hogan, D. Sternad: On rhythmic and discrete movements: Reflections, definitions and implications for motor control, Exp. Brain Res. **181**, 13–30 (2007)
35.39 S.T. Klapp, R.J. Jagacinski: Gestalt principles in the control of motor action, Psychol. Bull. **137**(3), 443–462 (2011)
35.40 H. Haken, J. Kelso, H. Bunz: A theoretical model of phase transitions in human hand movements, Biol. Cybern. **51**(5), 347–356 (1985)
35.41 R.I. Godøy: Understanding coarticulation in musical experience. In: *In: Sound, Music and Motion Lecture Notes in Computer Science*, ed. by M. Aramaki, M. Derrien, R. Kronland-Martinet, S. Ystad (Springer, Berlin 2014) pp. 535–547
35.42 S.T. Grafton, A.F. Hamilton: Evidence for a distributed hierarchy of action representation in the brain, Hum. Mov. Sci. **26**, 590–616 (2007)
35.43 R.I. Godøy: Quantal elements in musical experience. In: *Sound–Perception–Performance. Current Research in Systematic Musicology*, ed. by R. Bader (Springer, Berlin, Heidelberg 2013) pp. 113–128
35.44 S.T. Klapp, J.M. Nelson, R.J. Jagacinski: Can people tap concurrent bimanual rhythms independently?, J. Motor Behav. **30**(4), 301–322 (1998)
35.45 R.I. Godøy: Gestural affordances of musical sound. In: *Musical Gestures: Sound, Movement and Meaning*, ed. by R.I. Godøy, M. Leman (Routledge, New York 2010) pp. 103–125
35.46 D. Rocchesso, F. Fontana: *The Sounding Object* (Edizioni di Mondo Estremo, Firenze 2003)
35.47 U. Zölzer: *DAFX: Digital Audio Effects* (Wiley, Chichester 2011)
35.48 R. Cogan: *New Images of Musical Sound* (Harvard Univ. Press, Cambridge 1984)
35.49 B. Schäffer: *Introduction to Composition* (PWM, Warsaw 1976)
35.50 A.R. Jensenius: Some video abstraction techniques for displaying body movement in analysis and performance, Leonardo: J. Int. Soc. Arts Sci. Technol. **46**(1), 53–60 (2013)
35.51 R. Thom: *Paraboles et Catastrophes* (Flammarion, Paris 1983)
35.52 I. Xenakis: *Formalized Music* (Pendragon, Stuyvesant 1992)
35.53 L. Van Noorden: The functional role and bio-kinetics of basic and expressive gestures in activation and sonification. In: *Musical Gestures: Sound, Movement and Meaning*, ed. by R.I. Godøy, M. Leman (Routledge, New York 2010) pp. 154–179
35.54 F. Delalande, M. Formosa, M. Frémiot, P. Gobin, P. Malbosc, J. Mandelbrojt, E. Pedler: *Les Unités Sémiotiques Temporelles: Éléments Nouveaux d'analyse Musicale* (Marseille, Édition MIM-Documents Musurgia 1996)

36. Investigating Embodied Music Cognition for Health and Well-Being

Micheline Lesaffre

The aim of this chapter is to highlight challenges involved in the successful deployment of the rather young paradigm of embodied music cognition in the comprehensive domain of health and well-being. Both our current society and systematic musicology are experiencing transitions that have given rise to cross-disciplinary research collaboration between researchers in musicology, the sciences, and a variety of stakeholders in health and well-being. It has been shown that the interdisciplinary, empirical approach that typifies embodied music cognition research has the potential to bring new perspectives to therapeutic approaches for well-being. However, to bring this potential to fruition, researchers have to face the many challenges that arise from the difficulties of using new methods and technologies, especially when working in unfamiliar domains and contexts. In this chapter a framework is presented that provides support to the prominent question of how know-how from the paradigm of embodied music cognition can be efficiently transferred to the sectors of health, rehabilitation, and well-being.

36.1 Transitions in Musicology and Society

There are several transitions occurring both in systematic musicology and in society that provide a challenging basis for *cross-disciplinary* research studies between musicology, the sciences, and practitioners in healthcare. Here cross-disciplinary research refers to cooperation of a group of persons, academics and non-academics, trained in disciplines with different concepts, methods and data, organized into a common effort on a problem. Such teamwork is fundamental to the achievement of investigating goals, but to make sense of the issues of importance one has to face the strong challenges of dealing with the diversity of theoretical and practical perspectives of each partner in a team.

The link between music and health is a perennial knowledge that has been understood and practiced in varied ways across history. However, it was only recently that interdisciplinary scholarship began to research and theorize the richness of music, health and well-being, and made attempts to bridge gaps between scholarly and professional territories [36.1–5]. The renewed attention directed at using the power of music for health and well-being is associated with one of the major transitions that research in systematic musicology is currently experiencing. Because of the development of the *embodied music cognition* research paradigm [36.6], the way music is understood is changing completely, in the sense that it is replac-

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_36

ing the traditional theory for music that focused more on the mind rather than on the body. This approach offers a challenging framework, for example for thinking and setting up experiments on the mechanisms behind corporeal musical interaction and gestural communication [36.7]. Therefore, among scholars in systematic musicology, there is a shift of attention from the mind as a receiver of auditory stimuli to the human body and to the perception–action loops that are mediated by the body.

Furthermore, in our Western culture there are several societal issues, where novel paradigms in music research could make a difference. Typical problems that require attention are for instance the continuous growth in numbers of overweight persons and in young people with anxiety disorders. But, most striking is the ever-increasing phenomenon of an ageing population, which is affecting most countries around the world. Population aging has profound implications for many facets of human life that go beyond the elderly. The global share of older people (aged 60 years or over) increased from 9.2% in 1990 to 11.7% in 2013 and will continue to grow as a proportion of the world population, reaching 21.1% by 2050 [36.8]. Because chronic and neurodegenerative diseases are more common at older ages, there is an augmented prevalence of disorders such as memory loss and depression. Increasing life expectancy therefore raises the question of whether longer life spans result in more years of life in good health, or whether it is associated with increased morbidity and more years spent in prolonged disability and dependency.

Taking the above-mentioned transitions in musicology and society as a starting point, and given the extent of literature on the use of music in everyday life, in this chapter musical interaction is rather conceptualized as a therapeutic intervention as opposed to an everyday activity. Even within this constraint, the cross-disciplinary embodied music cognition approach has given rise to such a broad range of concepts that may be associated with music interventions for health and well-being, that it is hardly possible to define a unique framework. Throughout this chapter examples will be restricted to review papers on music therapeutic interventions and to specific studies that use music and movement to deal with problems of neurodegenerative diseases and physical or mental impairment. The emphasis will be on individual rather than social approaches to health and well-being.

Using music for therapeutic goals dates back a long time. In *Horden* [36.9] and *Bunt* and *Stige* [36.10] the history of music therapy is discussed from different perspectives, including ecological and anthropological viewpoints. *Antonietti* [36.11] provides an overview of ways of employing music for therapeutic purposes in *rehabilitation*. Here the field of *music therapy* is approached from a psychological viewpoint, as a tool of the mind, using cognition and emotion as the avenue towards accomplishing goals for rehabilitation. The advantage of using music in therapeutic programs for the elderly and beyond is that it has the benefit of being painless, nonintrusive, easily accessible, and cost-effective. For the elderly, music therapy may be an enjoyable means for the maintenance and improvement of cognitive, physical, and social functioning. In that perspective *Hays* [36.12] examines the importance of music in facilitating well-being for older people who have special needs and discusses how music can contribute to their *quality of life*. *Clair* and *Memmott* [36.13] detail the benefits of music therapy for aging populations. Their work addresses caregivers who use music to enhance the quality of life of older adults. An important remark that has to be made is that, although there are very many initiatives on the topic of *music for the elderly*, especially around musical activities with persons with *neurodegenerative diseases*, they are usually not supported by *evidence-based* theories.

Today, the challenges of using the power of music for the benefit of people's health and well-being are becoming the focus of extensive scientific discussions in disciplines beyond music therapy. A cross-disciplinary scientific approach to understanding the fundamentals of the benefits people can get from music has developed only recently. Unfortunately, there is practically no literature connecting music therapy with the fundamentals of *embodied music cognition*. Because of its emphasis on the tight relationship between perception and action, the young paradigm of musical embodiment offers a novel tool for the exploration of specific forms of *nonverbal communication*. Through physical behavior in response to music a person may for example reveal clues to unspoken intention or emotion. This is particularly the case for target populations, such as people with *dementia*, who tend to put bodily sensations immediately into action. One important and challenging characteristic shared by researchers who intend to apply embodied music cognition theory, is that embodiment in music experience and behavior is foremost investigated on an *empirical* basis using advanced technologies and tools to measure the effects that music can have on the human body. As a consequence, the evolution in systematic musicology toward the embodied music cognition approach goes along with two strongly connected shifts. Firstly, the focus of interest has moved toward the role that *tools and technologies* can have as *mediators* between experience and environment. Secondly, *user-centered research* involving a broad range of stakeholders in the research process has become an important topic in music research.

To summarize, the societal and musicological transitions mentioned above suggest that bringing the fundamental concepts of musical embodiment into practice, for the benefit of health and well-being, is not straightforward because:

- Challenges of cross-disciplinary research in the field have not been identified yet.
- The impact of new tools and technologies on research methods and strategies still remains underexplored.
- Involvement of a broad range of stakeholders throughout the research process is still not common in music and movement research.
- Current programs using music and movement are not sufficiently supported by evidence-based theories and practice.

So, the question is not to consider *whether* to engage with scientifically validated music applications for health and well-being, but *how* to engage with them.

This chapter has two major parts. The first part provides an understanding of how the field of health and well-being is defined and how the determinants thereof may be connected to current models in music research. The second part deals with the underpinnings of the embodied music cognition framework. It is subdivided into one section considering the theoretical perspectives of musical embodiment and another one discussing the use of tools and technology and user-related strategies involved in current empirical music research. The aim is to provide a framework that supports the deployment of the embodied music cognition paradigm in the field of health and well-being.

36.2 Models of Music, Health and Well-Being

Before exploiting the conceptual framework of embodied music cognition for the benefit of *health* and *well-being*, it is important first to have an understanding of how the field is defined and second what the existing models of music, health and well-being are.

36.2.1 Dimensions of Health and Well-Being

The most commonly accepted and straightforward definition of health was set out in the preamble to the Constitution of the *World Health Organization* (WHO) in 1946. WHO defines health as *a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity* [36.14]. These three dimensions of health are defined as follows:

- *Physical well-being* is concerned with the efficient functioning of the organs and the system.
- *Mental well-being* is the ability to cope successfully with the normal problems and stresses of life.
- *Social well-being* relates to people's roles within society as a whole.

However, the WHO definition has been subject to criticism and several other definitions have been put forward. For an overview see, for example, *Üstün* and *Jakob* [36.15]. And, although the WHO moved the focus beyond individual physical abilities or dysfunction, it is still often ignored in favor of a biomechanical focus on the physical health of individuals only. Over the last decades, the acquired knowledge that mind and body are not as separate as previously thought gave birth to new conceptualizations, such as the *psychobiological* model [36.16] which explores interactions between social, psychological and biological factors, and the *holistic* model [36.17] which includes a fourth dimension, the spiritual. According to *Chatterji* et al. [36.18], who approach the subject from a medical perspective, health must be understood as being inextricably tied to states of the human body and mind and as such distinct from extrinsic environmental features. In this discussion paper, health is not assigned to environments or behaviors, but to the health states or conditions that environments or behaviors produce in individuals. Interestingly, Chatterji et al. elaborate a conceptual framework in which *health states* are described in terms of levels of dimensions such as mobility, pain, hearing, and seeing.

But, how does health relate to well-being and vice versa? Health has been recognized as an important element of well-being because physical, mental, and social health have a direct impact on quality of life (QoL), especially for older people. Therefore, in the majority of models well-being is regarded as a concept related to, but separate from, the concept of QoL. Moreover, the term QoL is used differently in health literature than in sociological literature. In health literature, many QoL instruments have been developed. These generally contain items that relate to symptoms of impairment, functional status, and emotional states and affect. In sociological literature, quality of life refers to the feeling of how good, desirable, and enjoyable life as a whole is. From that perspective well-being

has been viewed variously as happiness, satisfaction, enjoyment, contentment, engagement, and fulfillment or a combination of these and other, hedonic and eudaimonic factors. *Kahneman* et al. [36.19] propose five conceptual levels as relevant for research on well-being. These are:

1. External (objective) conditions (e.g., income, neighborhood, housing)
2. Subjective well-being (e.g., self-reports of satisfaction/dissatisfaction)
3. Persistent mood level (e.g., optimism/pessimism)
4. Immediate pleasures/pains, transient emotional states (e.g., joy, anger)
5. Biochemical, neural bases of behavior.

Hjelm [36.20] distinguishes between and discusses the characteristics of five dimensions of health:

1. Physical
2. Social
3. Emotional
4. Intellectual
5. Spiritual.

Cutler and *Landrum* [36.21] characterize the multifaceted health of the elderly and aim at understanding how health along multiple dimensions has changed over time. They combine 19 measures of health into 3 broad categories representing:

1. Severe physical and social incapacity
2. Less severe difficulty
3. Vision and hearing impairment.

Notwithstanding the variety of approaches, it is possible to observe a general agreement concerning the main conditions shaping health and well-being. Researchers differentiate between well-being that incorporates objective measurable conditions and subjective well-being, which is well-being as defined, or assessed, by individuals themselves and which may include subjective response to objective conditions. The distinction between *objective* and *subjective well-being* is conceptual as well as methodological. External conditions can for example be assessed by self-report (subjective) as well as by independent (objective) observation.

The intertwined concepts of health and well-being are widely used in very many disciplines, but with little consistency. One important reason, that is not so often mentioned, why health and well-being research cannot easily be classified consistently is that its *dimensions* impact one another. For example, people with physical disability may also have problems with respect to mental health, such as feeling depressed or being anxious, or may be excluded from social life. The connectedness between the dimensions of health and well-being provides an imperative challenge for researchers who often restrict their studies to one single dimension without examining its relationship with the others. However, a person's health and well-being results from a complex interplay of the multiple dimensions of health. The knowledge that these dimensions are not unique and differ between target groups makes it even more complex.

36.2.2 Linking Music and Health and Well-Being

In recent years, there has been considerable interest in exploring and understanding the positive effects that music can have on health and well-being in diverse populations and settings. The majority of studies that connect music with health and well-being investigate indicators of social and emotional factors contributing to the overall quality of life of individuals. This approach is supported by the fact that healthcare services are evolving from a disease and symptom model into one of prevention and wellness that emphasizes quality of life and lifestyle in addition to pharmacological treatment. Music therapy is becoming more and more part of lifestyle-enhancing programmes that promote strategies for coping with stress and increasing pleasure in life. From that perspective, *Hanser* [36.22] gives a comprehensive overview of music therapy approaches that document how music experiences can promote and facilitate health and well-being. The author reports on the effects of music on stress, pain, immune and neurological functions, and devotes special attention to conditions of childbirth, depression, coronary heart disease, and cancer.

In the last decade, serious efforts have been made to map the practical and theoretical field of music and health and to determine the relationship between music and health and well-being. Authors have analyzed the field from different approaches, research contexts and methods, for instance health musicking, music in everyday life, and qualitative research. The availability of distinct models that have different types of interventions associated with them shows the complexity of the field. The following examples of studies provide an account of this difficulty. *Bonde* [36.23] starts from the concept of *health musicking*, which is any form of participation in musical activities and the influence thereof [36.24]. The notion of health musicking draws attention to human action and the performance of relationships. Bonde presents a theoretical quadrant model showing how health musicking relates to four major purposes. Here health musicking is understood as the common core of any use of music experiences to regulate emotional or relational states, or to promote

well-being, be it therapeutic or not, professionally assisted or self-made. The four goals of this model are:

1. The development of communities and values through musicking
2. The shaping and sharing of musical environments
3. Professional use of music(ing) and sound(ing) to help individuals
4. Formation and development of identity through musicking.

In Bonde's model health musicking is described as an interdisciplinary field of theory and practice including specific disciplines like music therapy, music medicine, community music (therapy), and professionals with many different backgrounds and qualifications.

In her book on well-being through music in everyday life, *DeNora* [36.25] develops an interdisciplinary framework for music, health and well-being. Considering health and illness both in medical contexts and in the realm of everyday life, DeNora argues that these identities are by no means mutually exclusive. It is suggested that the promotion of health and more specifically, mental health, involves a great deal more than a concern with medication, genetic predispositions, clinical and neuroscientific procedures. Interestingly, in DeNora's approach music is not adjunct to medicine but an equal partner in the performance of well-being.

In a special issue on music, health and well-being, *MacDonald* [36.4] maps the current field of research, introducing five areas that encompass a perspective on music and health. They are the following:

1. Music therapy
2. Community music
3. Music education
4. Everyday uses of music
5. Music medicine.

MacDonald places the emphasis on the benefits of *qualitative research methods*, because this approach facilitates the exploration of the subjective and phenomenological aspects of musical experience. Indeed, musicological research often goes together with a growth of qualitative research performance within healthcare contexts. However, besides relying on qualitative research methods, unraveling the relationship between music health and well-being also depends on the study of quantitatively measurable processes.

Lesaffre [36.5] presents a model as a *contextual* framework conceived from the viewpoint of music and movement applications for health and well-being. In that model, the three basic dimensions of health (i. e., physical, mental, and social well-being) are connected to contexts that can shape health. Activities in the domain of physical and mental well-being encompass musical interventions for the body and the mind. They mainly pertain to what is known as music therapy, movement therapy, and dance therapy. Social well-being is related to lifestyle and quality of life and more specifically to the extent to which music and movement activities can contribute to this.

A review of the literature has shown that studies that provide a framework for music, health and well-being have so far been short on accounting for the mechanisms that underlie the power of music to support health and well-being. Notwithstanding, scientific evidence of strong connections between mind and body and the study of action–perception coupling processes have opened new perspectives for musical therapeutic interventions. Supported by innovative theories in cognitive sciences and musicology, such as the embodied music cognition paradigm, new ways of exploring the extraordinary relationship between man and music are being uncovered. Especially, as more sophisticated technology facilitates the understanding of human functioning in both the body and the brain, there are more means to support the great impact of music.

36.3 From Theory to Therapeutic Approaches

A future key challenge of implementing the embodied music cognition paradigm will be to demonstrate what unique contributions its framework can make to the domain of health and well-being. This challenge can only be met if a broad range of *stakeholders* are involved through all stages of the process of developing evidence-based foundations for future programs and interventions. Stakeholders can be individuals (including scientists, therapists, developers, testers, sponsors, patients, and anyone impacted by a project) or organizations (e.g., caregiving centers, hospitals, sports clubs) that have an interest in a particular project and ideally are involved in the project from the very beginning. However, including a variety of stakeholders introduces music researchers to unfamiliar research approaches pertaining to unfamiliar disciplines. Therefore, it is essential to integrate the theoretical perspectives of embodied music cognition into the empirical framework in a way that enables stakeholders to make sense of the complexities of concepts and methods applied. An

additional challenge is the development and testing of technologies, methodologies, and associated tools in accordance with user approaches that are rather novel to the domain of music research. The concepts and ideas pertaining to these issues are discussed in the next subsections.

36.3.1 Theoretical Perspectives of Musical Embodiment

A significant feature of the embodied cognition research approach is that it views the body as being central to the way the world is experienced. This means that it is assumed that cognitive processes are ruled by sensorimotor capacities and their interactions with the environment. Because embodied cognition already is the product of various research fields including psychology, movement science and neuroscience, *Shapiro* [36.26] refers to it as a research program rather than a theory. *Gallagher* [36.27] provides an overview of the variety of theorists that have attempted to outline different approaches and meanings related to the concept of embodied cognition. Although there are various approaches to embodied cognition, its central claim can be summarized in the idea that there is a representational equivalence between perception and action [36.28].

With his theory of embodied music cognition, *Leman* [36.6] introduces the concept of embodiment in systematic musicology. He argues for the understanding of the human body as the most natural mediator between musical subjects and their musical environment. Leman's framework distinguishes between four fundamental components of musical embodiment, namely:

1. The body as mediator
2. The action-oriented ontology
3. Action–perception coupling
4. Embodied musical communication or expression.

These components provide the conceptual scaffolding for cross-disciplinary research between embodied music cognition, the sciences, and health and well-being. In the next paragraphs a brief description of the meaning of these fundamentals is given, together with references to studies and applications in the domain of music, health and well-being. For a more extensive explanation of these fundamentals, see *Leman* [36.6, 29, 30].

In embodied music cognition the idea of the *body as mediator* implies that the human body is conceived as a mediator intervening between a person's physical environment (i. e., the music) and a person's subjective experience of that environment. The physical environment can be described in an objective way (e.g., the waveform), whereas experience can only be described in a subjective way. Similarly, musical gestures can be described in an objective way as movement of body parts, but they have an important experiential component that is related to intentions, goals, and expressions. Starting from the concept of the body as mediator, *De Bruyn* [36.31] has, for example, developed an empirical methodology that uses motor tasks to assess music perception skills. A music application was developed that provides therapists with a methodology for music perception training, for example in auditory rehabilitation of persons with a cochlear implant. The application called *Sound Caterpillar* is conceived as a musical game that uses motor tasks to assess music perception skills.

The second component, the *action-oriented ontology* or action–gesture repertoire, consists of a set of gestures and related expressions that a person has built up over years. The effect music might have on a person is strongly dependent on the experiences that people have accumulated during their gestural interactions with music. As corporeal articulations and actions are carried out in space and time, it seems natural to conceive of the action repertoire as a container of spatial–temporal patterns. In current studies these *spatial–temporal* patterns have been examined through aspects of gesture [36.7] and *entrainment* [36.30, 32]. Entrainment is the process of synchronizing endogenous sensations of beat with an external rhythm of movement. These studies open new perspectives for understanding the relationship between entrainment and expressiveness. In a review paper, *Nombela* et al. [36.33] address the underlying mechanisms of musical entrainment and its effect on improvements in gait and motor improvements in people with Parkinson's disease. Especially neuroimaging studies provide valuable insights for developing potential movement therapies.

The *action–perception coupling system* is a complex mechanism that controls the interaction between environment and subjective experience. It is responsible for predicting the outcome of corporeal actions and musical intentions. Understanding the components and dynamics of the action–perception coupling system holds promising potential for motor rehabilitation. *Wolpert* et al. [36.34] review recent research in human motor learning with an emphasis on the computational mechanisms that are involved. In *Maes* et al. [36.35] the action–perception coupling mechanism is approached from the embodied music cognition perspective. Maes presents a theoretical framework that captures the ways in which the human motor system and its actions can reciprocally influence the perception of music. Such a framework offers a foundation for research that investigates the effect of musical activities, for example in people with motor disorders.

The fourth component, *embodied musical communication*, is based on the processes of encoding and decoding *expressiveness*. Traditionally, body movements have not been considered a parameter of *musical expression*. However, the work of *Davidson* [36.36] has shown, that body movement is in fact one of the most salient features when communicating expressive intentions. The idea that music may establish an intentional layer of communication between listener and performer, links the component *musical communication* to the concept of *musical intentionality*. Musical intentionality is based on an action–perception coupling system that uses actions as causes of perceived patterns. The system turns perceived sound patterns into action patterns that could have caused the perceived sounds. *Davidson* and *Emberly* [36.37], for example, provide a cross-cultural analysis that allows for the examination of distinct notions of embodied musical communication whilst building on developments that support the idea of musicality and its role in enhancing quality of life and feelings of well-being.

36.3.2 Tools, Technologies, and Their Users

Over the last decade, the omnipresence of music interaction tools has formed the basis for a shift in systematic musicology toward *technology-driven research*. In embodied music cognition, such research is of key importance, because it makes it possible to complement the subjective approach with objective studies of the phenomenon of musical empowerment, for example, by investigating body movement in music-related activities or studying the characteristics of the action–perception coupling system. This means that researchers in musicology not only enter the rather unfamiliar domain of tool use and tool design, they also have to take into account the specificities and needs of the various groups and subgroups in the domain of health and well-being they are targeting. The next paragraphs consider the challenges of working with novel *measurement tools*, *technologies*, and *user-related strategies* involved in current empirical music research.

Tools for Measurement and Analysis

Many aspects of music's role in relation to health and illness have been subject to measurement. Apart from *qualitative methods* such as interviews and focus groups, a range of *quantitative measurement* instruments in the form of *self-assessment* tools or *observer-rated* tools and clinical tests is typically employed. Prominent examples are the mini-mental state examination MMSE [36.38] for testing cognitive impairment and quality of life (QoL) surveys, such as the widely used short form health questionnaire SF-36 [36.39] for measuring self-reported physical and mental health status and the QUALIDEM scale [36.40] that was designed for persons with dementia.

These and many other measurement instruments have been subject to critique, for example concerning their *ecological validity* [36.25]. Indeed, the traditional situation of testing often involves responding to a survey guided by a trained assessor and is limited to a lab setting and thus outside the context of everyday life. Therefore, critical thinking about tool use is required in order to produce ecologically valid accounts of how music may promote health and well-being. Furthermore, such surveys offer retrospective perspectives and are therefore subjected to distortion. Moreover, in humanities, too many researchers studying the response to music solely rely on self-report and questionnaires as evidence and in some cases questions are used that were not validated. In any case, such information does not capture the complexity of feelings, intentions, and thoughts a person may have. Also, the accompanying concepts do not have an unambiguous meaning because most behaviors are the result of more than one condition. Therefore, this subjective *top-down* approach needs to be supported by additional objective *bottom-up* information like measurements of parameters related to body movement or brain activity.

Embodied music cognition research particularly concentrates on the empirical use of a variety of tools for measurement and on tool development. But, working with a combined set of tools and technologies produces *high-dimensional data* which require advanced skills to handle. It is for example very likely that apart from using surveys, an experimental design using optical motion capture will include measurements for several subjects of several body parts (e.g., hands, legs, heads, trunk). In addition it may include measurement of other information indirectly related to movement such as biometric data (e.g., heart rate, inhalation and expiration volume, and cerebral blood flow measured by techniques like fMRI (functional magnetic resonance imaging)). All these measurements result in huge multivariate datasets causing challenges not only for *data collection* and *storage* but also for methods of analysis and visualization. The *integrated analysis* of top-down and bottom-up measurements of human physical and/or mental movements in a musical context is a real challenge in cross-disciplinary science today. *Multimodal modeling* is an important goal in order to provide a combined representation and analysis of the collected data. In a study on monitoring and analyzing the effect of live music performances on people with dementia, *Lesaffre* et al. [36.41] combine the use of various questionnaires, the MMSE test, measurement of quantity of movement, and audio and video record-

ing of participants and performers. For this multimodal dataset, an analysis tool was developed that combines the bottom-up and top-down approach for comparative analysis. In this way it was possible to show that live music invites people with dementia to move more to the music than listening to music from loudspeakers and that there is a significant effect of the degree of cognitive impairment on their motor response.

Technology-Driven Research for Music, Health and Well-Being

Technology-driven research groups aim at developing health and well-being by means of technological applications rather than care itself. In many cases the main activities are service design and follow-up and evaluation methodology development. In embodied music cognition research the technology-driven approach is rather young. Its emphasis is more on using technologies to investigate the fundamentals of action–perception processes and on developing tools for monitoring or retraining purposes, for example of people who have had a stroke.

The elaboration of tools and technologies has introduced the concept of *mediation technology* in systematic musicology. It is based on the idea that technology mediates human perception and action. Mediation technology is the equipment by which the human body and consequently also the human mind, can be given an extension in the digital musical domain thereby enabling the performer/user to explore the domain of sound and music. From the embodied music cognition perspective the focus is on the acting body as the center of the mediation process that makes it possible to behave in resonance with music. During musical interaction, a musical instrument or an electronic device typically extends the human body. These body extensions allow mediation between subjective experience and physical (i. e., musical) reality. In regard to that process a musical instrument is considered to be the most natural corporeal–technological mediator [36.42]. But, we do not yet fully understand the way in which technology mediates perception and action, neither how mediation tools are dealt with. It is even more challenging when entering the domain of health and well-being, where users of mediation tools are foremost nonmusicians. Moreover, when designing tools for older people, one is likely to be challenged by difficulties of adaptability and openness to new experiences involving new devices.

Interaction with mediation technology for musical activities also confronts researchers with the challenge of building and testing of paradigms. So far there has been little research looking at benefits of musical interaction by means of body-extending devices for health and well-being. In a study that investigates musical interaction between normal-hearing people and hearing-impaired individuals with cochlear implants, *Lesaffre* et al. [36.43] present an experimental framework to test the user's sense of embodiment in sound identification and creation. Here attention is paid to the possibilities of coupling the well-known psychological concepts of *flow* and *presence* as a valuable top-down strategy for the design of embodied music interaction tools. It was shown that these states of mind are important factors in the experience of having fun or immersion in a shared environment and therefore are valuable concepts for exploring future applications.

Apart from the problems encountered by working with existing technologies, developing new tools or applications that expand the natural link between music and movement is even more thought provoking. For instance, *Moens* et al. [36.44] developed an interactive music player called D-Jogger that analyzes body movement in order to dynamically select music and adapt its tempo to the user's pace. This smart music player can use a wide range of sensors such as 3-D accelerometers, gyroscopes and pressure sensors, capable of measuring the users' movements. D-Jogger was used to explore how music can entrain human walkers to synchronize to the musical beat without being instructed to do so [36.45]. This study has shown that entrainment is controlled by brain mechanisms that work on time-differences between movement and music. Moreover, it was shown that perfect synchronization with the beat of the music has an enormous impact on the motivation to move. These insights show potential for performance enhancement in those who suffer from movement difficulties (e.g., patients with Parkinson's disease) or those who seek a movement boost (e.g., for sports performance).

However, to use such applications in a rehabilitation program it has to be demonstrated that these devices are effective through clinical trials. At this point in time, the barriers they are confronted with often discourage researchers. We report on a selection of barriers in the following. Time constraints mean that clinicians and therapists are not keen on adapting their program to the testing of new tools. They would rather hang on to a number of different therapeutic technologies that are available for use in clinics. But, the value of these technologies to the treatment program is not well defined. The use of technology in health and well-being is highly dependent on the application or therapy program for which it will be used. Besides, its effectiveness will also be determined by the user's receptiveness, attitude, and motivation. Furthermore, new and unfamiliar technologies can generate ethical concerns. Therefore, trust and confidence issues must be addressed.

User-Centered Research in MH and W/User Involvement

User-centered studies are more and more seen as the cornerstones of working with music mediation technology [36.46]. This transition implies an expansion of methodologies for addressing relevant features of musical action, musical tool use, and user-based tool development. The benefits of user-centered research methodologies in general are well documented in the literature [36.47]. However, apart from some investigations in the domain of *music information retrieval* (for an overview see [36.48]), using the principles of user-centered design is still underexplored in music research. Thus, along with the increased concentration on tools and mediation technologies goes the challenge of involving a broad range of stakeholders throughout the process of tool development, system design, and empirical studies.

The process of *user involvement* is a joint activity of a cross-disciplinary team of stakeholders that cooperate throughout the whole research procedure. However, it is not straightforward to involve users in empirical music research activities, especially with regard to health and rehabilitation. User involvement still faces many barriers, some of which are not easy to pass. A not exhaustive list of important reasons includes the following:

- In a caregiving context, users are time and time again approached as a patient, i. e., in terms of their illness or impairment. Being approached as a patient is an obstacle that discourages many people from being involved in empirical research.
- Ethical constraints force people with an impairment or disease to come to a hospital to test newly developed tools and programmes. Their personal experiences associated with illness, injury, and contact with doctors may hinder their participation.
- Professionals in healthcare are not always supportive of user involvement, for example because effective teamwork may require a restructuring of therapy programmes.
- Problems due to limited budgets are likely to be encountered, including time costs for the professionals and users themselves.
- Many of the products that are available on the market, such as smart phone applications for physical activities, have not been scientifically validated by user research.
- Involving users requires that designers cope with the huge variability among people's abilities and demands. There is a need for new testing paradigms that can cope with differences in personality and levels of impairment.
- Involving users with different ability levels, needs, and backgrounds requires tools to be developed and therapy programs to be conceived at an individual level.

Teamwork of stakeholders, including users, can be characterized as a multistage problem-solving process. It not only requires that researchers analyze and foresee how users in the field of health and well-being are likely to interact with a music interaction tool, the validity of the assumptions should be tested as well. Although this problem-solving activity is a simple reiterative process, given the schedule pressures of many research projects, in reality it is not widely implemented and therefore requires more attention. A supportive framework is presented that breaks the user-involvement process into five phases that are:

1. Defining user requirements
2. Outlining user strategies
3. Designing tools and experiments
4. Investigating user experience
5. Building a user interface (Fig. 36.1).

User requirements are identified early on in a research project. The process of defining user requirements starts from theoretical models as a foundation for case studies. Ideally, all researchers, experts, and beneficiary parties in the field of health and well-being (e.g., nursing, physiotherapy, gerontology, and ethics) that develop user strategies, user-experience designs, and/or user interfaces for a music interaction system

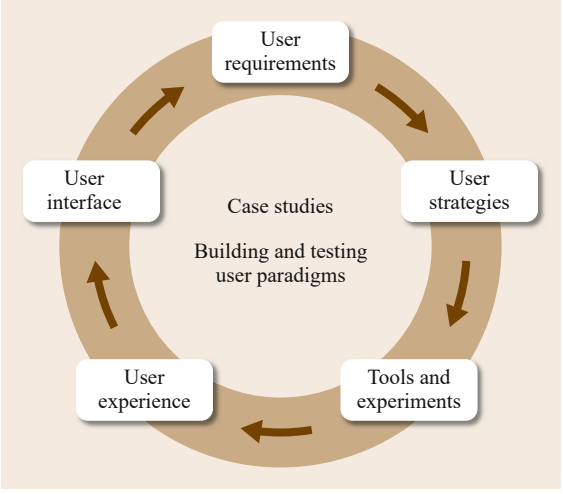


Fig. 36.1 Framework for the multistage problem-solving process of user involvement representing the reiterative procedure of defining user requirements, outlining user strategies, designing tools and experiments, investigating user experience, and building a user interface

will participate directly in gathering and defining the necessities of a specific target group. Although many researchers acknowledge the importance of identifying user requirements and their input into the process of designing tools and experiments, they often have no real plan as to how to achieve this. User requirement activities targeting the elderly may for example address questions concerning what kind of music devices they generally use, what kind of approach has the potential to add value to the daily lives of persons with impaired motor skills, or what musical preferences people with dementia have.

User strategy basically refers to the outlining of a clearly specified plan that flows from knowledge of user requirements of specific target groups and is initially created from the ideal user experience perspective. Typically, user strategy is embedded in *use cases* wherein certain assumptions are made, for example about features of action–perception coupling. Currently, user-oriented strategies have been developed in the direction of person-centered approaches where successful identification and interpretation of individual characteristics is at the focus of attention. Especially healthcare is becoming more and more individualized. The shift to practices where a patient becomes an active subject rather than a mere object of care is rooted in humanistic psychology and the person-centered therapy advocated by *Rogers* [36.49]. This type of therapy encompasses a nondirective, empathic approach that empowers and motivates a person in the therapeutic process. From that perspective the motivational power of music, for example to move our body, is still underexploited. *Van Noorden* [36.50] explains in his work the systemic principles of why music motivates movement.

Ideally, the process of *designing* (*music interaction*) *tools and experiments* for music applications for health and well-being is guided by the principles of *participatory design* and person-centered approaches. Participatory design [36.51] represents part of the cross-disciplinary collaborative work and implies that stakeholders with different expertise and knowledge have a deciding vote in the design process. However, inclusion of patients in tool design remains a challenge, especially when their characterization is atypical, as is the case with disabled persons who do not have the capabilities of persons with normal motor skills. The person-centered approach implies that music interaction and analysis tools need to be flexible enough to customize in situations where a paradigm for each different user type is required. In view of adapting the design process to disabled users, *Veytizou* et al. [36.52] for example propose a method that integrates parameters such as motor specificities and progression over time in the design process. *Holone* and *Herstad* [36.53] use the strategy of participatory design and *design for all* in a project that aims at improving health and quality of life for persons with severe disabilities by means of *co-creation* between children, their families, and caregivers.

User experiences deal with both experience related to the *usability of tools* and experience of embodied interaction with music. The study of effectiveness of tools and programs looks into users' corporeal articulations, behaviors, attitudes, and emotions about a particular system or tool. Experiences of musical interaction involve ways in which aspects of embodied experiences lived by means of music may support physical, mental, and social well-being. User experiences, more than anything else will determine *user satisfaction*. Interaction with embodied mediation technology for musical activities requires building and testing of paradigms of user experience. A study probing users' sense of embodiment in sound identification and creation has, for example, confirmed the importance of an approach that is concerned with embodied interdependency between the user and music mediation technology [36.43]. One problem that could be tackled through cross-disciplinary cooperation is the alignment of the functionality of technologies with the agency of the user. Designers often treat users as secondary to their tools or do not have sufficient time to follow the whole process. For example, when it comes to the development of musical applications for people with neurodegenerative diseases the design process requires analysis of user actions and experiences of the specific target population in order to move beyond merely addressing its functionality.

The *user interface* is the space where human–machine interactions occur. It may consist of visual, auditory, and/or other components through which users may interact directly or indirectly with the system. *User interaction* is facilitated through a musical user interface, which can be personalized to meet user's preferences and deliver the necessary cues. For example, *Lesaffre* et al. [36.41] (submitted) developed a musical balance board that is a musical user interface for retraining of balance in stroke patients. The system provides real-time sonification of weight distribution together with the motivational effect of music to retrain maintenance of equilibrium.

A last question to be asked is who the *potential users* for musical applications for health and well-being are. They are numerous, being everyone, disabled and not disabled, that can benefit from communicating, experimenting, playing, and training with music. Music tools may for example address sportsmen looking for new training possibilities that support prevention of injuries or individuals with specific impairments such as

the retraining of balance in hemiplegic patients. However, apart from finding and prioritizing the *right* user needs it is also important to address the scope of users from which to draw insights. From the perspective of the embodied music cognition paradigm, the scope of potential users (in therapeutic interventions and beyond) could be narrowed down to persons who comply with some or all of the following characteristics:

- Like, enjoy, or have an interest in music
- Being sensitive to musical reward experiences
- Have a positive attitude towards (new) technology
- Are willing to experiment and perform musical activities
- Have adequate motor skills
- Are willing to perform within a group (co-creation)
- Are able to hear sound (or at least feel it).

36.4 Conclusion

Starting from transitions in society and musicology, in this chapter a framework is provided that is meant as a support for providing an answer to the question of how to engage with scientifically validated music applications for health and well-being. The presented framework is stimulated by the belief that the scientific expertise to be gained from using the embodied music cognition approach in promoting music and movement interaction for health and well-being holds promise for future achievements. This chapter has discussed models of music, health and well-being. It was shown that the growing prominence of new tools, technology-driven research, and user approaches indicate key developments toward achieving demands of bringing the embodied music cognition approach into the practice of musical interventions and therapeutic programs for health and well-being. To conclude, this final summary brings together the key challenges that were explained in this chapter.

With respect to theory:

- Cross-disciplinary research in the context of the embodied music cognition framework requires the integration of theoretical principles into the empirical framework in a way that enables all stakeholders to make sense of the complexities of concepts and methods applied.
- The connectedness of the multiple dimensions of health requires the study of the relationships between these dimensions rather than focusing on one dimension alone.
- The development of musical interventions and musical therapeutic programs should be grounded in evidence-based theories.

With respect to technology:

- The shift to technology-driven research requires the development of new tools that expand the natural link between music and movement.
- Testing of technologies, methods, and associated tools needs to be done in accordance with predefined user requirements and strategies.

With respect to users:

- The multistage problem-solving process of user-centered research requires involvement of all stakeholders throughout the process.
- User-centered strategies need to be expanded to account for person-centered approaches.
- Ecological approaches are required that recognize the interconnectedness of individuals and their personal context.

With respect to modeling and analysis:

- A multimodal modeling approach is a prerequisite for allowing integrated representation and analysis of high-dimensional data.
- The analysis framework should be developed in such a way that the subjective top-down and the objective bottom-up approach become tightly connected to each other.

References

36.1 E. Ruud: *Music Therapy: A Perspective from the Humanities* (Barcelona, Gilsum 2010)

36.2 E. Ruud: The new health musicians. In: *Music, Health and Well-Being*, ed. by R. MacDonald, G. Kreutz, L. Mitchell (Oxford Univ. Press, Oxford 2012) pp. 87–96

36.3 T. De Nora: Music sociology: Getting the music into the action, Br. J. Music Educ. **20**(2), 165–177 (2003)

36.4 R.A.R. MacDonald, G. Greutz, L. Mitchell: *Music Health and Well-Being* (Oxford Univ. Press, Oxford 2013)

36.5 M. Lesaffre: The power of music and movement to reinforce well-being. In: *The Power of Music. Researching Musical Experiences: A Viewpoint from IPEM*, ed. by M. Lesaffre, M. Leman (Acco, Leuven 2013)

36.6 M. Leman: *Embodied Music Cognition and Mediation Technology* (MIT Press, Cambridge 2008)
36.7 R.I. Godøy, M. Leman (Eds.): *Musical Gestures: Sound, Movement and Meaning* (Routledge, New York 2010)
36.8 D.E.S.A.D.E.S.A. Report: *World Population Aging* (United Nations Department of Economic and Social Affairs, Population Division, New York 2013)
36.9 P. Horden (Ed.): *Music as Medicine. The History of Music Therapy Since Antiquity* (Ashgate, Aldershot 2000)
36.10 L. Bunt, B. Stige: *Music Therapy: An Art Beyond Words*, 2nd edn. (Routledge, London 2014)
36.11 A. Antonietti: Why is music effective in rehabilitation? In: *Advanced Technologies in Neurorehabilitation*, ed. by A. Gaggioli, E. Keshner, P.L. Weiss, G. Riva (IOS, Amsterdam 2009) pp. 179–194
36.12 T. Hays: Facilitating well-being through music for older people with special needs, Home Health Care Services Q. **25**, 55–73 (2006)
36.13 A.A. Clair, J. Memmott: *Therapeutic Uses of Music with Older Adults*, 2nd edn. (American Music Therapy Association, Silver Spring 2008)
36.14 WHO: Constitution of the World Health Organization, Am. J. Public Health Nations Health **36**(11), 1315–1323 (1946), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1625885/
36.15 B. Üstün, R. Jakob: Calling a spade a spade: Meaningful definitions of health conditions, Bull. World Health Organ. **83**(11), 802 (2005)
36.16 D.A. Dewsbury: Psychobiology, Am. Psychologist **46**(3), 198–205 (1991)
36.17 J.S. Gordon: Holistic medicine and mental health practice: Toward a new synthesis, Am. J. Orthopsychiatry **60**(3), 357–369 (1990)
36.18 S. Chatterji, B.L. Ustün, R. Sadana, J.A. Salomon, C.D. Mathers, C. Murray: *The Conceptual Basis for Measuring and Reporting on Health* (World Health Organization/Global Program on Evidence for Health Policy, Geneva 2002) Discussion Paper No. 45
36.19 D. Kahneman, E. Diener, N. Schwarz: *Well-Being: The Foundations of Hedonic Psychology* (Russell Sage, New York 1999)
36.20 J. Hjelm: *The Dimensions of Health: Conceptual Models* (Jones Bartlett, Sudbury 2010)
36.21 D. Cutler, M.B. Landrum: Dimensions of health in the elderly population. In: *Investigations in the Economics of Aging*, ed. by A. David (Univ. Chicago Press for National Bureau of Economic Research, Chicago 2012) pp. 179–201
36.22 S.B. Hanser: Music, health and well-being. In: *Handbook of Music and Emotion: Theory, Research, Applications*, ed. by P.N. Juslin, J.A. Sloboda (Oxford Univ. Press, Oxford 2010) pp. 849–877
36.23 L.O. Bonde: Health music(k)ing – Music therapy or music and health? A model, eight empirical examples and some personal reflections, Music Arts Action **3**(2), 120–140 (2011)
36.24 B. Stige: *Culture-Centered Music Therapy* (Barcelona, Gilsum 2002)
36.25 T. De Nora: *Music Asylums: Well-Being Through Music in Everyday Life* (Ashgate, Farnham 2013)
36.26 L. Shapiro: The embodied cognition research programme, Phil. Compass **2**, 338–346 (2007)
36.27 S. Gallagher: Interpretations of embodied cognition. In: *The Implications of Embodiment: Cognition and Communication*, ed. by W. Tschacher, C. Bergomi (Imprint Academic, Exeter 2011) pp. 59–71
36.28 G. Rizzolatti, C. Sinigaglia: *Mirrors in the Brain: How Our Minds Share Actions and Emotions* (Oxford Univ. Press, Oxford 2008)
36.29 M. Leman: Fundamentals of embodied music cognition: A basis for studying the power of music. In: *The Power of Music. Researching Musical Experiences: A Viewpoint from IPEM*, ed. by M. Lesaffre, M. Leman (Acco., Leuven 2013) pp. 17–34
36.30 M. Leman: Musical entrainment subsumes bodily gestures: Its definition needs a spatiotemporal dimension, Empirical Musicology Rev. **7**(1/2), 63–67 (2012)
36.31 L. De Bruyn, D. Moelants, M. Leman: Sound caterpillar: Assessment and training of sound and music perception skills in hearing impaired children. In: *Proc. 11th Int. Conf. Music Percept. Cogn. (ICMPC 11)* (2012)
36.32 P.J. Maes, D. Amelynck, M. Lesaffre, D.K. Arvind, M. Leman: The conducting master: An interactive, real-time gesture monitoring system based on spatiotemporal motion templates, Int. J. Hum.-Comput. Interaction **29**(7), 471–487 (2013)
36.33 C. Nombela, L.E. Hughes, A.M. Owen, J.A. Grahn: Into the groove: Can rhythm influence Parkinson's Disease?, Neurosci. Biobehav. Rev. **37**(10), 2564–2570 (2013)
36.34 D.M. Wolpert, J. Diedrichsen, J.R. Flanagan: Principles of sensorimotor learning, Nat. Rev. Neurosci. **12**(12), 739–751 (2011)
36.35 P.-J. Maes, E. Van Dyck, M. Lesaffre, P.M. Kroonenberg, M. Leman: The coupling of action and perception in musical meaning formation, Music Percept. **32**(1), 67–84 (2014)
36.36 J.W. Davidson: The role of the body in the production and perception of solo vocal performance: A case study of Annie Lennox, Musicae Scientiae **5**(2), 235–256 (2001)
36.37 J. Davidson, A. Emberly: Embodied musical communication across cultures: Singing and dancing for quality of life. In: *Music, Health and Well-being*, ed. by R.A.R. MacDonald, G. Kreutz, L. Mitchell (Oxford Univ. Press, Oxford 2012) pp. 163–149
36.38 M.F. Folstein, S.E. Folstein, P.R. McHugh: Mini-mental state: A practical method for grading the cognitive state of patients for the clinician, J. Psychiatric Res. **12**, 189–198 (1975)
36.39 J.E. Ware: SF-36 health survey (SF-36). In: *Handbook of Psychiatric Measures*, ed. by American Psychiatric Association, A.J. Rush (American Psychiatric Association, Washington DC 2000)
36.40 T.P. Ettema, R.M. Dröes, J.D. De Lange, M.W. Ribbe: QUALIDEM: Development of a dementia-specific

quality of life instrument. Scalability, reliability and internal structure, Int. J. Geriatr. Psychiatry **25**, 549–556 (2007)

36.41 M. Lesaffre, B. Moens, F. Desmet, M. Leman: Monitoring music and movement interaction in people with dementia. In: *The Routledge Companion to Embodied Music Interaction*, ed. by M. Lesaffre, P.-J. Maes, M. Leman (Routledge, New York 2017) pp. 294–303

36.42 L. Nijs, M. Lesaffre, M. Leman: The musical instrument as a natural extension of the musician. In: *Music and Its Instruments*, ed. by M. Castellengo, H. Genevois (Editions Delatour France, Sampzon 2013)

36.43 M. Lesaffre, L. Nijs, M. Leman: Interacting with music mediation technology for hearing impaired – First tests with normal hearing subjects. In: *Proc. Eur. Soc. Cogn. Sci. Music (ESCOM), Jyväskylä* (2009) pp. 266–270

36.44 B. Moens, L. van Noorden, M. Leman: D-Jogger: Syncing music with walking. In: *Proc. Sound Music Comput. Conf., Barcelona* (2010) pp. 451–456

36.45 B. Moens, C. Muller, L. van Noorden, M. Franec, B. Celie, J. Boone, J. Bougois, M. Leman: Encouraging spontaneous synchronisation with D-Jogger, an adaptive music player that aligns movement and music, PLOS ONE **9**(12), 40 (2014)

36.46 M. Leman, M. Lesaffre, A. Deweppe, L. Nijs: User-oriented studies in embodied music cognition research, Musicae Scientiae Special Issue **14**(Suppl 2), 203–224 (2010)

36.47 V.A. Lofthouse, D. Lilley: What they really, really want: User centred research methods for design. In: *Proc. Int. Design Conf.-Design, Dubrovnik* (2006)

36.48 M. Schedl, A. Flexer: Putting the user in the center of music information retrieval. In: *Proc. 13th Int. Soc. Music Info. Retrieval (ISMIR)* (2012) pp. 385–390

36.49 C. Rogers: A theory of therapy, personality and interpersonal relationships as developed in the client-centered framework. In: *Psychology: A Study of a Science. Formulations of the Person and the Social Context*, Vol. 3, ed. by S. Koch (McGraw Hill, New York 1959)

36.50 L. van Noorden: Fundamentals of music and movement: Towards an understanding of the motivational power of music. In: *The Power of Music. Researching Musical Experiences: A Viewpoint from IPEM*, ed. by M. Lesaffre, M. Leman (Acco, Leuven 2013) pp. 97–112

36.51 F. Kensing, J. Blomberg: Participatory design: Issues and concerns, Comput. Support. Co-op. Work **7**(3/4), 167–185 (1998)

36.52 J. Veytizou, C. Magnier, F. Velleneuve, T. Thomann: Integrating the human factors characterization of disabled users in a design method, Assoc. Adv. Model. Simul. Tech. Enterp. **73**(3), 173 (2012)

36.53 H. Holone, J. Herstad: RHYME: Musicking for all, J. Assist. Technol. **7**(2), 93–101 (2013)

37. A Conceptual Framework for Music-Based Interaction Systems

Pieter-Jan Maes, Luc Nijs, Marc Leman

Music affords a wide range of interactive behaviors involving social, cognitive, emotional, and motor skills. In this chapter, we consider the role of technologies in relation to these interactions afforded by music. A general conceptual model is introduced that forms a basis to frame and understand a vast number of music-based interactive systems. In this model, we consider the necessity of coupled action–perception processes, in combination with human reward, prediction and social interaction processes. In addition, we discuss three perspectives on how music-based interaction systems may involve users' actions (monitoring, motivation, and alteration). To conclude, we discuss two case studies of technologies to illustrate the most innovative aspects of the presented model.

A distinct characteristic of music is that it stimulates active behavior, individually and collectively. For instance, music performance brings people together to actively play an instrument in coordination with others. Also, listening to music spurs people to move along with the beat and other musical features. Accordingly, music can be considered an excellent medium that affords and reinforces sensorimotor interactions between humans, and their sensory environment. The impact of music is far-reaching and may have a bearing upon general skills that are conducive to the general development of human beings. For instance, both performing and listening to music have been shown to promote interpersonal entrainment, or *joint action*, which engenders social skills, such as imitation, collaboration, the nonverbal understanding and sharing of intentions, hierarchy formation, empathy, etc. [37.1, 2]. Also, music performance and music listening enable people to explore, experience, and develop human affective engagement with expanded complexity and phenomenal character [37.3]. In addition, it has been suggested that people's engagement with music calls on temporal processing systems, which in turn may develop domain-general cognitive abilities such as sequential learning, temporal integration, and serial recall [37.4, 5]. Finally, music may facilitate auditory cueing, motivation, and/or diversion, which may contribute to movement performance in sports and motor rehabilitation [37.6–8]. In short, music affords a wide range of interactive behaviors involving social, cognitive, emotional, and motor abilities.

In this chapter, we consider the role of technologies in relation to these interactions afforded by music. We claim that music-based interactive systems can reinforce interactions that contribute to a heightened experience of music, and to the improvement of domain-general social, cognitive, affective, and motor skills. Technologies are thereby considered extensions of the human body. Acoustic musical instruments are a basic example of technologies that extend the human body: endowed with its sensorimotor capabilities, into the external (musical) environment [37.9]. However, since the 1990s advances in electronic and digital technology development have drastically changed the landscape of music production; the emergence of sensors, motion-capture technology and digital sound

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_37

processing modules have expanded the possibilities for interaction. Moreover, this has not been limited to only sound production. For instance, many technologies are developed specifically to enrich people's experience of listening to music; think of the mobile phone, wearable digital media players, or sophisticated music streaming-on-demand services. Also, new music technology has created an immense potential for assisting and optimizing teaching methods in a music educational context [37.10–12], or for enhancing social interaction [37.13].

However, despite the possible benefits of using new electronic and digital technologies, they have also introduced new problems and challenges with regard to musical interaction and communication. For example, a fundamental problem inherent in digital music production is the fact that the mediation between the different modalities (basically from movement to sound) has an arbitrary component, which is due to the fact that the energies of the modalities are transformed into electronic signals. This is not the case with traditional instruments, where energetic modalities are mechanically mediated and where the user gets a natural feeling of the causality of the multimodal interface. Therefore, in the domain of electronic and digital music production, there is a need for more *transparent* mediation technologies that create a feeling of nonmediation; as if the mediation technology *disappears* when it is used [37.14]. Similar issues exist in other music-based interaction technologies. Researchers therefore believe that music-based technology development must be informed by knowledge on the cognitive, sensorimotor, and social dispositions and capabilities of human beings. This requires the integration of fundamental research in order to fully exploit the potential of electronic and digital music technologies.

This chapter introduces a general conceptual model that forms a basis to frame and understand a vast number of music-based interaction systems. The model is intended to serve as a guideline for practical developments in the future. Its focus is on music-based technologies that enable action and interaction (human–music interaction, and human–human interaction). These technologies are mediation technologies in the sense that they offer ways to extend the coupled action–perception mechanisms and sensorimotor integration processes that are *naturally* involved in people's engagement with music. The model expands the concept of action–perception couplings so as to incorporate the concept of the human reward system and social interaction processes. We then go on to discuss three perspectives on how music-based interaction systems may involve users' actions (monitoring, motivation, and alteration). To conclude, we discuss two case studies of technologies to illustrate the most innovative aspects of the presented model.

37.1 A Conceptual Model of Music-Based Interaction Systems

The classical, *information processing* approach in cognitive science considered behavior and perception as separate peripheral processes outside central cognition [37.15]. By now, this model has been further developed in favor of embodied accounts that emphasize the central role of sensorimotor interaction with the external environment in human behavior and perceptual-cognitive processes [37.16]. This interaction requires the integration (coupling) of action and perception processes. Acting in and onto the environment guides perceptual-cognitive processes, which reciprocally influence motor control and behavior. Moreover, this interaction process is driven by subjective aspects, such as intentionality, reward and affect, and is situated within a specific context (cultural, social, etc.).

Our conceptual model of music-based interaction systems considers technologies from the perspective of this action–perception loop and its role in human–environment interactions. We consider the aim of technology development to intervene the natural coupled action–perception processes, in order to extend the cognitive, sensory, sensorimotor and social capabilities, and to reinforce a human's interaction with the external environment (Fig. 37.1). Such *mediation* technologies enable an engagement into new realms and an opening up of exciting possibilities, such as those in the domain of electronic and digital music processing [37.14]. This approach to technology development is closely related to the theories of the *extended mind*, where humans use technology to connect with the musical environment [37.17], and *situated cognition* where the goal is to find a way of interacting with the environment such that meaning emerges from it [37.18].

Mediation technologies that intervene in the action–perception loop in support of sensorimotor interactions contain two components. First, systems must be allowed to capture actions and to align them with sensory outcomes. Typically, this is what traditional musical instruments do. However, the ability to extend actions into the electronic and digital domain, drastically increases the possibilities for sound production. Unlike traditional instrument teaching where the input–output

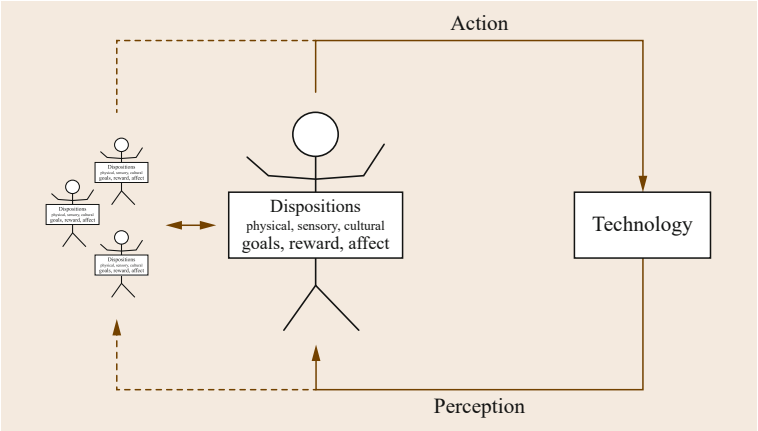


Fig. 37.1 Schematic overview of the conceptual model of music-based interaction systems

is determined, the use of electronic and digital devices allow for a complete rethink on the familiar relationships between physical input and sound output and to incorporate aspects of instrumental design related to sound design and interaction modes [37.19]. For example, obvious relationships can be reversed (e.g., small actions produce loud sounds, short notes trigger extended sounds). The possibility of rethinking the input–output model provokes musical exploration and development of novel emotional experiences [37.3]. Additionally, in transferring actions into sensory outcomes, systems could contribute to the multisensory monitoring of one's action and develop awareness of an individual's movement performance [37.20]. The availability of monitor movement performance could be exploited in various domains of research and practice, such as dance, sports and motor rehabilitation.

Second, mediation technologies may focus on the alignment of actions (i. e., movement responses) to specific musical and/or multisensory feedback or stimuli [37.21]. The applications are numerous. Movement responses to music have been shown to improve the perception of structural and expressive musical structures [37.22]. Additionally, music's ability to guide movement responses to specific goals may be used for sports and rehabilitation purposes, or it may facilitate social interaction and entrainment. Although musical mediation systems may capitalize on one or the other of the components just discussed, *real* interaction exists only when the two components are operating in combination. In that regard, actions are transferred to appropriate sensory outcomes and, in turn, these sensory outcomes are generated such that they support or adapt actions towards specific expressive, motor, or social goals.

In Sects. 37.2 and 37.3, we discuss in more detail how coupled action–perception processes can be further linked to the physiological processes related to reward, and to social interaction [37.23]. More importantly, we discuss these processes in light of our conceptual model, and how music-based interaction technologies can tap into the processes underlying reward and social interaction in order to reinforce musical experiences, as well as enhancing various skills (Fig. 37.1).

37.2 The Human Reward System

A somewhat disregarded aspect of interactive music systems is the role of the human reward system, and related subjective phenomena such as motivation, intentionality, affect, etc. However, a closer look at the role of reward in the design of interactive music systems shows that it is a fundamental component of the interaction dynamics. Current thinking on the reward system understands the process as a collection of neural structures that contain dopamine-secreting neurons in the midbrain with pathways to other brain structures such as the nuclei accumbens and prefrontal cortex [37.24, 25]. In general, reward serves as a *reinforcer*, motivating and regulating voluntary behavior control by inducing feelings of pleasure and happiness.

In the following, we provide arguments to support our claim that the reward system and reward values play an essential role in many aspects of electronic and digital music systems. First, we discuss the importance of reward values in learning processes that are required in many of the interactive music systems. Secondly, we discuss how reward, and musical pleasure during music listening relates to the ability to anticipate and predict

forthcoming events. Moving on the argument, we extend this idea to the coupled action–perception loop involved in music performance, and in (synchronized) movement responses to music.

Research suggests that dopaminergic projections to prefrontal cortices underpin executive functions related to attention, memory, and learning [37.26–30]. Of major importance are the findings that the close intertwining of the reward and cognitive systems in learning processes is modulated by prediction mechanisms. Traditional learning theories emphasize the role of temporal *contiguity* in associative learning processes. That is, whenever two events or phenomena occur quasiconcurrently, such as a stimulus and a reward, they become associated. However, current learning theories suggest that learning processes are grounded in the discrepancies that occur (errors) between what is expected/predicted and what actually happens [37.28, 30–32]. In these studies, it is shown that dopamine neurons encode reward learning prediction errors; a positive response of dopamine neurons occurs when a reward is given unexpectedly, while this response gradually decreases as that reward becomes increasingly predictable. Importantly, it has been demonstrated that high stimulus-reward predictability, and a corresponding decrease of dopaminergic activity, prevents the concurrent (behavioral and neuronal) learning of other stimuli–reward relationships (cf. blocking paradigm; [37.30]). These findings stress the necessity to integrate issues such as surprise and novelty into the design of music mediation technologies that almost always involve some degree of learning.

In a recent study, *Zatorre* and *Salimpoor* [37.33] investigated the neural bases of the feeling of pleasure during music listening. It was found that temporal expectancies, and their associated predictions again play a key role. When listening to music, incoming auditory information is processed by auditory cortices and, in interaction with frontal cortices, fit into *templates*, or a *musical lexicon* [37.34]. Based on these templates, representations of structural regularities in the music can be generated, which in turn facilitate the generation of expectancies and associated predictions. Emotional responses and the sense of reward then result from the degree of confirmation or violation of predictions. An important remark made by *Zatorre* and *Salimpoor* [37.33] is that the *templates* used to fit incoming auditory information, and eventually generate predictions and musical pleasure, are developed through prior musical experiences within a specific context. Accordingly, the authors suggest that these mechanisms could possibly explain the variation and diversity in peoples' musical choices.

The ability to anticipate or predict forthcoming events has been identified as a major source of reward in musical activities that involve sensorimotor coordination and synchronization [37.35, 36]. This can relate to the production of sound (as in music performance), as well as to (synchronized) movement responses to music (as in dance, or walking/running to music). Much like playing a *traditional* musical instrument, learning to play an electronic and/or digital instrument for gesture-based musical expression (e.g., http://www.nime.org/archive/) requires sensory-motor association learning in which action and perception become intricately interwoven. Playing such an instrument can be considered a goal-directed, intentional act, with the production of sounds being a primary goal. Reaching that goal requires obtaining knowledge about the relationship between the actions afforded by the instrument, and the auditory consequences of these actions. This knowledge is gradually acquired by exploring and manipulating the possibilities afforded by the instrument using (at first) arbitrary actions that lead to (at first) unexpected auditory events [37.37]. In that process of exploration and interaction, one systematically and repeatedly associates performed actions with heard sounds. At that point, playing an instrument may become a goal-directed act, in the sense that performers have the ability to exert control over the auditory outcome of performed actions. Similarly, body movement responses to music are often (spontaneously) performed in synchrony with particular musical structures, most often the musical beat. Auditory–motor synchronization, or *entrainment*, requires finely-tuned sensorimotor coordination to align spatiotemporal motor patterns with musical features. A successful alignment requires prediction mechanisms to anticipate forthcoming events. Accordingly, the successful prediction of how musical passages unfold may arouse feelings of reward, and even an imaginary sense of control, or *agency*; as if one produces the music oneself (e.g., air guitar).

To conclude, the proposed model incorporates aspects of reward, motivation, and affect into the design of music-based interaction technologies. Taking into account this motivational aspect in relation to the cognitive aspect (e.g., learning, prediction), an important challenge lies ahead. We have shown that the reward system capitalizes on prediction mechanisms that relate to music listening as well as to the (implicit or explicit) knowledge of how specific actions lead to specific sensory outcomes. Up until quite recently, it has been argued that reward outcomes are dependent on successes as well as on failures in prediction, or in other words, on confirmed as well as violated expectations [37.35]. However, it is of valid interest

to consider the roles of sensorimotor control (when sensorimotor expectancies are confirmed and thus provide a sense of control), and of surprise and novelty (when sensory outcomes do not match the expected outcomes) [37.36]. Additionally, it is important to note that prior experiences define, to a large extent, the reward values and outcomes of specific interactions. Therefore, music-based interaction technologies should ideally be capable of taking into account user-related backgrounds.

The focus on the human reward system extends our model, which views mediation technologies as tools to intervene and augment the action–perception loop involved in music-based interactions. In a sense, technologies may be considered tools that *grant access* to the human reward, and by extension the physiological response system. Sensorimotor interaction processes could arouse feelings of reward and pleasure, and in reverse feelings of reward and pleasure may evenly facilitate sensorimotor (learning) processes.

37.3 Social Interaction

Music-based interaction systems often function within a social context. Music is indeed a particularly powerful medium that brings people together, as it invites people to dance or to produce music together. Accordingly, music provides an excellent platform to express, share, and understand other people's emotions in a direct, nonverbal way. Research suggests that understanding the feelings and emotions of others (i. e., *empathy*) is, at least partly, an action-based process modulated by the human mirror neuron (HMN) system [37.38]. The main idea is that the mere observation of another person's behavior activates (simulates) corresponding motor representations in the brain of the observer, which facilitates psychological inference about the other person's mental state, and subsequently the person enters in an empathic relationship. Another characteristic of many social interactions, apart from empathy, is *joint action*, a concept referring to situations where actions are coordinated in conjunction with others in order to reach a shared goal. These two concepts, empathy and joint action share a common denominator in that they both promote prosocial behavior and, by extension, may positively contribute to people's happiness, well-being, and health. Multiuser social music systems are perfectly suited to stimulate communication and interaction in socially situated activities [37.39, 40]. For instance, listening to music solicits people to move in synchrony with each other, which may reinforce empathic relationships. Also, producing music engages and challenges people to jointly coordinate their actions in order to realize a particular musical (expressive) outcome [37.41–43]. Thereby, new technologies offer a wide range of opportunities to reinforce these interactions. For instance, motion capture technologies enable a transfer of body movements into multisensory representations, which may help people to align rhythmical body movements (visual, auditory, tactile, etc.). Digital audio signal processing technologies and (online) music databases also offer opportunities for people to experience a wide range of appealing auditory stimuli (for listening and production purposes). Additionally, online new media applications enable people to connect worldwide in their shared interests. Our main point here is to emphasize the idea that technologies serve not only to reinforce coupled action–perception processes and emotion-regulating processes in individuals, but also the interactions of these processes in social contexts.

37.4 Monitoring, Motivation, and Alteration

The above model suggests that engagement with music, in its broadest form, impels interactions between a wide range of brain functions involving cognitive, sensory, motor, and reward/emotion-related systems. The central point of this chapter is to consider the role new technologies take in reinforcing and extending these interactions, enriching people's engagement with music, and developing domain-general cognitive, social, and affective skills. In the following, building on the standpoint of recent overviews of embodied theories on perception, cognition and behavior [37.44], we consider the motor functions to be of central importance. Related to the design of music-based interaction technologies, motor functions can be approached from different perspectives. In the model that we introduce here, we consider three main perspectives; namely, how music-based interaction technologies can facilitate the monitoring, motivation, and alteration of actions and behavior [37.22].

Monitor Actions: Just as with any other type of human behavior, the active engagement with music

(e.g., dancing, performing) relies on an internal feedback mechanism that continuously monitors on-going behavior and provides the basis for a self-reflective psychological judgement about the appropriateness of one's musical behavior [37.45]. The feedback is generated in different sensory modalities (auditory, visual, tactile, kinaesthetic, and vestibular) and informs the degree to which musical intentions are successfully realized. This relies on the prereflexive control of actions, which is based on the combination of forward and inverse models: they enable a nonconscious comparison for the prediction of an action with the perceptual inputs [37.46]. When the coupling between actions and perception is experienced as matching, that is, when the outcome of an action satisfies the anticipated results, it generates a sense of being in control and of being the cause of altered inputs [37.47, 48]. Mediation technologies enable quantifying and displaying aspects of motor behavior. By externalizing internal feedback mechanisms, i. e., translating actions into various sensory information streams, they increase the perceptual experience of one's own body (*body image*) and stimulate the awareness of self-generated movements. As such, attention might be directed to previously nonconscious operations of the body (*body schema*) making it possible to affect these operations [37.49]. This is particularly interesting as the body tends to efface itself from conscious experience in most behavior. Therefore, motor aspects of behavior (e.g., performance) might disappear under the radar of bodily awareness, including aspects that might hinder or even prevent an intended result (e.g., a technically smooth and expressive performance). Augmenting the natural monitoring mechanism with technologies that enable the objective measurement of motor behavior and the translation of the measurement into different kinds of feedback might optimize the fine-tuning of action–perception couplings, the unconscious monitoring of the body (body schema) and, subsequently, the general movement performance quality and efficiency. The use of sound and music in sonification or auditory display systems has been shown to be of particular relevance to improve various forms of movement behavior (for reviews, see [37.50, 51]).

Motivate to (inter)act: Music is known for its motivational qualities in exercise and sports [37.6, 7]. The motivational qualities of music have been attributed to four hierarchically related factors: rhythm, musicality, cultural impact, and association [37.7]. This aspect of motivation is closely related to the concept of reward as discussed earlier.

Alter actions: In many music-based interaction systems, one tries to guide movement behavior towards specific goals. Here, we introduce two strategies that may be applied to realize this goal. Both are informed by fundamental knowledge of sensorimotor control processes. These two strategies, based on synchronization and prediction mechanisms, can be applied in a wide range of domains where the goal is to have an impact on people's movement behavior (e.g., sports, musical instrument practice, motor rehabilitation, motor retraining, interactive sonification, social interaction, etc.).

37.4.1 Spontaneous Synchronization

Humans have a natural tendency to synchronize their movements with external rhythmical stimuli or with the rhythmical movements of others. Research suggests that spontaneous synchronization is a dynamical, self-organizing principle that strives towards finding stability (e.g., phase-locking) between people and their external environment [37.52–54]. This spontaneous drive towards synchronization can be used in the design of music-based interaction systems to steer the users' movement behavior towards specific goals by means of a specific presentation and/or manipulation of musical stimuli and/or auditory feedback [37.55, 56].

37.4.2 Spontaneous Motor Adaptation

The mechanism of motor adaptation offers another strategy on which music-based interaction systems can capitalize. This adaptation mechanism relies on learned sound–movement relationships, typically named *internal models*. Once internal models are developed, they enable a prediction of the auditory outcome of planned actions [37.22]. When a mismatch occurs between the expected and the actual outcome of performed actions, these actions will be altered in order to reduce prediction errors [37.57–63]. Sensorimotor adaptation is exactly that: a process in which motor commands are updated in response to altered environmental conditions. Accordingly, one can take advantage of this adaptation mechanism to guide people's movement behavior, first by developing sound–gesture relationships (through associative learning processes), and second by a deliberate manipulation of the self-generated auditory feedback.

37.5 The Evaluation of Music-Based Interactive Systems

The above-described perspectives emphasize the possible benefits of music-based interaction systems. Indeed, from a theoretical point of view, they hold promising potential. Also, a growing number of empirical studies suggest that the realization of this potential is feasible. However, careful consideration and adopting a critical stance towards their design, development and implementation are needed. Each of the presented perspectives conceals, equally, a potential disadvantage to the development of the proposed aspects (e.g., learning and reward) of technology-based musical interactions. For example, interactive systems that focus on monitoring may increase cognitive load, create dependency or stimulate an internal focus [37.64]. Interactive systems that focus on motivating behavior may succeed in providing experiences that are fun at first glance, but fail to generate the desired long-term effects. By acknowledging these potential pitfalls, we believe that introducing empirical testing and evaluating music-based interactive systems is of vital importance. Indeed, there are currently several approaches at our disposal. For example, one of the emerging approaches for testing and evaluating interactive music systems is to borrow tools and techniques from human–computer interaction (HCI), often focusing on the usability of controllers and interfaces [37.65–67]. Another approach focuses on measuring (video observation, questionnaires) the quality of the subjective experience while engaging with an interactive music system (e.g., [37.68, 69]).

However, empirical studies are scarce and often encounter problems of methodological robustness. For example, with regard to sonification, *Dubus* and *Bresin* [37.50] state that a proper evaluation of mappings is performed only in a marginal proportion of publications. *Collins* and *d'Escrivan* [37.70] observed that the evaluation of IMSs has often been inadequately covered in reports. With regard to educational interactive systems, *Nijs* and *Leman* [37.64] argue that studies are often based on one-time experiences, a limited number of participants and a lack of statistical analysis to support the findings.

In our view, the presented conceptual model provides a framework for the development of adequate evaluation methods and robust experimental designs based on qualitative and quantitative measurement. Our work suggests that the development of interactive music systems needs to appeal to the cognitive architecture encompassing the basic mechanisms (e.g., reward, prediction) that underlie musical interaction. As such, the evaluation of interactive music systems can probe the degree to which aspects of the system (e.g., feedback, controller) intervene with these basic mechanisms. For example, concerning the use of visual feedback, evaluative methods can be developed in order to determine how different kinds of visual feedback (e.g., concurrent versus terminal, actual versus modified, static versus dynamic) affect prediction mechanisms and, consequently, the reward system. Concerning the use of different movement sensing controllers, evaluation methods can be developed that probe qualitative changes in the use of the body while performing.

Evidently, the development of technology evaluation methods work in tandem with the development of methods used to evaluate aspects of musical interaction such as musical understanding (e.g., structure, harmonic progression) and creativeness.

To conclude, the empirical testing and evaluation of interactive systems is in need of an elaborated framework that would form the basis upon which the construction of robust empirical designs can be set. We believe the presented model illustrated here serves as the basis for such a framework.

37.6 Some Case Studies of Applications and Supporting Research

The following provides two examples of music-based interaction applications to illustrate some of the key assets of the introduced conceptual model.

37.6.1 Music Paint Machine

The Music Paint Machine (MPM) is an educational technology that allows a musician to create a digital *painting* by playing music while making various movements on a colored pressure-sensitive mat [37.69, 71] (Fig. 37.2).

The overall objective of the MPM is to support instrumental music learning and teaching by assisting the development of (1) musical creativity, (2) an embodied understanding of music, and (3) an optimal relationship between musician and musical instrument [37.72]. Interaction with the MPM is based on the real-time monitoring of a player's music and movement, and on translating extracted sound and movement parameters into a clear, visual representation. In this way, the MPM intervenes in the player's action–perception loop by complementing the *natural* multisensory monitoring

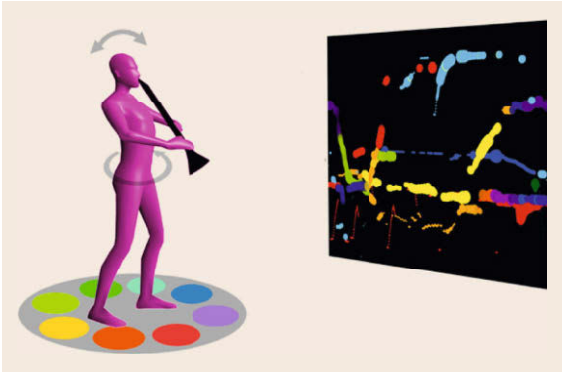


Fig. 37.2 The Music Paint Machine

of one's music playing with immediate (real-time) and unambiguous (based on objective measurement) visual feedback. Through the repeated multimodal experiences in which the visualization provides concurrent feedback on sound and movement, the sensorimotor couplings that occur during the initial trial and error experiences are gradually refined and consolidated into the mental schemas that realize the tight action perception couplings that underlie the interaction with the system (and the musical instrument). These mental schemas are composed of forward and inverse models [37.73]. Due to the tailor-made composition of the MPM's software (e.g., mapping, backgrounds), different activities can be designed in order to shape the development of these internal models [37.64]. For example, to train an inverse model, the system can provide learners with a target painting (e.g., a straight line that increases in stroke size). The learner then *copies* this while playing the instrument. Learners are then challenged to imagine the musical goal (play a long note with crescendo) and this would be shown in the model painting. To generate the appropriate motor command (e.g., blow harder) would be realized in the associated sensory outcome that would imitate the model. To train forward models, learners can be asked to perform specific tasks (e.g., play one note louder, increasing in volume), and the painting would provide information on the sound produced. Here the actual outcome (e.g., a somewhat curved line with increasing stroke size) might not concur with the expected outcome (e.g., a straight line with increasing stroke size), which will thus urge the player to adjust their motor commands (e.g., better breath support) in the next trial. Gradually, the learner will be able to predict the sensory outcome that follows a certain motor command with greater ease. Moreover, placing a visual representation of the desired and the actual outcome side by side on the screen can facilitate the interaction between the partnership of teacher and learner. Importantly, the MPM's visualization of music and movement goes beyond the mere provision of visual feedback. It invites users to creatively use music and movement in order to obtain a personalized outcome: namely, the digital *painting*. Accordingly, it allows a musician to reach out and engage with a new realm of experience, constituted by the combination of music, movement and visuals, and to discover novel ways of music making. It allows players to deliberately manipulate the self-generated visual feedback and, reinforced by straightforward mapping, stimulates implicit learning (because the focus is on the *painting* of auditory–motor couplings). Therefore, besides appealing to the motor adaptation mechanism to consciously guide music playing, the MPM may also enable a more accurate spontaneous motor adaptation of pre-established goals and artistic intentions. Furthermore, engaging with the system appeals to the human reward system. Next to supporting prediction mechanisms (based on the real-time visual feedback), the system could possibly facilitate the occurrence of an optimally pleasurable experience (flow experience, feeling of presence [37.69, 71]). Flow experience, which is a highly intrinsically rewarding state [37.74], has previously been related (mainly from a theoretical point of view) to the release of dopamine in the brain [37.75]. The MPM facilitates a broad range of activities (e.g., challenging games, creative painting) that foster surprise (*Hé, did you see that, I painted a . . .*) and injects added novelty into playing music, while at the same time, error prediction mechanisms are activated due to the system's mapping strategies. To conclude, the MPM has the potential to monitor, to motivate and to alter musicians' actions while playing a musical instrument. We believe this approach positively contributes to the development of genuine musicianship.

37.6.2 D-Jogger

D-Jogger is a technology that facilitates the synchronization of human movement to music [37.55, 56]. It is an example of a system that interrupts the human action–perception cycle (Fig. 37.3). It functions by spontaneously changing human activity at a subliminal (nonconscious) level. The technology has been developed primarily for walking and running applications, but it can also be coupled with biological rhythms, such as breathing and heart rate.

The overall objective of D-Jogger is to manipulate the tempo and phase of music such that it becomes possible to walk or run in synchrony with the beat of the music. D-Jogger can be understood as a technology that aligns two coupled rhythms: a musical rhythm with a movement rhythm (or a gait pattern). This alignment

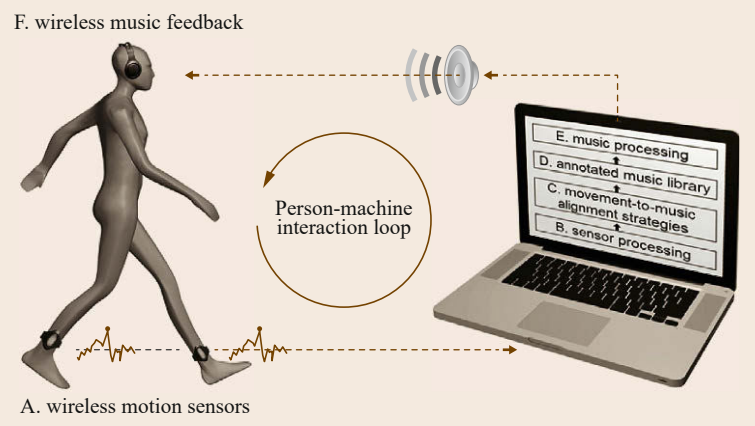


Fig. 37.3 Smart music player: person–machine interaction loop and the main components involved

is based on the timing of two salient moments of these rhythms: the musical beat and the footfall (i. e., the moment when the foot hits the ground). D-Jogger does this by monitoring the movement rhythm, using sensors (such as accelerometers or gyroscopes) that measure the timing of the gait pattern. Once the tempo is found, it looks for music which comes as close as possible to the tempo. Then it adapts the music tempo (by a stretching technique known as phase-vocoder) altering the music to match the gait tempo. The next step is then an adjustment of the phase such that beat and footfall can be precisely predicted.

D-Jogger intervenes in the human action–perception system, in particular with the internal models that predict the (perceptual) arrival of the musical beat, and thus the proper gait patterns are activated to allow for a match of the footfall with the beat. D-Jogger offers several options to control this match. One of the options is to ensure that the song starts in the same tempo as the tempo of the gait pattern, irrespective of whether the beat starts at the same moment of the footfall. For example, the music can have the same tempo of the gait, but the beat may not match the footfall. In that case it is likely that the human movement rhythm will be entrained and automatically adjust itself to match the beat. Apparently, internal forward models adjust the phase of the human movement rhythm at a subliminal level, which is an example of spontaneous motor adaptation. Interaction with D-Jogger has a strong rewarding effect because it facilitates synchronization. Being able to synchronize gives a feeling of control and satisfaction that is energizing and empowering.

37.7 Conclusion

Music is ubiquitous in people's daily lives. The accessibility of music and music services has grown exponentially in the past few years, and makes up an important part of a popular lifestyle. Music is a powerful medium because interaction with music taps into various cognitive, sensory, motor, and emotion regulation processes, and can lead to social interactions. In this chapter, we discussed the role that new electronic and digital technologies could play in reinforcing music-based interactions. In so doing, we have introduced a conceptual model that provides a framework upon which a vast number of music-based applications have already been developed, as well as offering a guide for future developments in this domain. The crux of the model is the idea that technologies function as extensions of the natural cognitive, perceptual, physical, physiological, and social dispositions and capabilities of human beings. Thus, technologies profoundly change and enrich people's experiences and interactions with their sensory and social environment, and help to develop people's social, cognitive, motor, and affective skills. The central mechanism in which music technologies can intervene is the coupled action–perception loop that is central to how people engage in a manifold of musical activities, such as music performance, music listening, and dance. In all of these activities, motor control and behavior, and sensory information processing become reciprocally linked. Further, we have discussed how this action–perception loop, and in particular the role of music-based interaction systems therein, further links to physiological processes related to reward, and to social communication and interaction. Following embodied theories on perception, cognition and behavior, we emphasized the role of the human body, and its actions, in

the design of music-based interaction systems. In that context, we have outlined three important perspectives that music-based systems may consider; monitoring actions, motivating actions, and altering actions. The last perspective is especially innovative, and it would be interesting to further explore this perspective in, for example, the context of interactive sonification systems.

This chapter was not intended as an overview of existing music-based technologies. Given the colossal number of realizations, this would be an impossible endeavor. Rather, we wanted to introduce a theoretical and conceptual perspective on the design of music-based interaction systems. Thus, we discussed some key concepts that could frame previous realizations, but more importantly, we hope that they can guide future developments in the field. Our discussion made clear that the design of music-based interaction systems is a highly complex task, as it incorporates knowledge and draws on expertise from several fields of practice and research. Alongside the engineering aspect, the design of music-based interaction systems must be informed by fundamental research on psychological, physiological, and social mechanisms, and incorporate viable educational methods. Therefore, the development of music technologies should ideally be realized in close collaboration with the sciences, the arts, and educational practices.

References

37.1 J. Phillips-Silver, P.E. Keller: Searching for roots of entrainment and joint action in early musical interactions, Front. Hum. Neurosci. **6**, 26 (2012)

37.2 A. D'Ausilio, G. Novembre, L. Fadiga, P.E. Keller: What can music tell us about social interaction?, Trends Cogn. Sci. **19**(3), 111–114 (2015)

37.3 J. Krueger: Affordances and the musically extended mind, Front. Psychol. **4**, 1003 (2013)

37.4 C.M. Conway, D.B. Pisoni, W.G. Kronenberger: The importance of sound for cognitive sequencing abilities the auditory scaffolding hypothesis, Curr. Dir. Psychol. Sci. **18**(5), 275–279 (2009)

37.5 B. Tillmann: Music and language perception: Expectations, structural integration, and cognitive sequencing, Topics Cogn. Sci. **4**(4), 568–584 (2012)

37.6 R.J. Bood, M. Nijssen, J. van der Kamp, M. Roerdink: The power of auditory-motor synchronization in sports: Enhancing running performance by coupling cadence with the right beats, PLoS ONE **8**(8), e70758 (2013)

37.7 C.I. Karageorghis, D.-L. Priest: Music in the exercise domain: A review and synthesis (Part I), Int. Rev. Sport Exerc. Psychol. **5**(1), 44–66 (2012)

37.8 A. Moussard, E. Bigand, S. Belleville, I. Peretz: Music as a mnemonic to learn gesture sequences in normal aging and Alzheimer's disease, Front. Hum. Neurosci. **8**, 294 (2014)

37.9 L. Nijs, M. Lesaffre, M. Leman: The musical instrument as a natural extension of the musician. In: *Music and Its Instruments*, ed. by M. Castellengo, H. Genevois (Editions Delatour France, Sampzon 2013)

37.10 W.I. Bauer: *Music Learning Today: Digital Pedagogy for Creating, Performing, and Responding to Music* (Oxford Univ. Press, Oxford 2014)

37.11 A. Brown: *Music Technology and Education: Amplifying Musicality* (Routledge, New York 2014)

37.12 J. Dorfman: *Theory and Practice of Technology-Based Music Instruction* (Oxford Univ. Press, Oxford 2013)

37.13 K.O. Hara, B. Brown: *Consuming Music Together: Social and Collaborative Aspects of Music Consumption Technologies* (Springer, Berlin, Heidelberg 2006)

37.14 M. Leman: *Embodied Music Cognition and Mediation Technology* (MIT Press, Cambridge 2007)

37.15 J.A. Fodor: *The Modularity of Mind: An Essay on Faculty Psychology* (MIT Press, Cambridge 1983)

37.16 A.K. Engel, A. Maye, M. Kurthen, P. König: Where's the action? The pragmatic turn in cognitive science, Trends Cogn. Sci. **17**(5), 202–209 (2013)

37.17 A. Clark, D. Chalmers: The extended mind, Analysis **58**(1), 7–19 (1998)

37.18 C. Addyman, R.M. French, D. Mareschal, E. Thomas: Learning to perceive time: A connectionist, memory-decay model of the development of interval timing in infants. In: *Proc. 33rd Annu. Conf. Cogn. Sci. Soc. (COGSCI)* (2011)

37.19 P. Craenen: Instruments for new ears, ISCM World Music Mag. **22**, 90–99 (2012)

37.20 M. Leman, L. Nijs: Music cognition and technology-enhanced learning for music playing. In: *The Routledge Companion to Music, Technology & Education*, ed. by A. King, A. Ruthmann, E. Himonides (Routledge, New York 2015)

37.21 F. Bevilacqua, B. Zamborlin, A. Sypniewski, N. Schnell, F. Guédy, N. Rasamimanana: Continuous realtime gesture following and recognition. In: *Embodied Communication and Human-Computer Interaction*, Vol. 5934, ed. by S. Kopp, I. Wachsmuth (Springer, Berlin, Heidelberg 2010)

37.22 P.-J. Maes, M. Leman, C. Palmer, M.M. Wanderley: Action-based effects on music perception, Front. Psychol. **4**, 1008 (2013)

37.23 G. Knoblich, N. Sebanz: The social nature of perception and action, Curr. Dir. Psychol. Sci. **15**(3), 99–104 (2006)

37.24 W. Schultz: The reward signal of midbrain dopamine neurons, Physiology **14**(6), 249–255 (1999)

37.25 R.A. Wise: Dopamine, learning and motivation, Nat. Rev. Neurosci. **5**(6), 483–494 (2004)
37.26 P.C. Fletcher, J.M. Anderson, D.R. Shanks, R. Honey, T.A. Carpenter, T. Donovan, E.T. Bullmore: Responses of human frontal cortex to surprising events are predicted by formal associative learning theory, Nature Neurosci. **4**(10), 1043–1048 (2001)
37.27 M.V. Puig, J. Rose, R. Schmidt, N. Freund: Dopamine modulation of learning and memory in the prefrontal cortex: Insights from studies in primates, rodents, and birds, Front. Neural Circuits **8**, 93 (2014)
37.28 W. Schultz: Behavioral dopamine signals, Trends Neurosci. **30**(5), 203–210 (2007)
37.29 E.E. Steinberg, R. Keiflin, J.R. Boivin, I.B. Witten, K. Deisseroth, P.H. Janak: A causal link between prediction errors, dopamine neurons and learning, Nat. Neurosci. **16**(7), 966–973 (2013)
37.30 P. Waelti, A. Dickinson, W. Schultz: Dopamine responses comply with basic assumptions of formal learning theory, Nature **412**(6842), 43–48 (2001)
37.31 T.E. Hazy, M.J. Frank, R.C. O'Reilly: Neural mechanisms of acquired phasic dopamine responses in learning, Neurosci. Biobehav. Rev. **34**(5), 701–720 (2010)
37.32 W. Schultz: Predictive reward signal of dopamine neurons, J. Neurophysiol. **80**(1), 1–27 (1998)
37.33 R.J. Zatorre, V.N. Salimpoor: From perception to pleasure: Music and its neural substrates, Proc. Natl. Acad. Sci. **110**(Supplement 2), 10430–10437 (2013)
37.34 I. Peretz, M. Coltheart: Modularity of music processing, Nat. Neurosci. **6**(7), 688–691 (2003)
37.35 D.B. Huron: *Sweet Anticipation: Music and the Psychology of Expectation* (MIT Press, Cambridge 2006)
37.36 P. Vuust, M.L. Kringelbach: The pleasure of making sense of music, Interdiscip. Sci. Rev. **35**(2), 166–182 (2010)
37.37 B. Hommel: Acquisition and control of voluntary action. In: *Voluntary Action: Brains, Minds, and Sociality*, ed. by S. Maasen, W. Prinz, G. Roth (Oxford Univ. Press, Oxford 2003) pp. 34–48
37.38 G. Rizzolatti, L. Craighero: Mirror neuron: A neurological approach to empathy. In: *Neurobiology of Human Values*, ed. by J.-P.P. Changeux, A. Damasio, W. Singer (Springer, Berlin, Heidelberg 2005) pp. 107–123
37.39 N. Bryan-Kinns: Mutual engagement in social music making. In: *Intelligent Technologies for Interactive Entertainment*, ed. by A. Camurri, C. Costa (Springer, Berlin, Heidelberg 2012) pp. 260–266
37.40 N. Moran: Social implications arise in embodied music cognition research which can counter musicological "individualism", Front. Psychol. **5**, 676 (2014)
37.41 L. Badino, A. D'Ausilio, D. Glowinski, A. Camurri, L. Fadiga: Sensorimotor communication in professional quartets, Neuropsychologia **55**, 98–104 (2014)
37.42 D. Glowinski, M. Mancini, A. Camurri: Studying the effect of creative joint action on musicians' behavior. In: *Arts and Technology*, Vol. 116, ed. by G. De Michelis, F. Tisato, A. Bene, D. Bernini (Springer, Berlin, Heidelberg 2013) pp. 113–119
37.43 M. Leman, M. Demey, M. Lesaffre, L. van Noorden, D. Moelants: Concepts, technology, and assessment of the social music game 'Sync-in-Team'. In: *Proc. Int. Conf. Comput. Sci. Eng. (CSE)* (2009)
37.44 L. Shapiro: *The Routledge Handbook of Embodied Cognition* (Routledge, New York 2014)
37.45 G.F. Welch, D.M. Howard, E. Himonides, J. Brereton: Real-time feedback in the singing studio: An innovatory action-research project using new voice technology, Music Educ. Res. **7**(2), 225–249 (2005)
37.46 G. Riva: Enacting interactivity: The role of presence. In: *Enacting Intersubjectivity: A Cognitive and Social Perspective on the Study of Interactions*, ed. by F. Morganti, A. Carassa, G. Riva (IOS Press, Amsterdam 2008) pp. 97–114
37.47 G. Riva, G. Castelnuovo, F. Mantovani: Transformation of flow in rehabilitation: The role of advanced communication technologies, Behav. Res. Methods **38**(2), 237–244 (2006)
37.48 J. Russell: *Agency: Its Role in Mental Development* (Psychology, Hove 1996)
37.49 T. Metzinger: Phenomenal transparency and cognitive self-reference, Phenomenol. Cogn. Sci. **2**(4), 353–393 (2003)
37.50 G. Dubus, R. Bresin: A systematic review of mapping strategies for the sonification of physical quantities, PLoS ONE **8**(12), e82491 (2013)
37.51 R. Sigrist, G. Rauter, R. Riener, P. Wolf: Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review, Psychon. Bull. Rev. **20**(1), 21–53 (2013)
37.52 H. Haken, J.A.S. Kelso, H. Bunz: A theoretical model of phase transitions in human hand movements, Biol. Cybern. **51**(5), 347–356 (1985)
37.53 J.A.S. Kelso: Coordination dynamics. In: *Encyclopedia of Complexity and Systems Science*, ed. by R.A. Meyers (Springer, Berlin, Heidelberg 2009) pp. 1537–1565
37.54 O. Oullier, G.C. De Guzman, K.J. Jantzen, J. Lagarde, J.A.S. Kelso: Social coordination dynamics: Measuring human bonding, Soc. Neurosci. **3**(2), 178–192 (2008)
37.55 B. Moens, M. Leman: Alignment strategies for the entrainment of music and movement rhythms, Ann. N.Y. Acad. Sci. **1337**, 86–93 (2015)
37.56 B. Moens, C. Muller, L. van Noorden, M. Franek, B. Celie, J. Boone, M. Leman: Encouraging spontaneous synchronisation with D-Jogger, an adaptive music player that aligns movement and music, PLoS ONE **9**(12), e114234 (2014)
37.57 K. Friston, J. Kilner, L. Harrison: A free energy principle for the brain, J. Physiol. Paris **100**(1), 70–87 (2006)
37.58 M.I. Jordan, D.E. Rumelhart: Forward models: Supervised learning with a distal teacher, Cogn. Sci. **16**(3), 307–354 (1992)
37.59 J.W. Krakauer, P. Mazzoni: Human sensorimotor learning: Adaptation, skill, and beyond, Curr. Opin. Neurobiol. **21**(4), 636–644 (2011)

37.60 H. Lalazar, E. Vaadia: Neural basis of sensorimotor learning: Modifying internal models, Curr. Opin. Neurobiol. **18**(6), 573–581 (2008)
37.61 R. Shadmehr, M.A. Smith, J.W. Krakauer: Error correction, sensory prediction, and adaptation in motor control, Annu. Rev. Neurosci. **33**, 89–108 (2010)
37.62 M.C.M. van der Steen, P.E. Keller: The ADaptation and Anticipation Model (ADAM) of sensorimotor synchronization, Front. Hum. Neurosci. **7**, 253 (2013)
37.63 D.M. Wolpert, Z. Ghahramani, M.I. Jordan: An internal model for sensorimotor integration, Sci. New Ser. **269**(5232), 1880–1882 (1995)
37.64 L. Nijs, M. Leman: Interactive technologies in the instrumental music classroom: A longitudinal study with the Music Paint Machine, Comput. Educ. **73**, 40–59 (2014)
37.65 W. Hsu, M. Sosnick: Evaluating interactive music systems: An HCI approach. In: *Proc. Int. Conf. New Interf. Music. Expr. (NIME)* (2009)
37.66 C. Kiefer, N. Collins, G. Fitzpatrick: HCI methodology for evaluating musical controllers: A case study. In: *Proc. Int. Conf. New Interf. Music. Expr. (NIME)* (2008)
37.67 M.M. Wanderley, N. Orio: Evaluation of input devices for musical expression: Borrowing tools from HCI, Comput. Music J. **26**(3), 62–76 (2002)
37.68 A.R. Addessi, F. Ferrari, F. Carugati: Observing and measuring the flow emotional state in children interacting with the MIROR platform. In: *12th Int. Conf. Music Percept. Cogn. (ICMPC)/8th Trienn. Conf. Eur. Soc. Cogn. Sci. Music (ESCOM), Thessaloniki* (2012)
37.69 L. Nijs, P. Coussement, B. Moens, D. Amelinck, M. Lesaffre, M. Leman: Interacting with the Music Paint Machine: Relating the constructs of flow experience and presence, Interact. Comput. **24**(4), 237–250 (2012)
37.70 N. Collins, J. d'Escriván: *The Cambridge Companion to Electronic Music* (Cambridge Univ. Press, Cambridge 2007)
37.71 L. Nijs, B. Moens, M. Lesaffre, M. Leman: The Music Paint Machine: Stimulating self-monitoring through the generation of creative visual output using a technology-enhanced learning tool, J. New Music Res. **41**(1), 79–101 (2012)
37.72 L. Nijs, M. Leman: Performing with the Music Paint Machine: Provoking an embodied approach to educational technology. In: *Music, Technology & Education: Critical Perspectives*, ed. by A. King, E. Himonides (Routledge, London 2014)
37.73 G. Pezzulo: Grounding procedural and declarative knowledge in sensorimotor anticipation, Mind Lang. **26**(1), 78–114 (2011)
37.74 M. Csikszentmihalyi: *Flow: The Psychology of Optimal Experience*, Vol. 41 (Harper Perennial, New York 1991)
37.75 C. Peifer: Psychophysiological correlates of flow-experience. In: *Advances in Flow Research*, ed. by S. Engeser (Springer, Berlin, Heidelberg 2012) pp. 139–164

38. Methods for Studying Music-Related Body Motion

Alexander Refsum Jensenius

This chapter presents an overview of some methodological approaches and technologies that can be used in the study of music-related body motion. The aim is not to cover all possible approaches, but rather to highlight some of the ones that are more relevant from a musicological point of view. This includes methods for video-based and sensor-based motion analyses, both qualitative and quantitative. It also includes discussions of the strengths and weaknesses of the different methods, and reflections on how the methods can be used in connection to other data in question, such as physiological or neurological data, symbolic notation, sound recordings and contextual data.

38.1 Some Key Challenges

The last decades have seen a rapid growth of interest in studying *music-related body motion*, that is, all types of human motion that appear in a musical context [38.1–4]. This includes the motion carried out by *performers*, such as musicians, conductors and dancers, and the motion of *perceivers*, such as that of audience members during concerts, people dancing at clubs, or people's spontaneous motion to music in everyday life. As such, music-related motion is a functionally diverse category, ranging from describing purely instrumental motion (such as hitting a piano key with a finger) to purely communicative motion (such as gesticulating in the air with the arms). Furthermore, music-related motion may occur in any type of location, for example in a concert hall, at home, or in the street.

The challenge for musicologists interested in studying motion as part of their empirical material is to choose methods that allow for studying such motion in a systematic manner. However, before deciding on a methodological approach, it is important to properly evaluate the content and context in which the motion is to be studied. Some questions to consider are:

- Aim: why is music-related motion interesting in this study? What kind of interaction is planned (sound–human, human–sound, human–human)?
- Subjects: how many subjects will be studied? What is their demography (gender, age, music/motor abilities) and personal context (familiar/unfamiliar with the task)? Will they move individually or in groups? What is the social context of the study?
- Motion: what type of motion is expected, and in which parts of the body? Are they large or small? Are they slow or fast? Will the subjects be stationary, or will they move about? Is it necessary to find the absolute position in space, or is relative motion information (such as acceleration) sufficient?
- Environment: will the study be carried out in a *controlled* environment (such as a lab) or in an *eco-*

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_38

logical setting (such as a concert hall)? Is power available? How much time is there to set up equipment? What are the lighting conditions?
- Artifacts: will there be any instruments, tools or other types of technologies used in the setup, and how will they be captured and synchronized?
- Audio: what type of sound recording is needed? How many channels? What sampling frequency and bit rate? What is the necessary level of synchronization between motion and sound data?
- Video: what type of video recording is needed? How many cameras? What resolution and frame rate?
- Data handling: how will the different types of data be synchronized? What software and data formats will be used? What type of storage, backup, and sharing solutions are planned?
- Analysis: what type of analysis is planned? What types of visualizations are needed? What types of features will be extracted?

When it comes to the latter – analysis – there are numerous approaches to choose from, but they can be broadly categorized into two main categories:

- Descriptive analysis: the motion is described through *kinematics* (such as velocity or acceleration), *spatial features* (such as size and position in the room) or *temporal features* (such as frequency).
- Functional analysis: the motion is described through functional properties, such as whether the actions are sound-producing or sound-accompanying (see [38.5] for an overview of functional categories).

The former may often be associated with quantitative analysis approaches, such as using statistical methods on numerical data, while the latter may be based on qualitative analysis approaches. In most cases, however, one would typically need to carry out both descriptive and functional analyses, and utilize both qualitative and quantitative methods. As such, the methods should be seen as complimentary rather than competing.

This chapter will not focus on the analytical methods per se, but rather on methods that prepare the ground for such analyses to be carried out. We will begin by presenting a few methods for qualitative motion analysis, before moving to some quantitative approaches.

38.2 Qualitative Motion Analysis

There exist numerous systems for systematic notation, analysis and exploration of body motion from a qualitative and observational point of view, including the Alexander technique [38.6], Rolfing [38.7], expressive motion [38.8], Dalcroze [38.9] and Benesh [38.10]. Several of these systems were developed in parallel to, or influenced by, the work of Rudolf Laban (1879–1954). Here we will look more closely at two of Laban's methods: *Labanotation* and *Laban Movement Analysis* (LMA).

38.2.1 Labanotation

Rudolf Laban worked as a dancer and choreographer before he became interested in the description and analysis of human motion at large, with the aim of developing a universal system for motion analysis. For that reason he spent time studying all sorts of human motion, including that of factory workers. In 1928 he presented a system called *Schrifttanz* [38.11], a method for the notation of motion as symbols along a vertical axis. This system was later to become *Labanotation*.

When writing a motion score through Labanotation, it is common to start with a description of *motifs*. These motifs are reduced versions of a full score, and they provide a rapid approach to writing down the main elements of a motion sequence. This first sketch is later used to write a more detailed Labanotation based on a structured set of symbols allowing for the notation of any type of motion of any body part. Figure 38.1 shows examples of Labanotation, with time running vertically, from bottom to top. The vertical line in the center of each notation system marks the center of the body, and

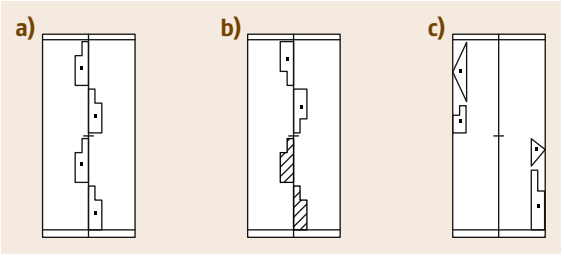


Fig. 38.1a–c Three short motion sequences written in Labanotation, with time running vertically from bottom to top: **(a)** normal walking forwards with right and left feet; **(b)** right and left feet moving forwards on their toes, followed by walking backwards; **(c)** right side of the body moves forwards and then to the right, followed by the left side of the body leaning forwards and then to the left

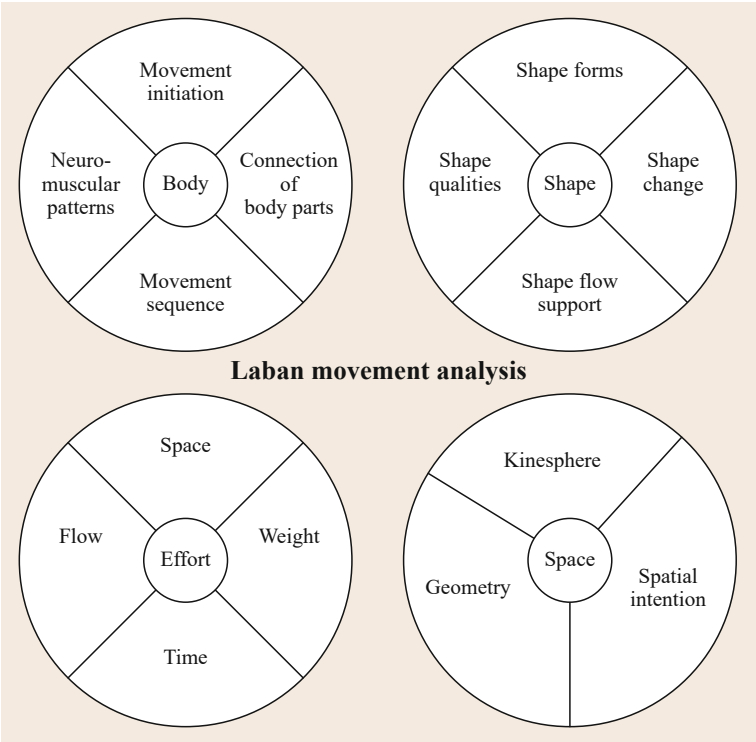


Fig. 38.2 Overview of all the categories in Laban Movement Analysis (LMA)

motion is notated through symbols on the left and right sides of the center line. Each of the body parts has their own symbol, and the duration of motion and position are given through different types of symbols inside the shapes of the body parts.

Although not very widespread, Labanotation is still in active use today, particularly in larger ballet and dance companies. There are also examples of the scholarly use of Labanotation, with an increasing interest from researchers within human–computer interaction and machine-assisted motion analysis [38.12]. However, the complexity of the system means that very few people really master Labanotation at a level with which they can read and write scores fluently. So, as opposed to music notation, which one can easily assume that most musicians and music researchers master well, it is hard to find people that can comfortably read and write Labanotation outside specialized circles.

38.2.2 Laban Movement Analysis

The theoretical basis for the Labanotation system was developed within a framework that has later been called *Laban Movement Analysis* (LMA). Even though Labanotation and LMA coexist, they can be used independently of each other. In fact, the LMA has received more widespread usage than Labanotation, also within the study of music.

As opposed to Labanotation's focus on writing motion structures in time, the LMA system is based on describing motion qualities. Fundamental to using the LMA is that the analysis should always start from the observer's point of view, and by asking the question *how does this motion feel from my own body?* [38.13]. As such, the LMA is by default subjective, even though it aims at being a general method for the observation of motion.

The LMA system is based on four main categories: *body*, *space*, *shape* and *effort*, each of which are subdivided into categories describing different motion qualities, such as outlined in Fig. 38.2. In this context we will mainly focus on one of the four main categories, the *effort* element [38.14], which has proven to be particularly relevant for musicological studies.

The effort category can be further divided into four subcategories, each with a descriptive axis:

- Space (direct–indirect). Space describes how one moves through the physical space, and how one relates to the body's *kinesphere*, the maximal volume we can reach around our body when standing fixed on the floor.

- Time (quick–sustained). Time is used to describe the rhythmic character of motion. Laban was concerned that time and rhythm should not be split. We experience rhythmical patterns all the time, and we have our own bodily rhythms defined by our pulse and breathing.
- Weight (strong–light). Weight is related to gravity, and the fact that we need to use muscular activity to work against gravity when moving upwards. We also use gravity to help us when moving downwards, and it is this constant interplay between our body and the Earth's gravitational pull that help shape our motion.
- Flow (bound–free). Flow describes how motion unfolds in time and space. Free-flow motion is both relaxed and continuous at the same time, while bound flow is when the motion is hindered.

Even though the four effort parameters are but one small part of the full LMA system, they have been used separately in several musicological studies, including sound–motion correlations in free dance to music [38.15], the phrasing motion of clarinetists [38.16], and relationships between music, emotion and dance [38.17].

38.3 Video-Based Analyses

If thinking about an axis from purely qualitative methods to purely quantitative, video-based analyses can be argued to cover quite a large part of the axis. For example, having a video recording that can be played back multiple times, and at various speeds, is very useful for carrying out LMA analysis. And, as we shall see in the next section, even a regular video recording can be used to extract meaningful quantitative motion data. Furthermore, it is also common to use video recordings as a reference when recording motion with sensor-based motion tracking technologies. In such cases, the video recording can be used to help the qualitative interpretation of numerical results.

For many applications, a regular video recording is often the easiest, fastest and cheapest solution to start working systematically with the study of music-related motion. Nowadays, everyone has access to high quality video cameras even in their mobile phones, and the cost of professional-quality video cameras is also within the reach for many. The main challenge for musicologists, then, is to record the video in a manner suitable for later analysis.

38.3.1 Recording Video for Analysis

One thing to bear in mind is that a video recording meant for analytical purposes is quite different from a video recording shot for documentary or artistic purposes. The latter type of video is usually based on the idea of creating an aesthetically pleasing result, which often includes continuous variation in the shots through changes in the lighting, background, zooming, panning, etc. A video recording for analysis, on the other hand, is quite the opposite: it is best to record it in a controlled studio or lab setting with as few camera changes as possible. This is to ensure that it is the *content* of the recording, that is, the human motion, which is in focus, not the motion of the camera or the environment.

Even though a controlled recording environment may be the best choice from a purely scientific point of view, it is possible to obtain useful recordings for analytical purposes also out in the field. This, however, requires some planning and attention to detail. Here are a few things to consider:

- Foreground/background: place the subject in front of a background that is as plain as possible, so it is possible to easily discern between the important and nonimportant elements in the image. For computer vision recordings it is particularly important to avoid backgrounds with moving objects, since these may influence the analysis.
- Lighting: avoid changing lights, as they will influence the final video. In dark locations, or if the lights are changing rapidly (such as in a disco or club concert), it may be worth recording with an infrared camera. Some consumer cameras come with a *night mode* that serves the same purpose. Even though the visual result of such recordings may be aesthetically unsatisfactory, they can still work well for computer-based motion analysis.
- Camera placement: place the camera on a tripod, and avoid moving the camera while recording. Both panning and zooming makes it more difficult to analyze the content of the recordings later. If both overview images and close-ups are needed, it is better to use two (or more) cameras to capture different parts of the scene in question.
- Image quality: it is always best to record at the highest possible spatial (number of pixels), temporal (frames per second) and compression (format and ratio) settings the camera allows for. However, the

Fig. 38.3 A *motion image* is created by subtracting subsequent frames in a video file (Frame (2) minus Frame (1))

Fig. 38.4 Individual motion history images of 14 separate percussion strokes allow for studying stroke heights and patterns. The images have been made by adding a motion history image on top of a picture of the scene, thus showing both motion features and the contextual information

most important is to find a balance between image quality, file size and processing time.

As mentioned earlier, a video recording can be used as the starting point for both qualitative and quantitative analysis. We will here look at a couple of different possibilities, moving from more qualitative visualization methods to advanced motion capture techniques.

38.3.2 Video Visualization

Videos can be watched as they are, but they can also be used to develop new visualizations to be used for analysis. The aim of creating such alternate displays from video recordings is to uncover features, structures and similarities within the material itself, and in relation to, for example, score material. Three useful visualization techniques here are *motion images*, *motion history images* and *motiongrams*.

Motion Images

One of the most common techniques when working with motion analysis from video files is to create a *motion image* by calculating the absolute pixel difference between subsequent frames in the video file (Fig. 38.3). The end result is an image where only the pixels that have changed between the frames are displayed. This can be interesting in itself, but motion images are also the starting point for many other video visualization and analysis techniques.

Motion History Images

Motion images only display the motion happening between two frames in a video file, but often it is desirable to visualize motion over a period of time, say a few seconds. This can be done through *motion history images* that display the temporal development of a motion sequence. There are numerous ways of creating such displays [38.18], but many of the most common techniques are based on adding several motion image frames together. The usefulness of motion history images depends on carefully selecting a time window that fits the content of the motion, as can be seen in the examples of percussion strokes in Fig. 38.4.

Motion history images have been used in various types of music analysis, such as in the study of music and dance [38.19], and are also often to be seen in visual arts and creative practice.

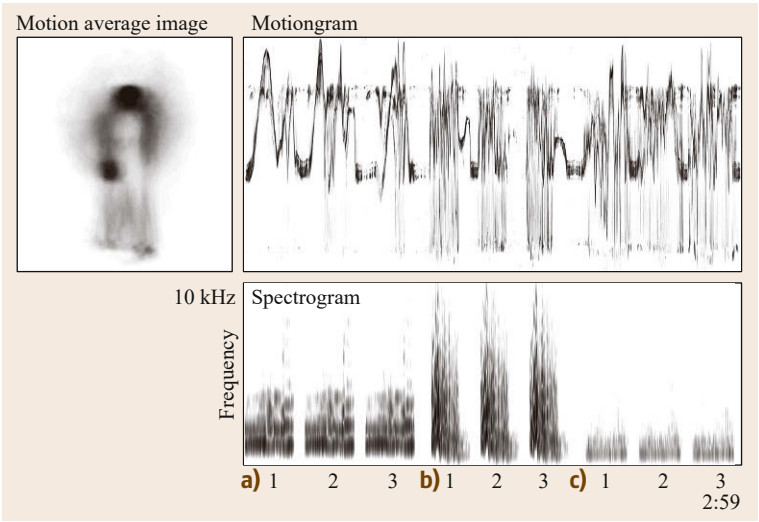


Fig. 38.5 A motion average image and a motiongram of a three-minute free-dance sequence to music make it possible to study both spatial and temporal features of a performance in connection to the spectrogram of the audio

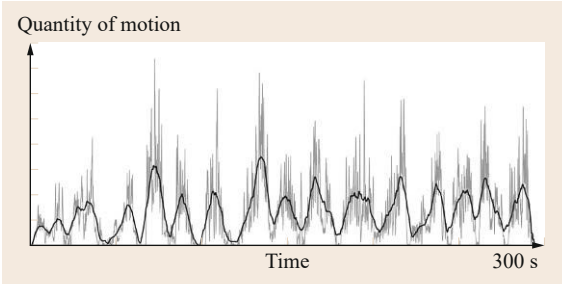


Fig. 38.6 A plot of the quantity of motion for a five-minute long dance sequence. The gray line is a plot of the tracked data, and the black line is a filtered version of the same dataset

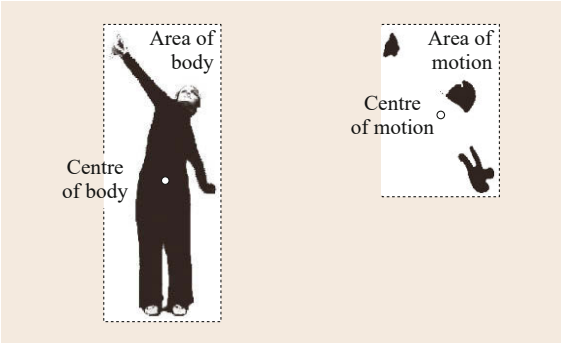


Fig. 38.7 Illustrations of the area and centroid of body and motion

Motiongrams

While a motion history image may reveal information about the spatial aspects of a motion sequence over a fairly short period of time, it is possible to use a *motiongram* to display longer sequences [38.20]. Figure 38.5 shows a motiongram created from a dance improvisation recording, and this display is created by plotting the normalized mean values of the rows of a series of motion images. The motiongram makes it possible to see both the location and quantity of motion of a video sequence over time, and is thus an efficient way of visualizing longer motion sequences.

A motiongram is only a reduced display of a series of motion images, with no analysis being done. It might help to think of the motiongram as a display of a collapsed series of pictures, or *stripes*, where each *stripe* summaries the content of a whole motion image.

Depending on the frame rate of the video file, motiongrams can be created from recordings as short as a few seconds to several hours. For short recordings it is possible to follow detailed parts of a body, particularly if there are relevant colors in the image, while motiongrams of longer recordings will mainly reveal larger sections of motion. Motiongrams work well together with audio spectrograms and other types of temporal displays such as graphs of motion or sound features.

38.3.3 Computer Vision

The broad field of *computer vision* (CV) is concerned with extracting useful information from video recordings. There is a lot of progress in the field, as summarized in [38.21–23], and we will here only look at a few possible methods.

Some basic motion features that are commonly used in music research are derived directly from the motion image. Since the motion image only shows pixels that have changed between the two last frames in a video sequence, the sum of the values of all these individual pixels will give an estimate of the *quantity of motion*

(QoM). Calculating the QoM for each frame will give a numeric series that can be plotted and used as an indicator of the activity, such as illustrated in the graph of a dance sequence in Fig. 38.6. Here it is possible to see where the dancer moved or stood still.

The *centroid of motion* (CoM) and *area of motion* (AoM) are other basic features that can easily be extracted from a motion image, and the differences between them are illustrated in Fig. 38.7. The CoM and AoM features can be used to illustrate *where* in an image the motion occurs as well as the spatial displacement of motion over time.

The field of computer vision has diverged into a number of different directions over the years. Most notably, there are now numerous methods available for tracking bodies and body parts in space, many of which are also being used in music interaction and analysis through tools such as EyesWeb [38.24], GEM for PureData [38.25] and Jitter for Max [38.26]. In addition to using regular video cameras for such analyses, there are several different types of specialized cameras that are made particularly for computer vision methods, including:

- Infrared cameras capturing only light in the infrared (nonvisible) range. Such cameras can be very useful in musical applications, since they work well also in dark spaces (such as a concert hall) or in locations with changing lights (for example club or stage lights).
- Time-of-flight cameras emitting a modulated acoustic signal and receiving the reflected signal from which it is possible to measure the time it took for the signal to return to the camera.
- Stereo cameras employing the same strategy as human vision, using two cameras next to each other (like our eyes), and then using the differences between the two images to estimate motion and rotation.

There are also examples of the combination of these three capturing methods, as well as the use of multiple cameras around the capture space. By combining multiple cameras distributed in space it is possible to create true three-dimensional recordings. This can now be done without markers on the body [38.27], but still the state-of-the-art when it comes to camera-based motion tracking are the marker-based systems using infrared cameras.

38.3.4 Infrared, Marker-Based Motion Capture

What many people refer to as *motion capture*, can more precisely be described as optical, infrared, marker-based systems. Such systems usually consist of at least six cameras positioned around the capture space (Fig. 38.8). Each of the cameras contains a ring of infrared light sources, and this infrared light is reflected on small markers and captured by the cameras. The system then calculates the exact position in space based on triangulating all the marker positions from each individual camera. The end result is a three-dimensional tracking of the markers in space, often captured at high speeds (more than 100 Hz) and at a high spatial resolution (in the range of millimeters). The captured points can be visualized directly or used as the basis for further analysis.

Fig. 38.8 Setup for motion capture of pianist Christina Kobb. In addition to the cameras placed in front, there are also cameras placed behind the pianist to capture the motion in three dimensions

An advantage of marker-based motion capture systems is that they allow for reliable tracking of the position (and sometimes orientation) of individual body joints. Furthermore, together with force plates and physiological sensors such motion capture systems may provide very precise and accurate data about the kinematics and kinetics of human motion.

There are, however, also some drawbacks of infrared, marker-based motion capture systems. The price tag is one, although such systems have become more affordable in recent years. More problematic from a user's perspective are the constraints enforced by the need of a controlled recording environment. Also, even though the markers used in such systems are small and lightweight, a recording session necessarily feels somewhat unnatural for the subjects. So when deciding on

whether to use such a system for a study, it is important to weigh the benefits of a high-quality motion recording against the obvious limitations of such an unecological setup.

38.4 Sensor-Based Motion Capture

Sensor-based systems are very different technologically to the camera-based solutions mentioned above. One main difference is that sensor systems are often modular, allowing for more flexibility when it comes to the types and numbers of *sensors* being connected to a sensor *interface*.

38.4.1 Sensor Interfaces

The sensor interface is the unit that digitizes the electrical signals coming from the sensors through an analog-to-digital converter (ADC), and which also contains the electronics needed for recording and/or connecting to a computer. Many sensors are analog and therefore have a theoretically infinite resolution, so the sensor interface is often the limiting factor of a digital sensor-based system. It is therefore important to choose a suitable sensor interface for the task [38.28], and some elements to consider are:

- Sampling rate: the speed at which the sensors are read, often ranging from a few to several thousands of samples per second.
- Bit rate: the resolution of each recorded sample, often ranging from 7-bit (like MIDI) to 10-bit or even higher.
- Connectivity: whether the interface is cabled or wireless, and the types of wired (USB, ethernet, etc.) or wireless (Wi-Fi, Bluetooth, etc.) connection being used.
- Power: whether the system needs external power or can run on batteries.
- Expandability: how many sensors that can be connected at the same time, and with what types of connectors.
- Size: anything from the smallest imaginable to fairly large-scale devices, typically dependent on all of the above together with the price.

Fortunately, sensor interfaces are constantly becoming smaller, faster, cheaper and more reliable, so there are numerous solutions to choose from, both ready-made and modular systems [38.28].

When it comes to the sensors themselves, there are many ways of classifying and evaluating these [38.29–32]. The presentation below follows the classification of tracking technologies used by *Bishop* et al. [38.33], in which the different systems are sorted according to the physical medium of the technology: *acoustic*, *mechanical*, *magnetic*, *inertial* and *electrical*.

38.4.2 Acoustic Tracking

Motion capture systems based on acoustic sensing use the principle of time-of-flight mentioned above, which measures the round-trip an acoustic signal takes from being transmitted, reflected on an object and received back at the sensor. In the same way as for infrared, marker-based systems, it is necessary to have multiple transmitter/receiver pairs to be able to estimate the three-dimensional position of an object through triangulation of the individual points.

Acoustic sensor systems usually work in the ultrasonic range, and can therefore be used unobtrusively for music-related applications. A challenge, however, is that such systems are often less accurate and precise for the spatial range of music-related motion. Thus acoustic tracking has received relatively little attention in music research, although there are some examples of its application in new electronic instruments [38.34, 35].

There are also some examples of the use of audible sound for motion tracking. One example is the walking experiment presented in [38.36], in which microphones were placed in the socks of participants, and periodicity peaks from the waveforms' amplitudes were used to measure the frequency of walking in relation to music.

38.4.3 Mechanical Tracking

Mechanical tracking systems are based on measuring angles, distances or forces, using potentiometers or sensors measuring flexing, stretching or force (Fig. 38.9). One advantage with mechanical tracking is that the sensors can easily be added to or embedded in suits, shoes and gloves, thus allowing for fairly unobtrusive tracking of body motion. Due to this versatility, mechanical systems have been popular for a lot of music applications, both for analysis and particularly for various types of electronic instruments and in interactive performance systems [38.37].

38.4.4 Magnetic Sensors

Motion capture systems using magnetic sensing are based on the idea of measuring disturbances in the mag-

Fig. 38.9 Examples of small and large force and bend sensors (*left*) and a custom-built glove with similar sensors built in (*right*)

netic field around the sensor. The perhaps simplest and cheapest of such sensors, *magnetometers*, are similar to a compass and measure the direction and strength of the Earth's magnetic field. Magnetometers are widely used in combination with inertial sensors (Sect. 38.4.5).

Active magnetic systems, on the other hand, are based on setting up a magnetic field around an electromagnetic source. Then one or more sensors can be used to measure the position of an object through sensing the induced electric current at a point in space (Fig. 38.10).

One advantage with magnetic tracking is that such systems can often track the three-dimensional orientation of an object (with each of the rotation axes often being referred to as *yaw*, *pitch* and *roll*) in addition to the three-dimensional position (with the axes being called *X*, *Y*, *Z*). Furthermore, magnetic tracking is generally quite accurate and precise, and allows for high sampling rates. Since such systems are cable-based, there are no problems related to occlusion of markers such as for camera-based systems.

Having to deal with cables is also one of the main negative aspects of magnetic systems, since the cables severely limit the physical range of the tracking. The perhaps biggest challenge, still, is that the systems are suspect to interference from ferromagnetic objects within the magnetic field [38.38]. This can be highly problematic in a musical context, since many musical instruments, including pianos, are constructed fully or partly with metal components.

Magnetic sensing has been used for a number of musical applications, both performance [38.39–41] and analysis [38.42–44]. However, with advancements in other tracking types, and particularly that of optical systems, the use of magnetic sensing has been in decline in recent years.

38.4.5 Inertial Sensors

Besides camera-based systems, *inertial* sensing is currently one of the most popular types of motion tracking.

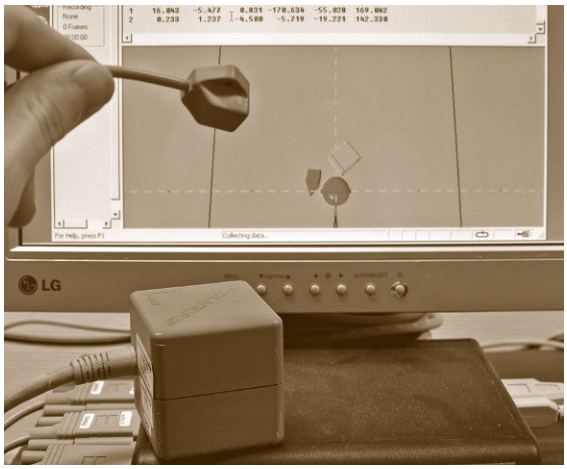


Fig. 38.10 One six-dimensional sensor (held in the hand) and the electromagnetic transmitter (on table) from a Polhemus Liberty magnetic tracking system. The image on the screen displays the position and orientation of the object in real time

The two main types of inertial sensors are *accelerometers* and *gyroscopes*, and both of these sensor types are based on measuring the displacement of a small *proof mass* inside the sensor. Accelerometers measure the positional displacement of an object, while gyroscopes the rotational. By combining three accelerometers and three gyroscopes it is possible to capture both three-dimensional position and three-dimensional rotation in one sensor unit.

One of the most compelling features of inertial sensors is that they rely on physical laws (gravity), which are not affected by external factors such as ferromagnetic objects or lighting. Since they do not contain any transmitters (like infrared cameras and acoustic and magnetic sensors) they can also be embedded in very small and self-contained units (Fig. 38.11), with low power consumption, and high sampling rates. These are probably some of the reasons why inertial sensors are now becoming integrated in a lot of technologies, further driving down the cost of single units and securing even broader integration in all sorts of electronic devices.

The downside to inertial sensing is that accelerometers do not measure the *position* of objects, but rather the rate of change of the objects. It is possible to estimate the position through integration, and, combined with the data from gyroscopes and magnetometers, this can lead to satisfactory results [38.45]. However, while the relative position estimates may be good, such position data often suffer from a considerable amount of drift [38.46]. One way to overcome some of the

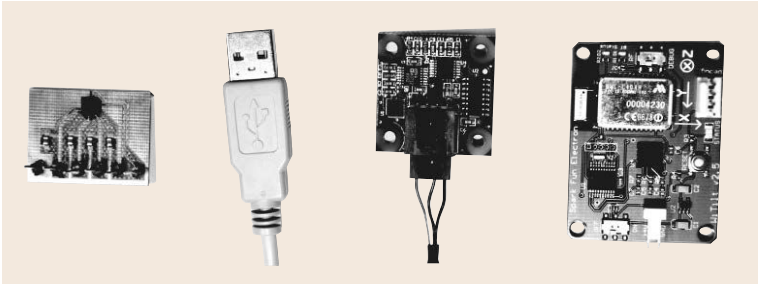


Fig. 38.11 Different types of home-made and commercial inertial sensing systems

Fig. 38.12 Yago de Quay performing with the Xsens MVN suit in Oslo, Norway (after [38.48])

Fig. 38.13 An example of music interaction with the commercially available Myo armband, which contains eight EMG sensors, accelerometers and gyroscopes (Photo: Kristian Nymoen)

drift problems is by combining inertial sensors with one or more of the other sensing solutions mentioned above [38.47]. One example of this can be found in the commercially available motion capture suits from Xsens, that combine a number of accelerometers, gyroscopes and magnetometers with a kinematic model of the body (Fig. 38.12).

38.4.6 Electrical Sensors

Sensors measuring the electrical current of *biosignals* have become increasingly popular for musical applications over the last decades [38.49]. Many such sensors, often also called *physiological* sensors, share the same sensing principle but are optimized to detect different types of biosignals:

- *Electromyograms* (EMGs) are used to measure muscle activity, and are particularly effective on the arms of musicians to pick up information about hand and finger motion [38.50]. Figure 38.13 shows an example of the commercially available Myo armband, with integrated EMG sensors, used for musical interaction [38.51].
- *Electrocardiograms* (ECGs) measure the electrical pulses from the heart, and can be used to extract information about heart rate and heart rate variability. The latter has been shown to correlate with emotional state [38.52].
- *Galvanic skin response* (GSR) refers to changes in skin conductance, and is often used on the fingers. The GSR signal is highly correlated with emotional changes, and such sensors have been used to some extent in music research [38.53, 54] as well as in music performance [38.49, 55]. However, the signals may not be entirely straightforward to interpret, and elements like sweat may become an issue when worn for longer periods of time.

- *Electroencephalograms* (EEGs) are used to measure electrical pulses from the brain, using either a few sensors placed on the forehead, or hats with numerous sensors included. Due to the weak brain signals, such sensors need to have strong amplifiers and are therefore also subject to a lot of interference and noise. Nevertheless, such sensors have also been applied in both music analysis and performance [38.56].

38.5 Synchronization and Storage

As mentioned in the introduction, some important challenges for music researchers working with any of the above-mentioned motion tracking solutions are to ensure synchronization with other types of data and to store the data in a structured way [38.57]. We will here briefly look into some possible solutions to these problems.

38.5.1 Motion Data Formats and Protocols

A number of formats exist for storing motion data, most of which were designed to solve the needs of a specific hardware system and/or research problem. The BRD format, for example, is used with electromagnetic trackers from Flock of Birds, while the C3D format is used for infrared motion capture systems from Vicon [38.58]. Several formats have also emerged for using motion capture data in animation tools, such as the BVA and BVH formats from Biovision, and the ASF and AMC formats from Acclaim, as well as formats used by animation software, such as the CSM format used by 3D Studio Max. There have been several attempts at creating XML-based (XML) standards for motion capture and animation data, such as the Motion Capture Markup Language (MCML) [38.59] and Multimodal Presentation Markup Language (MPML) [38.60]. But none of these nor the structures included in the MPEG-4 [38.61] and MPEG-7 [38.62] formats seem to have achieved widespread usage. Of all these, the C3D format seems to be the one that is supported by most software, although its implementation and number of features varies somewhat between different software solutions.

Since none of the other formats available solve the problems of music researchers, there have been several initiatives to create new formats and standards specifically targeted at music applications. The Gesture Motion Signal (GMS) format [38.63] is a binary format based on the Interchange File Format (IFF) standard [38.64], and was mainly developed for storing raw motion data. The Gesture Description Interchange File Format (GDIF) was proposed as a structure for handling everything from raw motion data to higher-level descriptors within other data formats [38.65], and has been successfully used within namespaces for the Open Sound Control (OSC) protocol [38.66] and as an extension to the Sound Description Interchange Format (SDIF) [38.67]. The Performance Markup Language (PML) was developed as an extension to the Music Encoding Initiative (MEI) [38.68] and focused on creating a structured approach to annotate performance data in relation to musical notation.

All of these formats and protocols solve some problems, even though none of them have emerged as de facto standards within the music research community.

38.5.2 Structuring Multimodal Data

Besides choosing a format for handling data, there are numerous conceptual and practical issues to deal with when working with music-related motion data. Different hardware devices use different protocols, formats and standards. They also work at different sampling rates, bit rates, with different numbers of sensors, different degrees of freedom, etc. Adding to this is the need to synchronize and store such data together with other relevant data, including MIDI data, audio and video files, as well as qualitative descriptors and various layers of analysis.

Fortunately, several software solutions have been developed for multimodal data acquisition and analysis, including the EyesWeb platform [38.69, 70], the MoCap Toolbox for Matlab [38.71], and various toolboxes for the graphical programming environment Max [38.44, 72]. There are also initiatives for creating online repositories for multimodal data storage and analysis, such as Repovizz [38.73]. In addition to helping in structuring and storing data, such systems also allow for carrying out collaborative and comparative studies on the same material.

38.6 Conclusion

In this chapter we have looked at a number of analytical approaches and technologies available for studying music-related motion, ranging from qualitative to quantitative, cheap to expensive, small to large, and simple to advanced. To simplify, we may differentiate between two main types of technologies used for motion tracking:

- Camera-based systems: this includes cameras of any sorts, recording in gray-scale, color or infrared, and recording with or without markers on the body.
- Sensor-based systems: this includes all sorts of sensors, including acoustic, mechanical, magnetic, inertial and electrical.

It is impossible to give one answer to what type of method or technology to use; they all have their strengths and weaknesses. For that reason it is important to decide on the right analysis method for the research question at hand. Perhaps the most decisive factor is whether to work in a laboratory setting or in a more ecological setting, say a concert hall, as this will to a large extent guide which methods and tools to use.

More specialized motion capture solutions, like infrared, marker-based systems, electromagnetic systems and various types of inertial sensors, often provide high recording speeds, and high spatial accuracy and precision, but they also come with several drawbacks. Price is one, although such systems have quickly become more affordable. More problematic is that the person being studied has to wear markers/sensors on the body, something that may be both alienating to the performer and obtrusive to the motion being performed. This is particularly problematic when using electromagnetic and mechanical systems with fairly large and heavy sensors and cables. But even the lightweight, reflective markers typically used in optical infrared systems are noticed by performers, albeit to a lesser degree.

Perhaps the biggest challenge with the larger systems is the nonecological setting they require. Small inertial systems, on the other hand, can be mobile, wireless, and more or less invisible, although the physical sensors need to be placed on the body. So if the aim is to study musicians in a real-world concert situation, regular video recordings may be the only realistic solution.

Apart from selecting the right technology, however, it is important to have a clear plan for what to record, have a structured approach to synchronizing the recordings with other data and media and to store the data using formats and protocols that maximize the potential for a diverse range of analytical approaches.

References

38.1 M.M. Wanderley, M. Battier (Eds.): *Trends in Gestural Control of Music* (IRCAM – Centre Pompidou, Paris 2000)

38.2 A. Gritten, E. King (Eds.): *Music and gesture* (Ashgate, Hampshire 2006)

38.3 A. Gritten, E. King (Eds.): *New Perspectives on Music and Gesture* (Ashgate, Hampshire 2011)

38.4 R.I. Godoy, M. Leman (Eds.): *Musical Gestures: Sound, Movement, and Meaning* (Routledge, New York 2010)

38.5 A.R. Jensenius, M.M. Wanderley, R.I. Godøy, M. Leman: Musical gestures: Concepts and methods in research. In: *Musical Gestures: Sound, Movement, and Meaning*, ed. by R.I. Godøy, M. Leman (Routledge, New York 2010) pp. 12–35

38.6 W. Barlow: *Alexander-princippet* (Borgen forlag, Copenhagen 1975)

38.7 R. Feitis: *Ida Rolf Talks about Rolfing and Physical Reality* (Harper and Row, New York 1978)

38.8 A. Pierce, R. Pierce: *Expressive Movement: Posture and Action in Daily Life, Sports, and the Performing Arts* (Perseus, Cambridge 1989)

38.9 E. Findlay: *Rhythm and Movement – Applications of Dalcroze Eurhythmics* (Summy-Birchard, Miami 1971)

38.10 M. Parker: *Benesh Movement Notation for Ballet* (Royal Academy of Dance, London 1996)

38.11 A.H. Guest: *Labanotation* (Routledge, New York 2004)

38.12 W. Choensawat, M. Nakamura, K. Hachimura: GenLaban: A tool for generating Labanotation from motion capture data, Multimed. Tools Appl. **74**(23), 10823–10846 (2014)

38.13 C.A. Schrader: *A Sense of Dance: Exploring Your Movement Potential* (Human Kinetics, Champaign 2004)

38.14 R. Laban, F.C. Lawrence: *Effort* (Macdonald Evans, London 1947)

38.15 E. Haga: *Correspondences Between Music and Body Movement*, Ph.D. Thesis (University of Oslo, Oslo 2008)

38.16 L. Campbell, M. Wanderley: *The Observation of Movement*, MUMT 609 Report (McGill University, Montreal 2005)

38.17 E. Van Dyck, P.-J. Maes, J. Hargreaves, M. Lesaffre, M. Leman: Expressing induced emotions through free dance movement, J. Nonverbal Behav. **37**(3), 175–190 (2013)

38.18 M.A.R. Ahad, J.K. Tan, H. Kim, S. Ishikawa: Motion history image: Its variants and applications, Mach.

Vis. Appl. **23**(2), 255–281 (2012)
38.19 A. Camurri, I. Lagerlöf, G. Volpe: Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques, Appl. Affect. Comput. Hum.-Comput. Interact. **59**(1/2), 213–225 (2003)
38.20 A.R. Jensenius: Some video abstraction techniques for displaying body movement in analysis and performance, Leonardo **46**(1), 53–60 (2013)
38.21 T.B. Moeslund, E. Granum: A survey of computer vision-based human motion capture, Comput. Vis. Image Underst. **81**(3), 231–268 (2001)
38.22 T.B. Moeslund, A. Hilton, V. Krüger: A survey of advances in vision-based human motion capture and analysis, Comput. Vis. Image Underst. **104**(2/3), 90–126 (2006)
38.23 S.S. Rautaray, A. Agrawal: Vision based hand gesture recognition for human computer interaction: A survey, Artif. Intell. Rev. **43**(1), 1–54 (2015)
38.24 A. Camurri, B. Mazzarino, G. Volpe: Analysis of expressive gesture: The EyesWeb expressive gesture processing library. In: *Gesture-based Communication in Human-Computer Interaction*, Lecture Notes in Computer Science, Vol. 2915, ed. by A. Camurri, G. Volpe (Springer, Berlin, Heidelberg 2004) pp. 460–467
38.25 J.M. Zmölnig: Gem for pd – Recent progress. In: *Proc. Int. Comput. Music Conf.*, *Miami* (2004)
38.26 G. Levin: Computer vision for artists and designers: Pedagogic tools and techniques for novice programmers, AI Society **20**(4), 462–482 (2006)
38.27 L. Sigal, A. Balan, M. Black: Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion, Int. J. Comput. Vis. **87**(1), 4–27 (2010)
38.28 M.M. Wanderley, D. Birnbaum, J. Malloch, E. Sinyor, J. Boissinot: SensorWiki.org: A collaborative resource for researchers and interface designers. In: *Proc. Int. Conf. New Interfaces Music. Expr.*, *Paris* (2006) pp. 180–183
38.29 R. Begg, M. Palaniswami: *Computational intelligence for movement sciences: Neural networks and other emerging techniques* (IGI Global, Hershey 2006)
38.30 H. Zhou, H. Hu: Human motion tracking for rehabilitation – A survey, Biomed. Signal Process. Control **3**(1), 1–18 (2008)
38.31 W.M. Richard: A sensor classification scheme, IEEE Trans. Ultrason. Ferroelectr. Freq. Control **UFFC-34**(2), 124–126 (1987)
38.32 S. Patel, H. Park, P. Bonato, L. Chan, M. Rodgers: A review of wearable sensors and systems with application in rehabilitation, J. NeuroEng. Rehabil. **9**(1), 21 (2012)
38.33 G. Bishop, G. Welch, B.D. Allen: Tracking: Beyond 15 minutes of thought. In: *SIGGRAPH Course 11* (ACM, Los Angeles 2001) pp. 6–11
38.34 F. Vogt, G. Mccaig, M.A. Ali, S.S. Fels: Tongue 'n' groove: An ultrasound based music controller. In: *Proc. Int. Conf. New Interfaces Music. Expr.*, *Dublin* (2002) pp. 181–185
38.35 M. Ciglar: An ultrasound based instrument generating audible and tactile sound. In: *Proc. Int. Conf. New Interfaces Music. Expr* (2010) pp. 19–22
38.36 F. Styns, L. van Noorden, D. Moelants, M. Leman: Walking on music, Hum. Mov. Sci. **26**(5), 769–785 (2007)
38.37 E.R. Miranda, M.M. Wanderley: *New Digital Musical Instruments: Control and Interaction Beyond the Keyboard* (A-R Editions, Middleton 2006)
38.38 G. Viglensoni, M.M. Wanderley: A quantitative comparison of position trackers for the development of a touch-less musical interface. In: *Proc. Int. Conf. New Interfaces Music. Expr.*, *Ann Arbor* (2012)
38.39 T. Marrin, R. Picard: The 'Conductor's Jacket': A device for recording expressive musical gestures. In: *Proc. Int. Comput. Music Conf* (1998) pp. 215–219
38.40 E. Lin, P. Wu: Jam Master, a music composing interface. In: *Proc. Hum. Interface Technol.*, *Vancouver* (2000) pp. 21–28
38.41 M.T. Marshall, J. Malloch, M.M. Wanderley: Gesture control of spatialization. In: *7th Int. Workshop Gesture Human-Comput. Interact. Simul.*, *Lisbon* (2007)
38.42 M.T. Marshall, M. Rath, B. Moynihan: The Virtual Bodhran – The Vodhran. In: *Proc. Int. Conf. New Interfaces Music. Expr.*, *Dublin* (2002) pp. 118–119
38.43 E. Maestre, J. Janer, M. Blaauw, A. Pérez, E. Guaus: Acquisition of violin instrumental gestures using a commercial EMF tracking device. In: *Proc. Int. Comput. Music Conf.*, *Copenhagen* (2007)
38.44 A.R. Jensenius, K. Nymoen, R.I. Godøy: A multilayered GDIF-based setup for studying coarticulation in the movements of musicians. In: *Proc. Int. Comput. Music Conf* (2008) pp. 743–746
38.45 H. Wilmers: Bowsense – An open wireless motion sensing platform. In: *Proc. Int. Comput. Music Conf.*, *Montreal* (2009) pp. 287–290
38.46 S. Skogstad, K. Nymoen, M.E. Høvin: Comparing inertial and optical MoCap technologies for synthesis control. In: *Proc. Sound Music Comput.*, *Padova* (2011) pp. 421–426
38.47 G. Welch, E. Foxlin: Motion tracking: No silver bullet, but a respectable arsenal, IEEE Comput. Graph. Appl. **22**(6), 24–38 (2002)
38.48 Y. de Quay, S. Skogstad, A.R. Jensenius: Dance Jockey: Performing electronic music by dancing, Leonardo Music J. **21**, 11–12 (2011)
38.49 M.A.O. Pérez, R.B. Knapp: BioTools: A biosignal toolbox for composers and performers. *Computer Music Modeling and Retrieval*. In: *Sense of Sounds*, Lecture Notes in Computer Science, Vol. 4969, ed. by R. Kronland-Martinet, S. Ystad, K. Jensen (Springer, Berlin, Heidelberg 2008) pp. 441–452
38.50 A. Tanaka: Musical technical issues in using interactive instrument technology with application to the BioMuse. In: *Proc. Int. Comput. Music Conf.*, *Waseda* (1993) pp. 124–124
38.51 K. Nymoen, M.R. Haugen, A.R. Jensenius: MuMYO – Evaluating and exploring the MYO armband for musical interaction. In: *Proc. Int. Conf. New Interfaces Music. Expr.*, *Baton Rouge* (2015)

38.52 C. Lee, S.K. Yoo, Y.J. Park, N.H. Kim, K.S. Jeong, B.C. Lee: Using neural network to recognize human emotions from heart rate variability and skin resistance. In: *Proc. IEEE Eng. Med. Biol., Shanghai* (2005) pp. 5523–5525

38.53 G.H. Zimny, E.W. Weidenfeller: Effects of music upon GSR and heart-rate, Am. J. Psychol. **76**(2), 311–314 (1963)

38.54 D.G. Craig: An Exploratory Study of Physiological Changes during "Chills" Induced by Music, Musicae Scientiae **9**(2), 273–287 (2005)

38.55 M. Ojanen, J. Suominen, T. Kallio, K. Lassfolk: Design principles and user interfaces of Erkki Kurenniemi's electronic musical instruments of the 1960's and 1970's. In: *Proc. Int. Conf. New Interfaces Music. Expr., New York* (2007) pp. 88–93

38.56 E.R. Miranda, B. Boskamp: Steering generative rules with the EEG: An approach to brain-computer music interfacing. In: *Proc. Sound Music Comput., Salerno* (2005)

38.57 A.R. Jensenius, A. Camurri, N. Castagne, E. Maestre, J. Malloch, D. McGilvray, D. Schwarz, M. Wright: Panel: The need of formats for streaming and storing music-related movement and gesture data. In: *Proc. Int. Comput. Music Conf* (2007) pp. 13–16

38.58 Motion Lab Systems: *The C3D File Format: User Guide* (Motion Lab Systems, Baton Rouge 2008)

38.59 H. Chung, Y. Lee: MCML: Motion capture markup language for integration of heterogeneous motion capture data, Comput. Stand. Interfaces **26**(2), 113–130 (2004)

38.60 T. Tsutsui, S. Saeyor, M. Ishizuka: MPML: A multimodal presentation markup language with character agent control functions. In: *Proc. (CD-ROM) WebNet, San Antonio* (2000) pp. 30–37

38.61 E. Hartman, J. Cooper, K. Spratt: Swing set: Musical controllers with inherent physical dynamics. In: *Proc. Int. Conf. New Interfaces Music. Expr* (2008) pp. 356–357

38.62 B. Manjunath, P. Salembier, T. Sikora: *Introduction to MPEG-7: Multimedia Content Description Interface* (Wiley, New York 2002)

38.63 M. Evrard, D. Couroussé, N. Castagné, C. Cadoz, J.-L. Florens, A. Luciani: The GMS File Format: Specifications of the version 0.1 of the format, Technical report, INPG, ACROE/ICA, Grenoble, France (2006)

38.64 J. Morrison: EA IFF 85: Standard for Interchange Format Files. Technical report, Electronic Arts (1985)

38.65 A.R. Jensenius, T. Kvifte, R.I. Godøy: Towards a gesture description interchange format. In: *Proc. Int. Conf. New Interfaces for Music. Expr* (2006) pp. 176–179

38.66 M. Wright, A. Freed, A. Momeni: OpenSound control: State of the art 2003. In: *Proc. Int. Conf. New Interfaces Music. Expr., Montreal* (2003)

38.67 J.J. Burred, C.E. Cella, G. Peeters, A. Roebel, D. Schwarz: Using the SDIF sound description interchange format for audio features. In: *Proc. Int. Conf. Music Inf. Retr.* (2008) pp. 427–432

38.68 P. Roland: The Music Encoding Initiative (MEI). In: *Proc. 1st Int. Conf. Music. Appl. using XML* (2002) pp. 55–59

38.69 A. Camurri, P. Coletta, A. Massari, B. Mazzarino, M. Peri, M. Ricchetti, A. Ricci, G. Volpe: Toward real-time multimodal processing: EyesWeb 4.0. In: *Proc. Artif. Intell. Simul. Behav. Conv., Leeds* (2004) pp. 22–26

38.70 A. Camurri, P. Coletta, G. Varni, S. Ghisio: Developing multimodal interactive systems with EyesWeb XMI. In: *Proc. Int. Conf. New Interfaces for Music. Expr., New York* (2007) pp. 305–308

38.71 B. Burger, P. Toiviainen: MoCap Toolbox – A Matlab toolbox for computational analysis of movement data. In: *Proc. Sound Music Comput. Conf.* (2013) pp. 172–178

38.72 J. Jaimovich, B. Knapp: Synchronization of multimodal recordings for musical performance research. In: *Proc. Int. Conf. New Interfaces Music. Expr., Sydney* (2010) pp. 372–374

38.73 O. Mayor, J. Llop, E. Maestre: RepoVizz: A multimodal on-line database and browsing tool for music performance research. In: *Int. Soc. Music Inform. Retr. Conf* (2011)

Part F Music and Media

Ed. by Isabel Barbancho

Music has been part of human beings since the very beginning. The opportunities for playing music, enjoying music, searching for music, producing music, learning music, composing music, etc. have evolved and changed through history in parallel with the developments of new technological possibilities and human knowledge itself.

Nowadays, digital technology has great impact on all aspects of our lives and music is one of them. Technology can enhance music in many ways. One of the key ways in which technology impacts music is that it makes lots of process related to music cheaper.

This section is devoted to showing how music, and people interacting with music in any sense, are affected by today's technology.

Digital technology used for music can be included in the music information retrieval (MIR) discipline. MIR is a truly interdisciplinary area, involving researchers, developers, educators, librarians, students and professionals from the disciplines of musicology, cognitive science, library and information science, computer science, electrical engineering and many others. This diversity gives an idea of how big the influence of music is in many aspects of life.

Specifically, in this section we will learn new approaches to finding music, new technologically driven possibilities for music learners, how technology can help hearing-impaired persons enjoy music, how the music industry has changed due to technology, new audio-signal processing approaches to better understand music evolution, new interactive possibilities for music, and how traditional music instruments can be enhanced to provide us with new possibilities.

Following this contextualization, a brief synopsis of each chapter now follows.

Chapter 39: *Juan Pablo Bello* et al., in their chapter *Content-Based Methods for Knowledge Discovery in Music*, present several computational approaches aimed at supporting knowledge discovery in music. They combine data mining, signal processing and data visualization techniques for the automatic analysis of digital music collections. They show how these techniques can be used to extract musically meaningful information from audio signals in order to identify their repetitive structures and representative patterns, and to characterize similarities within collections. With the help of illustrative examples, they show how these content-based methods facilitate the development of novel modes of access, analysis and interaction with digital content that can empower the study and appreciation of music.

Chapter 40: *Marshall Chasin* and *Neil S. Hockley*'s chapter, entitled *Hearing Aids and Musicians: Some Theoretical and Practical Issues*, focuses on the hearing assessment of musicians as well as how to recommend and specify the exact parameters for hearing aid amplification for hard-of-hearing people who either play musical instruments or merely like to listen to music. Music is typically listened to or played at a higher sound level than speech and there are some spectral and temporal differences between music and speech that have implications for differing electroacoustic hearing aid technologies for the two types of input. This involves a discussion of some hearing aid technologies best suited to amplified music as well as some clinical strategies for the hearing health care professional to optimize hearing aids for music as an input.

Chapter 41: *Estefanía Cano* et al. address the application of music information retrieval (MIR) technologies in the development of music education tools in their chapter *Music Technology and Education*. First, the relationship between technology and music education is described from a historical point of view. Then, three MIR technologies used within a music education context are presented: the use of pitch-informed solo and accompaniment separation as a tool for practice content creation, drum transcription for real-time music practice, and guitar transcription with plucking style and expression style detection. To conclude the chapter, some remaining challenges that need to be addressed to more effectively use MIR technologies in the development of music education applications are described.

Chapter 42: *Lorenzo J. Tardón* et al., in their chapter *Music Learning: Automatic Music Composition and Singing Voice Assessment*, focus on diverse aspects of how technology can be used for music learning. In this chapter, the goal is to describe a virtual environment that partially resembles the traditional music learning process and the music teacher's role, allowing for a complete interactive self-learning process. The complete chain of an interactive singing-learning system including tools and concrete techniques is presented. First, a set of training exercises is provided by the system and the user's performance is assessed, before the system provides the user with new exercises selected or created according to the results of the evaluation. Therefore, methods for the creation of user-adapted exercises and the automatic evaluation of singing skill is presented, together with techniques for the dynamic generation of musically

meaningful singing exercises, adapted to the user's skill.

Chapter 43: *Nadine Kroher* et al., in their chapter *Computational Ethnomusicology: A Study of Flamenco and Arab-Andalusian Vocal Music*, approach flamenco and Arab-Andalusian vocal music through the analysis of two representative pieces. They apply a hybrid methodology consisting of audio-signal processing to describe and contrast their melodic characteristics followed by musicological analysis. The use of such computational analysis tools complements a musicological-historical study with the aim of supporting the discovery and understanding of the specific characteristics of these musical traditions and their similarities and differences, while offering solutions to more general music information retrieval (MIR) research challenges.

Chapter 44: *Alexander Lerch*'s contribution is devoted to *The Relationship Between Music Technology and Music Industry*. The music industry has changed drastically over the last century and most of its changes and transformations have been technology driven. Music technology – encompassing musical instruments, sound generators, studio equipment and software, perceptual audio coding algorithms, and reproduction software and devices – has shaped the way music is produced, performed, distributed and consumed. This chapter describes technological innovations and trends in the past and their impact on musicians, listeners, and the music industry itself.

Chapter 45: *Jesús Corral García*, *Panos Kudumakis* et al. present a chapter entitled *Enabling Interactive and Interoperable Semantic Music Applications*. New interactive music services have emerged, although they currently use proprietary file formats. In order to enable interoperability among these services, the ISO/IEC Moving Picture Experts Group (MPEG) issued a new standard, the so-called, MPEG-A: Interactive Music Application Format (IM AF). This chapter reviews the IM AF standard and its features and provides a detailed description of the design and implementation of an IM AF codec and its integration into a popular open-source analysis, annotation and visualization audio tool, known as Sonic Visualiser. This is followed by a discussion highlighting the benefits of their combined features, such as automatic chord or melody extraction time-aligned with the song's lyrics. This integration provides the semantic music research community with a test-bed for enabling further development and comparison of new Sonic Visualiser plugins.

Chapter 46: *George Tzanetakis* and *Peter Driessen*, in their chapter *Digital Sensing of Musical Instruments*, describe hyperinstruments, which are acoustic instruments that are augmented with digital sensors for capturing performance information and in some cases offering additional playing possibilities. Direct sensors are integrated onto the physical instrument, which possibly requiring modifications. Indirect sensors such as cameras and microphones can be used to analyze performer gestures without requiring modifications to the instrument. Until recently, hyperinstruments were mostly used for electroacoustic music creation but, as stated in this chapter, they have a lot of potential in systematic musicological applications involving music performance analysis.

39. Content-Based Methods for Knowledge Discovery in Music

Juan Pablo Bello, Peter Grosche, Meinard Müller, Ron Weiss

This chapter presents several computational approaches aimed at supporting knowledge discovery in music. Our work combines data mining, signal processing and data visualization techniques for the automatic analysis of digital music collections, with a focus on retrieving and understanding musical structure.

We discuss the extraction of midlevel feature representations that convey musically meaningful information from audio signals, and show how such representations can be used to synchronize different instances of a musical work and enable new modes of music content browsing and navigation. Moreover, we utilize these representations to identify repetitive structures and representative patterns in the signal, via self-similarity analysis and matrix decomposition techniques that can be made invariant to changes of local tempo and key. We discuss how structural information can serve to highlight relationships within music collections, and explore the use of information visualization tools to characterize the patterns of similarity and dissimilarity that underpin such relationships.

With the help of illustrative examples computed on a collection of recordings of Frédéric Chopin's Mazurkas, we aim to show how these content-based methods can facilitate the development of novel modes of access, analysis and interaction with digital content that can empower the study and appreciation of music.

The rapid and sustained growth of digital music sharing and distribution is nothing less than astounding. A multitude of digital music services provide access to tens of millions of tracks, both legally and illegally. Such abundance of content, coupled with the relative ease of access and storage afforded by recent technologies, means that music is shared and listened to more than ever before in history.

Using computational methods to help users find and organize music information is a widely researched topic in industry and academia. Existing approaches can be coarsely divided into two types: in *content-based* methods the information is obtained directly from the analysis of audio signals, scores and other representations of the music, whereas *context-based* methods are based on information surrounding the music content, such as usage patterns, tags and structured metadata. While a significant amount of research has been devoted to the former strategy – see [39.1] for an early review – the latter has been historically favored in industrial applications such as music recommendation and playlist generation. Content-based analysis is sometimes seen as providing too little *bang for the buck*, with some observers going as far as wondering whether it is at all necessary for the retrieval of music information [39.2].

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_39

Yet, we argue that there are numerous data-mining problems for which context-based analysis is insufficient, as it tends to be low in specifics and unevenly distributed across artists and styles. Consider for example the issue of tracing back the original sources of samples used in electronic or hip-hop recordings; or of identifying the many derivations of George Gershwin's *I Got Rhythm* in the jazz catalog; or of finding quotations of a given Wagner motif in 20th century modernist music; or of quantifying which movements, artists and compositions are cited most often and are therefore the most influential. These problems are motivated by the needs of sophisticated users such as media producers, Foley artists, sound designers, film and game composers, copyright lawyers, musicologists, and professional and amateur musicians, for whom music search necessarily goes beyond the passive act of music recommendation. We believe that the development of robust and scalable solutions to these problems, and many others that could be listed instead, passes through the automated analysis of the musical content.

This chapter aims to introduce the reader to computational approaches to content-based analysis of digital music recordings. More specifically, we give an overview of a number of techniques for music structure analysis, i. e., the identification of the patterns and relationships that govern the organization of sounds in music, and discuss how the outcomes of this analysis can facilitate data mining in music. This review is intended for an audience interested in music search, organization and discovery, that is not steeped in the field of music information retrieval (MIR). Therefore the emphasis is not on technical details, which are published elsewhere in the literature, but on the presentation of examples and qualitative results that illustrate the operation and potential of the presented approaches.

The chapter is organized as follows: Section 39.1 familiarizes the reader with the basics of music structure analysis, discusses the fundamental role that repetition plays in it, and introduces the corpus of music that will be used throughout this chapter. Section 39.2 presents standard methods for music signals analysis. Section 39.3 introduces methods for temporal alignment of different representations of a given musical piece and shows how such alignments can be used to create novel user interfaces. Section 39.4 demonstrates how to characterize the patterns of repetition in music via self-similarity analysis, which has numerous applications in segmenting, organizing and visualizing music recordings. Section 39.5 introduces a powerful technique for structure analysis using matrix factorization with applications in identifying representative patterns and segmenting music signals. Finally, Sect. 39.6 presents our conclusions and outlook on the field.

39.1 Music Structure Analysis

The architectural structure of a musical piece, its *form*, can be described in terms of a concatenation of sectional units. The amount of repetition amongst these units defines the spectrum of possible forms, ranging from strophic pieces, where a single section is continuously repeated (as is the case for most lullabies), to through-composed pieces, where no section ever recurs. While repetition is not a precondition to music, it undeniably plays a central role (*the basis of music as an art form* according to [39.3]), and is closely related to notions of coherence, intelligibility and enjoyment in its perception [39.4]. Indeed, some observers estimate that more than 99% of music listening involves repetition, both internal to the work and of familiar passages [39.5].

However, the notion of repetition in structural analysis is not rigid and, depending on the composition, might include significant variations in the musical content of the *repeated* parts. This is even more true for recordings featuring changes in instrumentation, ornamentation and expressive variations of tempo and dynamics. In other words, an exact recapitulation of the content is not required in order for a part to be considered to be repeated. As a consequence, the analysis and annotation of musical structure is, to a certain degree, ambiguous [39.6, 7].

Take for example Frédéric Chopin's Mazurka in F major, Opus 68, No. 3, a piano work to which we will refer throughout this chapter as M68-3. One way to describe the structure of this piece is as: $\mathcal{A}_1\mathcal{A}_2\mathcal{B}_1\mathcal{B}_2\mathcal{A}_3\mathcal{T}\mathcal{C}_1\mathcal{C}_2\mathcal{A}_4\mathcal{A}_5$. In this description, each letter denotes a pattern and each subscript the instance number of a pattern's repetition. $\mathcal{T}$ is a special symbol denoting a transitional section. The four patterns are depicted, in score format, in Fig. 39.1.

Note that this annotation is by no means unique and implies a number of choices. First, grouping the music into a relatively small number of parts requires tolerating slight variations across repetitions. For example, pattern $\mathcal{A}$ presents with two alternative endings, with the last bar of segments $\mathcal{A}_1$ and $\mathcal{A}_4$ differing harmonically from the last bar of segments $\mathcal{A}_2$, $\mathcal{A}_3$

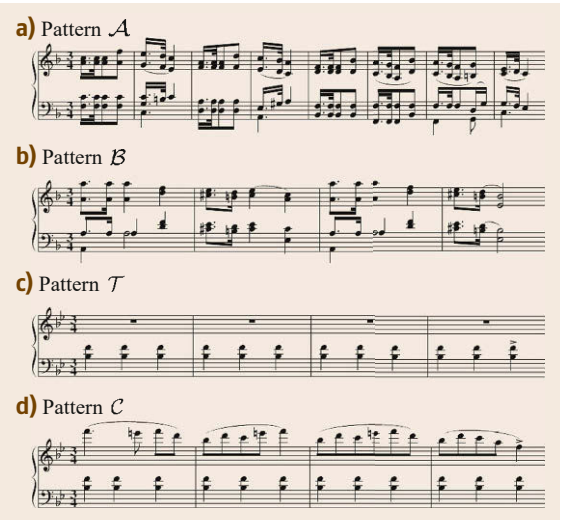


Fig. 39.1a–d Mazurka Op. 68, No. 3 (M68-3) in F major

and $\mathcal{A}_5$. We could have chosen to break these occurrences into two groups, but that would ignore the high degree of overlap that otherwise exists between them. Second, the description avoids patterns consisting of a small number of repeating subpatterns. For example, $\mathcal{T}$ and $\mathcal{C}$ are considered to be two different parts despite their strong harmonic similarities. The alternative would be to merge $\mathcal{T}\mathcal{C}_1\mathcal{C}_2$ into a single pattern of highly repetitive subpatterns. In the following sections we will illustrate how these decisions relate to the information within the music signals, and their implications for the proposed analyses.

Throughout this chapter we draw examples from the Mazurka dataset, compiled by the Center for the History and Analysis of Recorded Music in London [39.8]. The set includes 2919 recorded performances of the 49 Frédéric Chopin's Mazurkas, resulting in an average of 58 renditions per Mazurka. These recordings, featuring 135 different pianists, cover a range of more than 100 years beginning in 1902 and ending in 2008. This makes the dataset a rich and unique resource for the analysis of style changes and expressivity in piano performance, and of the evolution of recording techniques and practices [39.9].

Our analysis is mainly focused on a subset of 298 recordings that correspond to five Mazurkas. In addition to M68-3, mentioned above, these include: Opus 17, No. 4 in A minor (M17-4); Op. 24, No. 2 in C major (M24-2); Op. 30, No. 2 in B minor (M30-2); and Op. 63, No. 3 in C$^\sharp$ minor (M63-3). These recordings are chosen because they have been the subject of extensive musicological studies that resulted, amongst other things, in manually annotated beat positions [39.10]. Additionally, the musical structures of these Mazurkas are comparatively well defined, a fact that we have exploited to manually annotate their forms, as depicted in Fig. 39.2. Please note that the annotations were performed only once on the score representation, and then propagated to all recordings using the beat annotations mentioned above.

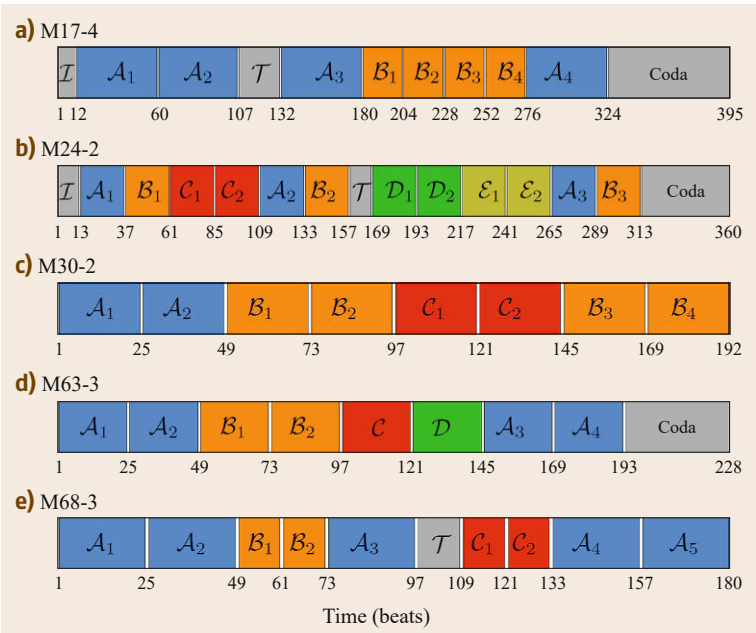


Fig. 39.2a–e Illustration of manually generated structural descriptions for five different Mazurkas

39.2 Feature Representation

All content-based music analysis begins with the extraction of a meaningful feature representation from the audio signal. This representation typically encodes information about one or more musical characteristics, e.g., harmony, melody, rhythm, or timbre as required by the specific task. There are numerous signal processing techniques that can be used to this end, such as the ubiquitous Mel-frequency Cepstral coefficients (MFCC), which are used for the analysis of timbre and texture (e.g., by [39.11, 12]). For a comprehensive list of audio features and their implementation, the reader is referred to [39.13].

In our work we use chroma features, introduced in Sect. 39.2.1, which can be used to derive information about chords. As shown in Sect. 39.2.2, these features are a powerful tool for music structure analysis, because repeating segments are often characterized by common chord progressions.

39.2.1 Chroma Features

Most musical parts are characterized by a particular melody and harmony. In order to identify repeating patterns in music recordings it is therefore useful to convert the audio signal into a feature representation that reliably captures these elements of the signal, but is not sensitive to other components of the signal such as instrumentation or timbre. In this context, *chroma features*, also referred to as *pitch class profiles* (PCPs), have turned out to be a powerful midlevel representation for describing harmonic content. They are widely used for various music signal analysis tasks, such as chord recognition [39.14], cover song identification [39.15, 16], and many others [39.17–19].

It is well known that human perception of pitch is cyclical in the sense that two pitches an integer number of octaves apart are perceived to be of the same type or class. This is the basis for the helical model of pitch perception, where pitch is separated into two dimensions: *tone height* and *chroma* [39.20]. Assuming the equal-tempered scale, and enharmonic equivalence, the chroma dimension corresponds to the twelve pitch classes used in Western music notation, denoted by $\{\mathrm{C}, \mathrm{C}^\sharp, \mathrm{D}, \ldots, \mathrm{B}\}$, where different pitch spellings such as $\mathrm{C}^\sharp$ and $\mathrm{D}^\flat$ refer to the same chroma. A *pitch class* is defined to be the set of all pitches that share the same chroma. For example, the pitch class corresponding to the chroma C is the set $\{\ldots, \mathrm{C0}, \mathrm{C1}, \mathrm{C2}, \mathrm{C3}, \ldots\}$.

There are several methods available for the computation of chroma features from audio, usually involving the warping of the signal's short-time spectrum or its decomposition into log-spaced subbands. This is followed by a weighted summation of energy across spectral bins corresponding to the same pitch class [39.15, 17, 19, 21]. Each chroma vector characterizes the distribution of the signal's local energy across the twelve pitch classes. Just as with the short-time Fourier transform (STFT), chroma vectors can be calculated sequentially on partially overlapped blocks of signal data, resulting in a so-called *chromagram*. The literature also proposes a number of ways in which the standard chroma feature representation can be improved by minimizing the effects of harmonic noise [39.22], by timbre and dynamic changes [39.19, 23], or by tempo variations via beat synchronization [39.15].

Figure 39.3a depicts the waveform of a 1976 recording of M68-3 performed by Sviatoslav Richter. Pattern labels are shown on the horizontal axis with boundaries marked as vertical red lines. Figure 39.3b shows the corresponding sequence of normalized chroma feature vectors, at a resolution of ten vectors per second, with the pitch classes of the chromatic scale starting in A, and ordered from bottom to top. As expected, most of the signal energy is concentrated on the classes corresponding to the F major key: F, G, A, $\mathrm{B}^\flat/\mathrm{A}^\sharp$, C, D and E. Additionally, careful inspection of the chromagram shows that the different patterns of this piece can be associated with distinct subsequences of chroma vectors.

39.2.2 Feature Trajectories

To make the latter point more evident we can use an alternative visualization of the chromagram in Fig. 39.3b. The values in a single chroma feature vector can be interpreted as a set of coordinates describing a point in the 12-dimensional space of pitch classes. Connecting those points over time results in a trajectory in feature space. Since this trajectory cannot be directly visualized in 12 dimensions, we use principal component analysis [39.24] to project this information onto its two principal components, resulting in the black line shown in Fig. 39.4. Since time is not explicitly encoded in this visualization, Fig. 39.4 also shows the two-dimensional (2-D) histograms of trajectory values, color-coded such that points contained within instances of patterns $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{T}$ are in blue, red, orange and gray respectively.

The projection clearly characterizes the main repetitions in the feature sequence as similar subtrajectories, i. e., segments of the main trajectory that are in close proximity to each other. Note that repetitions do not

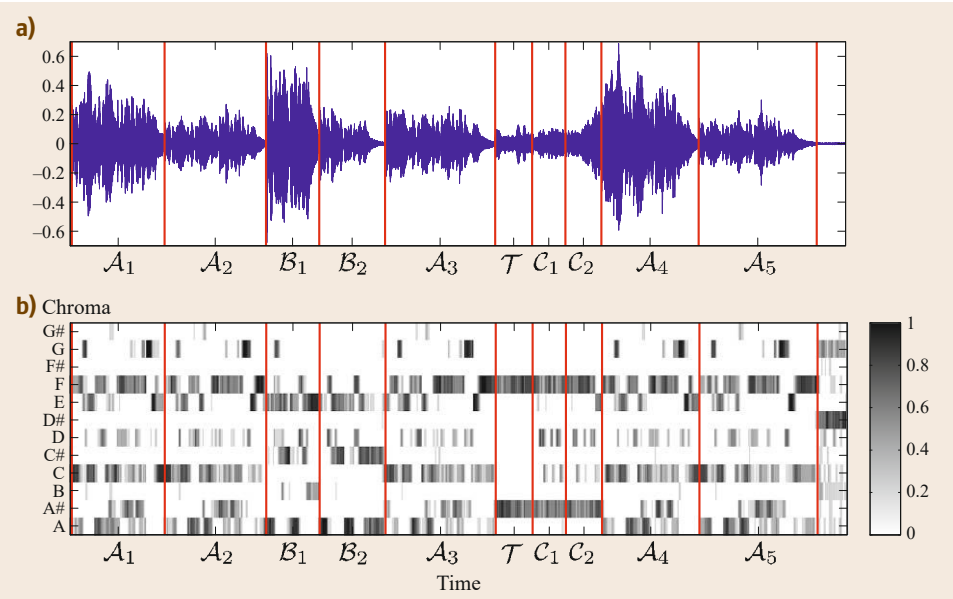


Fig. 39.3a,b Feature representations for a performance by Richter (1976) of M68-3. **(a)** Waveform. **(b)** Corresponding chromagram. Pattern boundaries are marked with *vertical red lines* and pattern labels are shown on the *time axis*

form fully overlapping subtrajectories due to expressive changes in tempo, timbre and dynamics. Despite this variability the histograms clearly show how each pattern results in distinct trajectory shapes. For example, pattern $\mathcal{A}$ has a multimodal distribution resulting from the six different chord types that are part of its progression: F major, C major, B♭ major and G major as well as D minor and A minor. In contrast, pattern $\mathcal{B}$, which is strongly dominated by A major chords (while also featuring some D minor and E major chords) has a much more narrow distribution. This is even more extreme for patterns $\mathcal{C}$ and $\mathcal{T}$, which are highly overlapping and almost exclusively dominated by a B♭-F dyad played ostinato, as can be seen in the score representation of Fig. 39.1.

This example illustrates how chromagrams successfully capture harmonic information in the signal. In the following sections we show how these features can be used to synchronize different instances of a musical work and enable new modes of music content browsing and navigation.

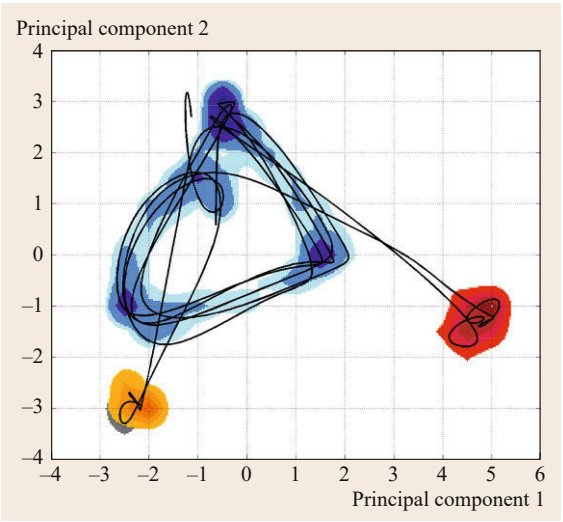


Fig. 39.4 Trajectory of 12-dimensional chroma features projected onto two dimensions using the chromagram shown in Fig. 39.3b

39.3 Music Synchronization and Navigation

Musical works can be represented in many different domains, e.g., in different audio recordings, MIDI (Musical Instrument Digital Interface) files, or as digitized sheet music. The general goal of music synchronization is to automatically align multiple information sources related to the same musical work. Here, *music synchronization* denotes a procedure which, for a given position in one representation of a piece of music, determines the corresponding position within another representation [39.19, 25]. The linking information produced by the synchronization process can be used to propagate information across these representations.

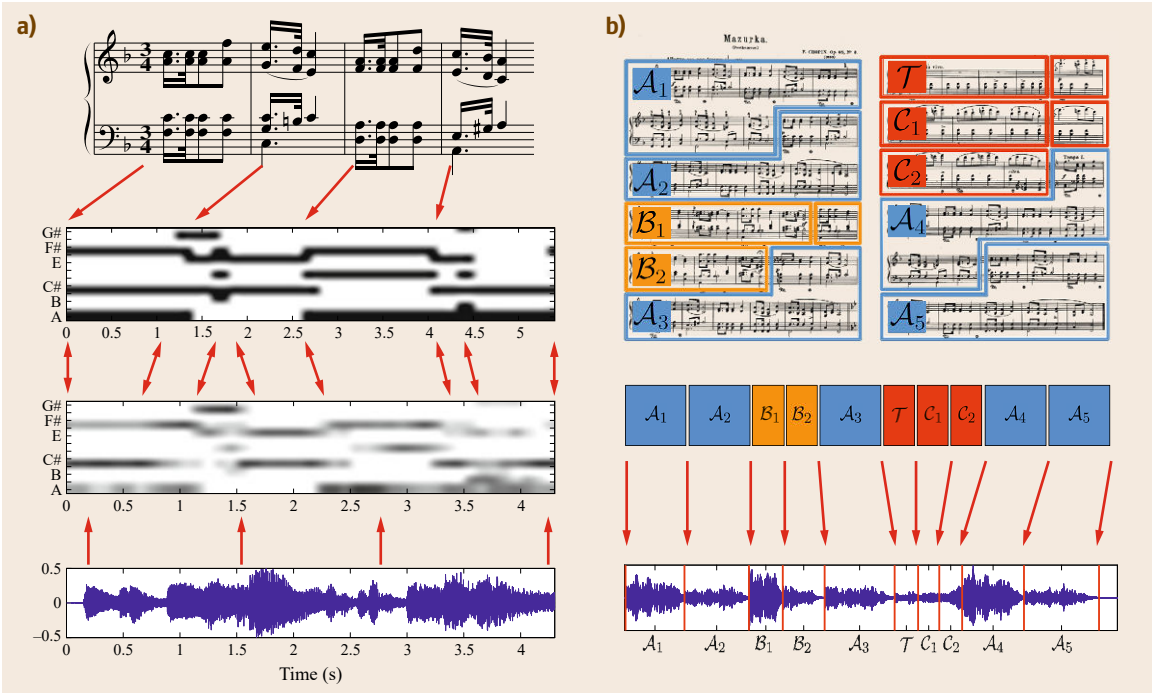


Fig. 39.5a,b Illustration of the score-audio synchronization pipeline used to transfer structure annotations from the score to the audio domain. **(a)** Score and audio representation with the corresponding chromagrams. The *red arrows* indicate the score-audio synchronization result. **(b)** The score-based annotations are transferred to the audio domain using the linking information supplied by the synchronization procedure

Consider the example in Fig. 39.5. The left column demonstrates the synchronization process between a musical score (top) and a recording of the same piece (bottom). Both the score and audio are converted to a common midlevel representation: the chromagram. While one would naturally expect the audio chromagram to be much noisier than the chromagram derived from the score, both representations are expected to contain energy in the pitch classes corresponding to the played notes, and to be organized according to the sequence of notes in the score. Therefore, standard alignment procedures based on dynamic time warping can be used to synchronize the two feature sequences. See [39.19, 25, 26] for details.

There are many potential applications of the synchronization process. For example, if the musical form of the piece has been manually annotated using the score representation – as shown in the right-hand side of Fig. 39.5 – the synchronization result can be used to transfer this, or any other annotation, to a recording of the same piece, regardless of variations of local or global tempo. This alleviates the need for the laborious process of manually annotating multiple performances of a given work. The same techniques can be applied to score following for automatic page turning or to adding subtitles to a video recording of an opera performance, to name only some further applications.

Another application is the creation of novel user interfaces for inter- and intradocument navigation in music collections. Figure 39.6 shows an interface (from [39.27–29]) that allows the user to interact with several recordings of the same piece of music, which have been previously synchronized. The timeline of each recording is represented by a slider bar, whose indicator (a small down arrow) points to the current playback position. Note that these positions are synchronized across performances. A user may listen to a specific recording by activating the corresponding slider bars and then, at any time during playback, seamlessly switch to another recording. Additionally, the slider bars can be segmented and color coded to visualize symbolic annotations common to all recordings, such as a chord transcription or a structure segmentation as shown in the figure. Furthermore, users can jump directly to the beginning of any annotated element simply by clicking on the corresponding block, thus greatly facilitating intradocument navigation. A similar functionality was introduced in [39.30].

The interface offers three different timeline modes. In *absolute mode*, shown in the top of Fig. 39.6, the

Fig. 39.6a,b User interface for intra- and interdocument navigation of music collections. The *slider bars* correspond to four different performances of M68-3. They are segmented and color-coded according to the structural annotations in Fig. 39.2. The *down arrows* point to the current playback time, synchronized across performances. The *absolute* (**a**) and *relative* (**b**) timeline modes are shown

length of a particular slider bar is proportional to the duration of the respective recording. In *relative mode*, shown in the bottom of the figure, all timelines are linearly stretched to the same length. The third and final mode, referred to as *reference mode*, uses a single recording as a reference. All other timelines are then temporally warped to run synchronous to the reference. In all cases, the annotations are adjusted according to the selected mode. Such functionalities open up new possibilities for viewing, comparing, interacting, and evaluating analysis results within a multiversion, and multimode, framework [39.29].

39.4 Self-Similarity in Music Recordings

Since their introduction to the music information community in [39.31], self-similarity matrices (SSM) have become one of the most widely used tools for music structure analysis. Their appeal resides in their ability to characterize patterns of recurrence in a feature sequence, which are closely related to the structure of musical pieces.

39.4.1 Self-Similarity Revisited

Given a sequence of N feature vectors, and a function to measure the pairwise similarity between them, the *self-similarity matrix* $\mathbf{S}$ is defined to be the $N \times N$ matrix of pairwise similarities between all feature vectors in the sequence. For the examples in this section we use the chroma features as introduced in Sect. 39.2.1 and the cosine similarity function, i. e., the inner product between the normalized chroma vectors.

Figure 39.7a shows the matrix $\mathbf{S}$ for the chromagram in Fig. 39.3. Both the horizontal and vertical axes represent time, beginning at the bottom-left corner of the plot. Hand-annotated pattern boundaries are marked by vertical and horizontal red lines. Note that if the similarity function is symmetric, the matrix $\mathbf{S}$ is also

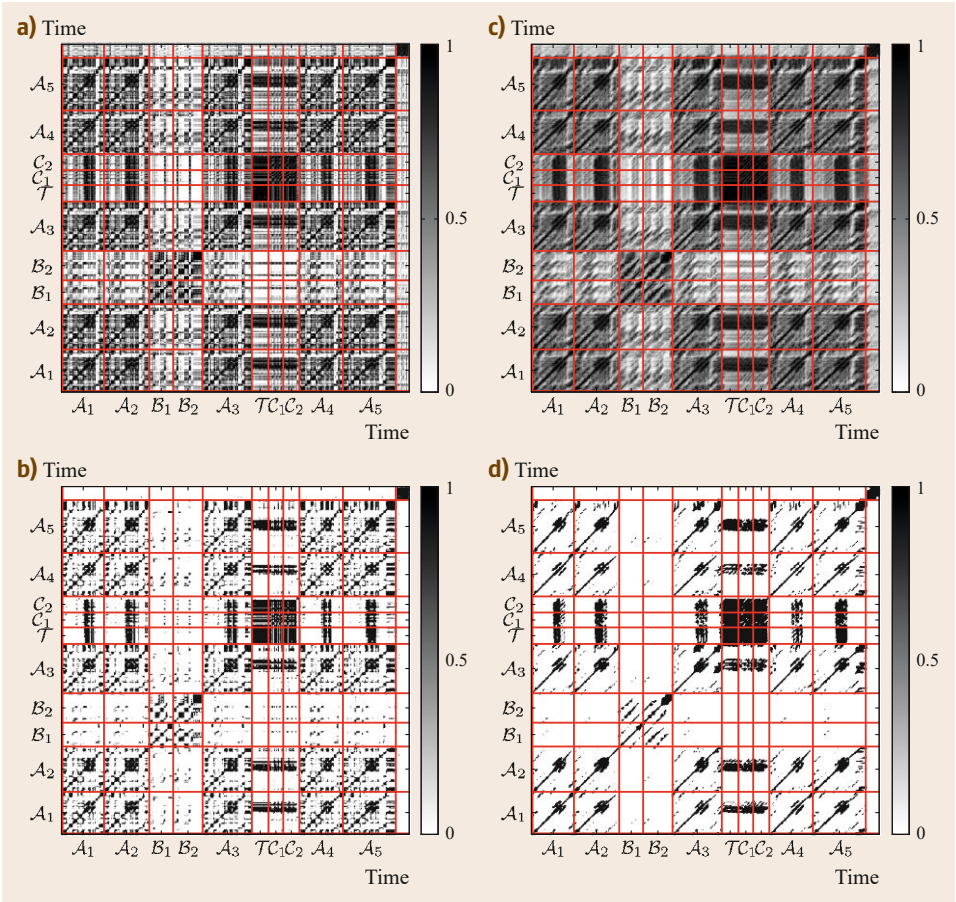


Fig. 39.7a–d Self-similarity matrices for M68-3 performed by Richter, 1976: **(a)** SSM, **(b)** thresholded SSM, **(c)** smoothed SSM, **(d)** thresholded smoothed SSM. Both vertical and horizontal axes represent time

symmetric with respect to the main diagonal, and that larger/smaller similarity values are represented respectively by darker/lighter colors in the plot. It can be readily observed that similar feature subsequences result in dark diagonal lines or stripes of high similarity. For example, all submatrices corresponding to the two instances of the $\mathcal{A}$ part contain a dark diagonal stripe. The segments corresponding to the $\mathcal{C}$ and $\mathcal{T}$ parts are rather homogeneous with respect to their harmonic content. As a result, entire blocks of high similarity are formed, indicating that each feature vector is similar to every other feature vector within these segments.

This example clearly illustrates how SSMs capture the repetitive structure of music recordings and reveal the location of repeating patterns in the form of diagonal stripes of high similarity. It follows that this information can be exploited for automatic segmentation [39.32], the extraction of musical form [39.19, 33], the detection of chorus sections [39.34], or music thumbnailing [39.35]. For a comprehensive review the reader is referred to [39.36]. However, the reliable extraction of such information from the SSM is problematic in the presence of distortions caused by variations in dynamics, timbre, note ornaments (e.g., grace notes, trills, arpeggios), modulation, articulation, or tempo. The following section describes how SSMs can be processed to minimize their sensitivity to such distortions.

39.4.2 Enhancing Self-Similarity Matrices

A number of strategies have been proposed to emphasize the diagonal structure information in SSMs. These strategies commonly include a combination of some form of filtering along the diagonals, and the thresholding of spurious, off-diagonal values. While many alternative approaches have been proposed in the literature (e.g., [39.33–35]), here we briefly discuss methods based on time-delay embedding, contextual similarity and transposition invariance.

The process of filtering SSMs along the diagonals can be framed in terms of *time-delay embedding*, a process that has been widely used for the analysis of dynamical systems [39.37]. In this process, the feature

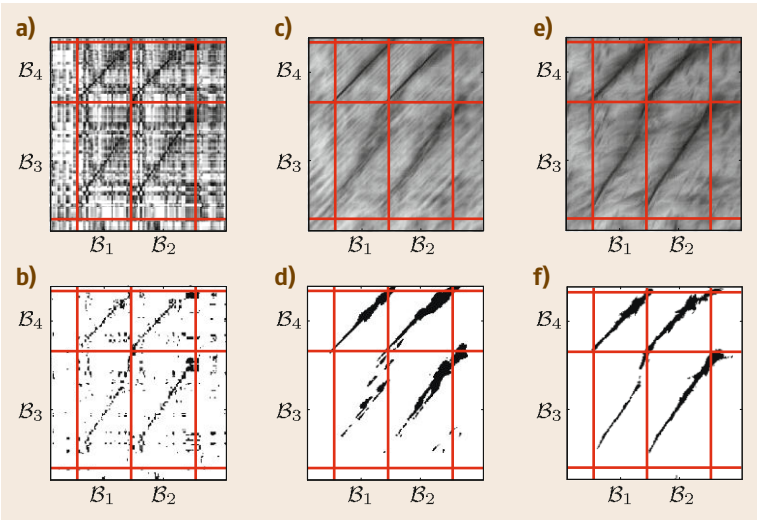


Fig. 39.8a–f SSM smoothing and thresholding in the presence of tempo variations for M30-2 performed by Jonas, 1947: **(a,b)** SSM, **(c,d)** smoothed SSM, and **(e,f)** smoothed SSM with contextual similarity

sequence is converted into a series of feature m-grams, each of which is composed of a stack of m feature vectors. These vectors can be taken from a contiguous window, or spaced by a fixed sample delay τ. In the context of the example in Sect. 39.2.2, setting $m > 1$ implies that pairwise similarities in the SSM computation are computed between subtrajectories instead of between individual feature vectors. Subtrajectories that are parallel to each other result in large self-similarity, and the effect of random crossings is minimized, generally resulting in a smoothed matrix. For $\tau = 1$, the time-delay embedding process is similar to textural windows [39.38] and audio shingles [39.39].

Figure 39.7a,b shows the traditional SSM (i. e., $m = \tau = 1$) computed from M68-3 before and after thresholding, respectively. Similarly, Fig. 39.7c,d shows the corresponding similarity matrices computed using $m = 15$ and $\tau = 1$. It is immediately apparent how the time-delay embedding emphasizes the diagonal line structure of the matrix and significantly minimizes the amount of off-diagonal noise that obscures the structure in the nonembedded matrix. It is worth noting that different methods can be used for thresholding, e.g., by simply ignoring distances larger than a predefined threshold (as was done in this example), by fixing the number of nearest neighbors per sequence element, or by enforcing a rate of recurrence in the thresholded SSM [39.37]. Recent studies show that the latter strategies can improve the applicability of SSMs to a variety of music analysis tasks [39.40, 41].

This form of filtering only works well if the tempo is (close to) constant across the music recording, i. e., where repeating segments have roughly the same length. Stripes corresponding to repeated patterns at the same tempo run diagonally with a slope of 1. Music structure analysis research has most commonly operated on popular music, in which this constant-tempo assumption is often reasonable. However, in other genres such as romantic piano music, expressive performances often result in significant tempo variations. Take for example Jonas' performance of M30-2, shown in Fig. 39.8, where $\mathcal{B}_3$ is played at nearly half the tempo of $\mathcal{B}_1$. This results in a diagonal stripe in the SSM with slope close to 2. Applying time-delay embedding to this data can result in broken stripes, as shown in Fig. 39.8c,d. One way to avoid this loss of information is to use a contextual similarity measure that filters the SSM along various slopes around 1 [39.42]. The resulting SSM, illustrated in Fig. 39.8e,f, has filtered out the off-diagonal noise present in the unprocessed SSM shown in Fig. 39.8a while successfully preserving diagonal structure in the presence of local tempo variations.

It is further possible to make the SSM transposition-invariant by computing the similarity between the original chromagram and its 12 cyclically shifted versions, corresponding to all possible transpositions. The invariant SSM is calculated by taking the point-wise minimum over the 12 resulting matrices [39.34, 43]. An example of the process from an excerpt of M7-5 is shown in Fig. 39.9. The excerpt contains repetitions of two patterns, $\mathcal{A}$ and $\mathcal{B}$, which are actually transpositions of each other, as shown in Fig. 39.9a,b. The standard SSM, shown in Fig. 39.9c,d, differentiates between the repetitions of $\mathcal{A}$ and $\mathcal{B}$, while the transposition-invariant SSM shown in Fig. 39.9e,f, successfully identifies the relationship between the two patterns.

The enhancement procedures described in this section are aimed at emphasizing the diagonal stripe structure of the SSM. In music structure analysis, the

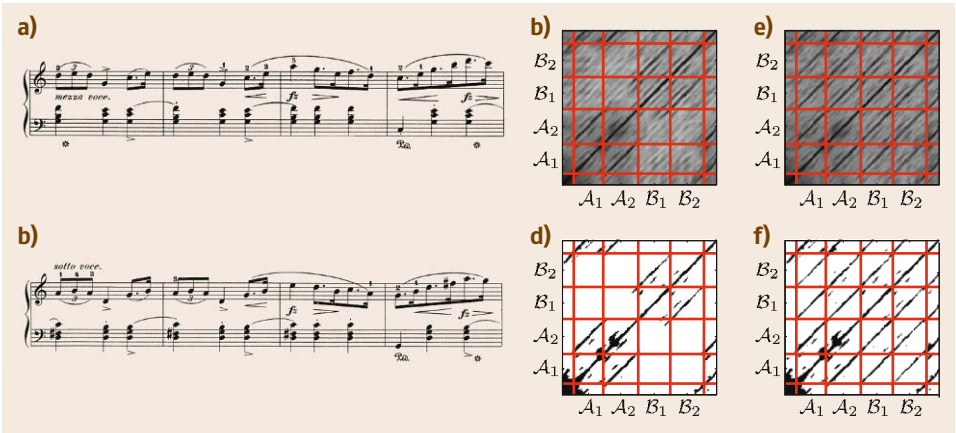


Fig. 39.9a–f Transposition invariance for M7-5 performed by Cohen, 1997: **(a,b)** Scores of patterns $\mathcal{A}$ and $\mathcal{B}$, **(c,d)** SSM before and after thresholding, and **(e,f)** transposition-invariant SSM before and after thresholding

computation of the self-similarity matrix is usually followed by a *grouping* step, where clustering techniques and heuristic rules are used to identify the global repetitive structure from the matrix, or to select representative segments for e.g., music thumbnailing. This is a large topic that we do not aim to fully review in this chapter. For a detailed review of these solutions and discussions of their advantages and disadvantages, the reader is referred to, e.g., [39.36, 44].

39.4.3 Structure-Based Similarity

In the previous sections, we discussed the computation of SSMs as an intermediate step towards the extraction of a piece's musical form or identification of representative excerpts. In this section, we use SSMs directly as a midlevel representation of a recording's structure. Distances measured in this domain can therefore be used to characterize the *structural similarity* between different music recordings [39.45, 46].

A similar problem in bioinformatics is concerned with measuring the similarity between protein structures on the basis of SSM-like representations known as *contact maps* [39.47]. One solution to this problem makes use of the *normalized compression distance* (NCD), an approximation of the joint Kolmogorov complexity between binary objects, to measure the amount of information overlap between contact maps [39.48]. The NCD is versatile, easy to implement using standard compression algorithms, and does not require alignment between the representations. In [39.46, 49], these ideas are transferred and applied to music information retrieval. Experimental results show that these measures successfully characterize global structural similarity for large and small music collections.

Figure 39.10 depicts the entire Mazurka dataset organized by structural similarity. The plot shows a two-

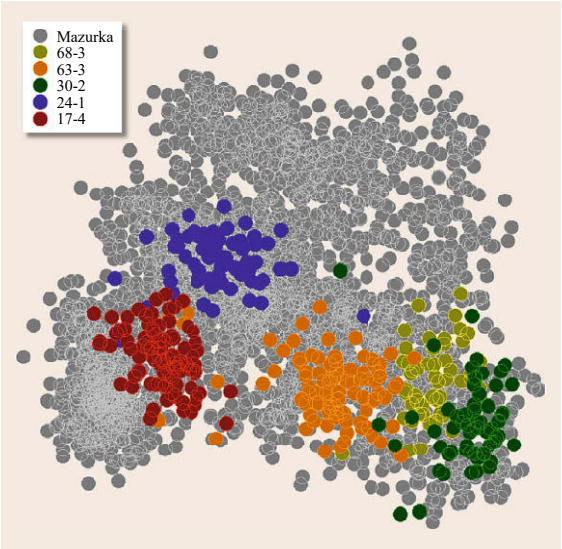


Fig. 39.10 The Mazurka dataset (*gray*) organized according to the NCD-based similarity of SSMs. *Colored points* correspond to recordings of the five Mazurkas considered in this chapter

dimensional projection, obtained using multidimensional scaling [39.50], of the matrix of pairwise NCDs between the SSMs of all 2919 recordings. Each recording is shown as a gray circle, except for those corresponding to performances of M17-4 (depicted in red), M24-1 (blue), M30-2 (green), M63-3 (orange) and M68-3 (olive green).

An informal inspection of this plot shows that most performances of a given work, which feature little or no variation in global structure, naturally form strongly knit clusters in this space. Note that the projection is deceptive in that it makes some of the clusters, e.g., the olive green and orange groups, appear closer to each other than they actually are. As

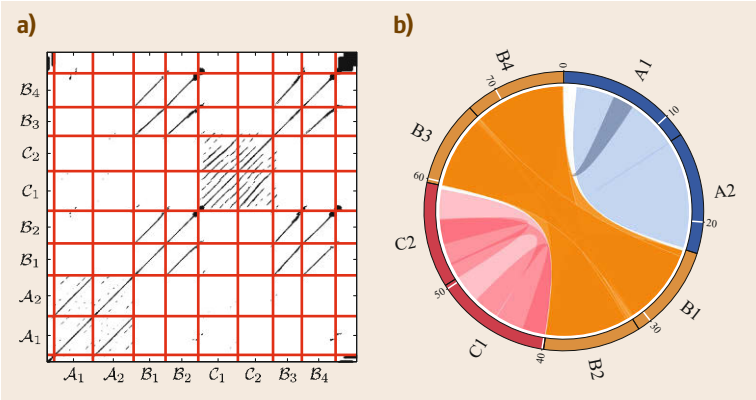


Fig. 39.11a,b Arthur Rubinstein's 1939 performance of M30-2: **(a)** thresholded SSM, and **(b)** radial convergence diagram

a result, works can be easily grouped by means of a simple nearest-neighbor search. However, the NCD is, predictably, sensitive to structural changes. This is typified by the M24-1 performance that appears the furthest from the blue cluster. This recording is part of a 1957 release of a master class by Alfred Cortot, where the pianist's spoken commentary alternates (and at times overlaps) with a partial performance of the piece including errors, nonscored pauses and repetitions. This results in a structure that diverges considerably from those of other performances. Likewise, the M30-2 (green) outlier corresponds to a 1945 performance by Vladimir Horowitz, in which the famous pianist repeats the entire second half of the piece, resulting in a unique $\mathcal{A}_1\mathcal{A}_2\mathcal{B}_1\mathcal{B}_2\mathcal{C}_1\mathcal{C}_2\mathcal{B}_3\mathcal{B}_4\mathcal{C}_3\mathcal{C}_4\mathcal{B}_5\mathcal{B}_6$ structure, see Fig. 39.2c. There are other such examples in this collection, whereby the performer took the liberty to deviate from the notated musical score by playing additional repetitions or leaving out certain parts. A structure-based similarity metric is a natural way to identify these variations.

39.4.4 Visualizing Structure

Self-similarity matrices have been used for visualizing structure in music [39.31, 51] and other domains such as programming [39.52] and computational biology [39.53]. However, SSMs are rarely used for data visualization outside of scientific research, possibly due to the nonintuitive presence of two temporal axes, which results in a complex and redundant representation.

Radial convergence diagrams [39.54] (RCDs) are an alternate vehicle for visualizing structural information derived from SSMs. These diagrams consist of a network of data points arranged on a circle. Groups of related points in the sequence are emphasized by connecting them via links or ribbons. We use the *circos* toolbox [39.55], a powerful and popular tool for information visualization based on circular graphs, to generate the RCDs shown in this section [39.54].

An example RCD and the associated SSM computed from a recording of M30-2 is shown in Fig. 39.11. To generate the RCD, the matrix is postprocessed as follows: the lower triangular part as well as the main diagonal and all points within a prespecified distance from it are ignored. The remaining repetitions (*ones* in the matrix) are stored as a list of *links* connecting pairs of elements in the feature sequence. Nearby links in time (i. e., those that form diagonal strips in the SSM) are grouped together. Small groups, either in temporal scope or in membership, are filtered out.

The remaining groups are used to construct the radial convergence diagram where time increases clockwise along the perimeter of the circle, beginning and ending at the topmost point. The outer ring shows the sequence of annotated sections as blocks, colored according to the schema in Fig. 39.2. Time markers are drawn as white lines and are placed at ten-second intervals along the ring, and section boundaries are drawn in black. Groups are represented using translucent ribbons connecting two time segments such that each ribbon edge connects the beginning of the first segment to the end of the second. If they link two instances of the same pattern, these ribbons assume the color of that pattern. Otherwise they are depicted in gray, which is also used to denote special sections such as introductions, interludes, transitions and codas.

As is shown in Fig. 39.11b, the RCD seamlessly combines information from the SSM with the piece's structural annotation, resulting in a rich and appealing visualization. The annotation in this example was manually generated on the basis of a musical score and propagated to audio recordings using the synchronization approach described in Sect. 39.3. Alternatively, the annotations can be automatically generated

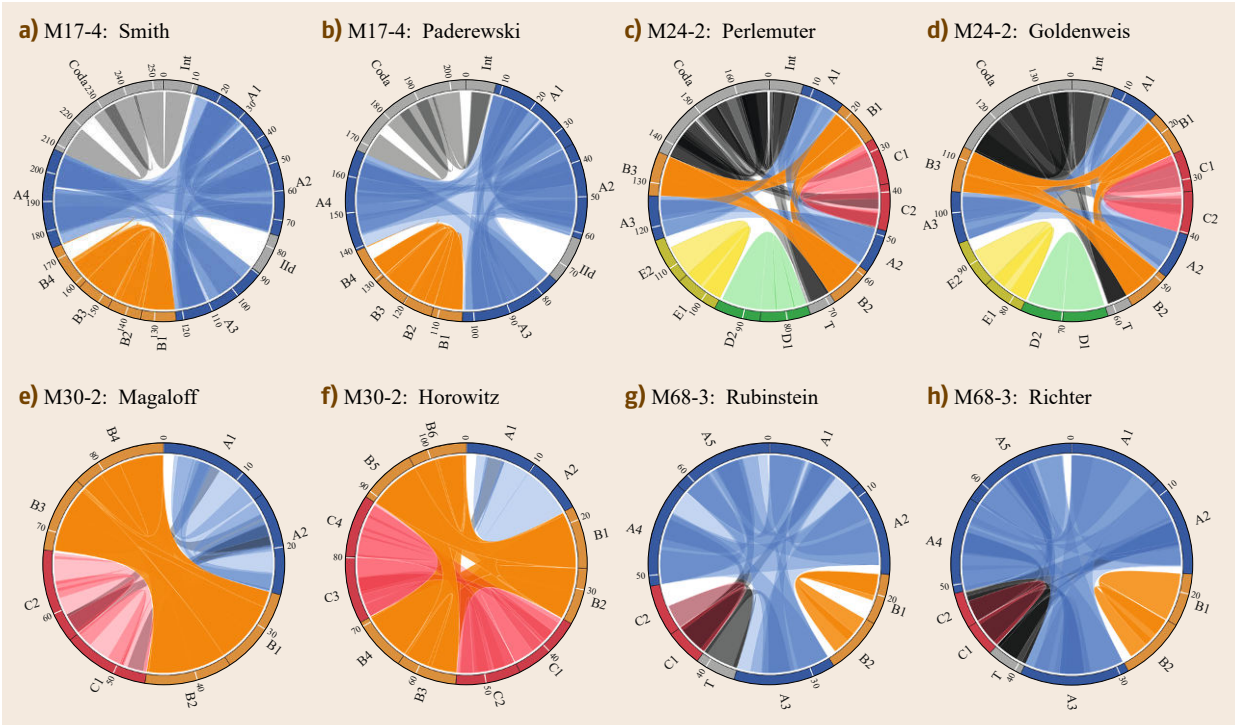


Fig. 39.12a–h Radial convergence diagrams for various Mazurka performances

by, e.g., analyzing the SSM (Sect. 39.4.2), or using the factorization approach that will be introduced in Sect. 39.5.

To illustrate the RCD's potential for qualitative analysis, Fig. 39.12 shows diagrams for two recordings of each of the following four Mazurkas: M17-4 in Fig. 39.12a,b, M24-2 in Fig. 39.12c,d, M30-2 in Fig. 39.12e,f, and M68-3 in Fig. 39.12g,h. The presence or absence of repetitive patterns results in distinctive shapes inside the diagrams that complement the information in the block-based structural annotations. These shapes highlight, for instance, the unique (nonrepetitive) character of the interlude (bottom-right gray segment) in M17-4, the significant overlap between introduction, transition and coda (gray sections) in M24-2, and the strong predominance of pattern $\mathcal{A}$ in the structure of M68-3.

Note that the gray ribbons linking different patterns depict relationships that are absent from the structural annotations. For example, they highlight the high similarity between instances of pattern C and the transition section in M68-3, both of which are dominated by a B$^\flat$ major chord as described previously in Sects. 39.1 and 39.2. The gray ribbons also connect the C and $\mathcal{T}$ sections with instances of the same chord in pattern $\mathcal{A}$, a link that results from the piece's only key modulation, in the transition, from F major to B$^\flat$ major.

The RCD representation can also be used to qualify stylistic differences in performance practice. Take for example one of the outliers identified in the analysis in Sect. 39.4.3. Figure 39.12f shows the unique structure of Vladimir Horowitz's performance of M30-2 when compared to performances that more closely follow the original score in Figs. 39.12e or 39.11b. Likewise, the varying-length gaps at the end of section C_2 (the final red segment) in M68-3, indicate a ritardando and an expressive pause in Rubinstein's performance (Fig. 39.12g), which is not present in Richter's (Fig. 39.12h).

Please note that an in-depth musicological analysis based on these diagrams is beyond the scope of this chapter. These simple observations are only intended to illustrate some of the capabilities of the RCD representation.

39.5 Automated Extraction of Repetitive Structures

Section 39.4.2 describes some general strategies for identifying repetitive structure within music recordings using self-similarity matrices. Example approaches include the clustering of diagonal elements in the SSM representation into sets of disjoint segments, and selecting the first of the most similar pair of segments, or choosing an arbitrary occurrence of the most frequent segment as a representative pattern. In this section we present an alternative approach based on matrix factorization, which jointly identifies repetitive patterns and derives a global structure segmentation from audio features.

39.5.1 Structure Analysis Using Matrix Factorization

In [39.56, 57] we propose a novel approach for the localization and extraction of repeating patterns in music audio. The idea underlying this method is that a song can be represented through repetitions of a small number of patterns in feature space. For example, recall that M68-3 has the form $\mathcal{A}_1\mathcal{A}_2\mathcal{B}_1\mathcal{B}_2\mathcal{A}_3\mathcal{T}\,C_1C_2\mathcal{A}_4\mathcal{A}_5$, i. e., the piece exhibits repetitions of three parts ($\mathcal{A}$, $\mathcal{B}$, and C) as well as a transition $\mathcal{T}$ that occurs only once. The only information needed to represent this piece are the feature sequences corresponding to each of the four patterns and the points in time of their occurrences.

This observation can be exploited to identify the main parts of a music recording and its overall temporal structure using an approach known as *shift-invariant probabilistic latent component analysis* (SI-PLCA). The process is illustrated in Fig. 39.13 for a performance of M68-3 by Gábor Csalog from 1996. The chromagram in Fig. 39.13a is approximated as the weighted sum of K ($K = 4$ in this example) components, each of which corresponds to a different part. Each component is further decomposed into a short chroma *basis pattern* (shown in Fig. 39.13b) and an activation function defining the location of each repetition of that pattern in the chromagram (shown in Fig. 39.13c). The peak positions in the activation function denote occurrences of the corresponding basis pattern in the feature sequence. Given a chromagram, this decomposition into basis patterns and activations can be computed iteratively using well-known optimization techniques [39.57].

There are many advantages to this model. First, it operates in a purely data-driven, unsupervised fashion: outside of a few functional parameters, the only prior information it requires is the length of the patterns L and the number of patterns K to extract. Second, the probabilistic formulation makes it straightforward to impose sparse priors on the distribution of basis patterns and mixing weights, enabling the model to learn optimal values for L and K. In addition, the framework can be easily extended to be invariant to key transpositions using a technique similar to that described in Sect. 39.4.2. Finally, the method jointly estimates the set of patterns and their activations, thus avoiding the need for heuristics for pattern extraction, grouping and selection. The versatility of the model is demonstrated in [39.57], where it is successfully applied to riff finding, meter identification, and segmentation of popular music.

39.5.2 Representative Patterns

The SI-PLCA model can be easily extended to extract patterns jointly from all available recordings of a piece by sharing the bases across recordings but using different activation functions for each one. This

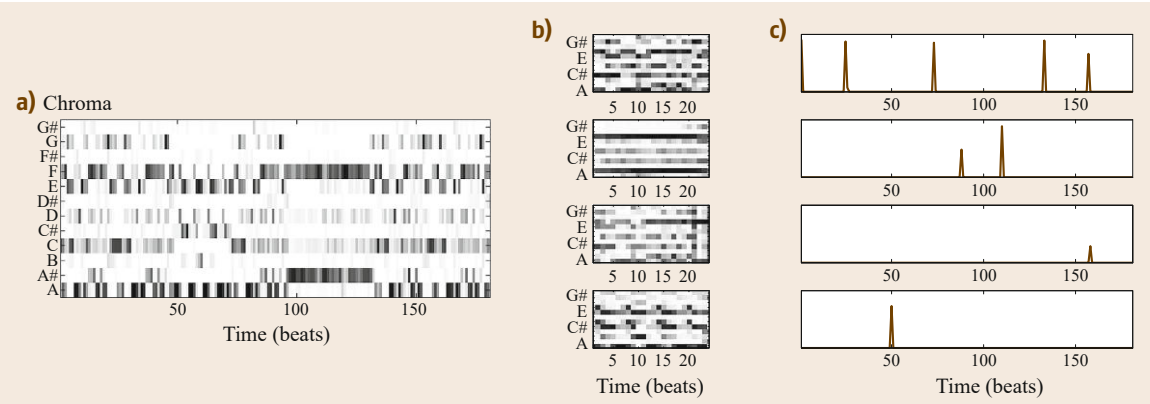


Fig. 39.13a–c Illustration of the SI-PLCA decomposition. **(a)** Chromagram. **(b)** Set of four basis patterns. **(c)** Respective activations in time

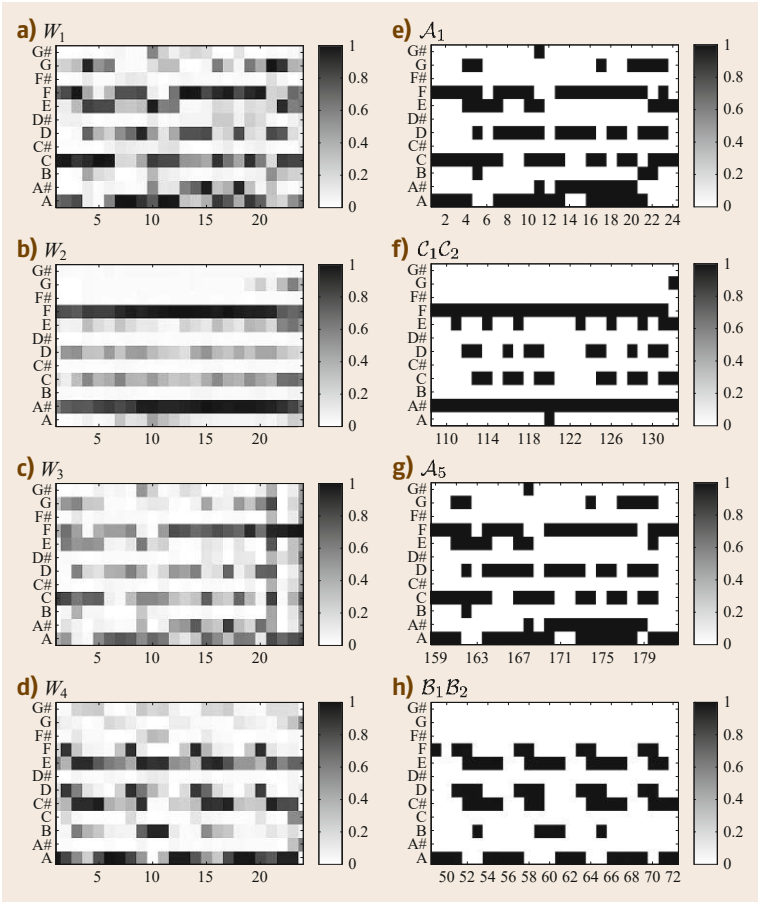


Fig. 39.14a–h Extracted patterns $W_k, k \in [1:4]$ for M68-3 and the corresponding MIDI-generated ground-truth patterns for the different parts. To facilitate comparison, **(f)** and **(h)** show both instances of the parts $\mathcal{C}$ and $\mathcal{B}$ respectively. The vertical axis shows time information given in beats. For the ideal patterns, the absolute beat numbers indicate the position of the pattern in the piece, see Fig. 39.2e

allows the model to identify the patterns of the piece rather than capture the idiosyncrasies of a single performance. In order to facilitate this cross-recording analysis we use beat-synchronous chromagrams, where features are averaged within beat segments [39.15]. This leads to a representation containing one feature vector per beat. We employ the manually annotated beat positions available for the Mazurka dataset to generate the chromagrams shown in this section.

Figure 39.14 shows patterns extracted for M68-3. The number of patterns in the decomposition is set to $K = 4$ and the pattern length is set to $L = 24$ beats, the length of $\mathcal{A}$, the longest part in the piece. The extracted patterns, denoted W_k, are shown in the left column of Fig. 39.14. The right column shows *ideal* chromagrams for the parts of M68-3, extracted from a symbolic (MIDI) representation. These sequences are paired with the basis that is most similar.

Upon close examination, it can be seen that W_1 matches every chord in $\mathcal{A}_1$, aside from some ambiguity in beats 21, 23 and 24. This is due to the alternation of endings of different instances of the $\mathcal{A}$ pattern: $\mathcal{A}_1$ and $\mathcal{A}_4$ finish with Gmaj-Cmaj-Cmaj-Cmaj chords at beats 21–24; while $\mathcal{A}_2$, $\mathcal{A}_3$ and $\mathcal{A}_5$ end with a Gmin-Cmaj-Fmaj-Fmaj sequence. Correspondingly, beat 21 shows energy in pitch classes G, B, $B^{\flat}/A^{\sharp}$ and D; while beats 23 and 24 show activity for classes A, C, E and F. Likewise, W_4 matches $\mathcal{B}_1\mathcal{B}_2$ almost perfectly, aside from a slight shift: W_4 starts in beat 2 of $\mathcal{B}_1$ and finishes in the first beat of $\mathcal{A}_3$, which contains a transitional C major seventh chord.

The melodic components of W_2 are also shifted by a couple of beats, although the transitional C dominant seventh chord at the end is well aligned. However, most of the energy of the matrix is concentrated in the $B^{\flat}$ fifth chord that dominates the $\mathcal{C}$ and $\mathcal{T}$ patterns. As a result, W_2 matches activations of both these patterns, as can be seen in the corresponding activation vectors in Fig. 39.13c. The unintended consequence of this is that the remaining pattern W_3, which must be assigned to some portion of the signal, absorbs a small amount of the energy from the last instance of pattern $\mathcal{A}$, mostly from the long F major chord at the end. W_3 is largely redundant with W_1, and could therefore be easily discarded. This is also indicated by the low value of the corresponding activation in Fig. 39.13c.

39.5.3 Segmentation Analysis

The work in [39.56] also describes a method for segmenting an audio signal based on the contribution of each basis to the chromagram. This is done by computing a temporal localization function ℓ_k from the extracted bases and activations that represents the energy contribution of each component to the overall chromagram at each point in time. Figure 39.15 shows the localization functions for the five Mazurkas in Fig. 39.2. These results are obtained for $L = 24$ beats and the optimal setting of K for each piece.

Figure 39.15e shows ℓ_k for M68-3, our running example. Given the similarity between patterns C and $\mathcal{T}$, discussed in previous sections, we have chosen to set $K = 3$, which removes the spurious component depicted in Fig. 39.14c. The resulting analysis shows a strong correspondence between the localization function and the annotated parts. The annotations for all $\mathcal{A}$ and $\mathcal{B}$ parts align perfectly with ℓ_2 and ℓ_1 respectively. ℓ_3 corresponds primarily to the C and $\mathcal{T}$ parts, and also makes a small contribution to $\mathcal{A}_3$.

As shown in Fig. 39.15c, the localization functions for M30-2 correlate closely with the manual annotation, given an optimal choice of $K = 3$. Similarly for M17-4 in Fig. 39.15a, there is high correlation between ℓ_k and the annotated parts. However, there is a constant shift in the segmentation that is mainly caused by the 12-beat long introduction. Additionally, the choice of $K = 3$ results in the grouping of the introduction and interlude sections (and some of the coda) with the $\mathcal{A}$ pattern.

The segmentation is less straightforward for M24-2 and M63-3, shown in Fig. 39.15b,d. In the case of M24-2, the choice of $K = 7$ closely matches the number of distinct parts in this piece, also visualized in the diagrams in Fig. 39.12c,d. As a result, C, $\mathcal{D}$, $\mathcal{E}$, the introduction and coda are slightly shifted but well segregated. However, the decomposition combines the contributions of $\mathcal{A}$ and $\mathcal{B}$, which always occur in sequence, into ℓ_2. Similarly, the ending of $\mathcal{B}$ and the transition are merged together in ℓ_4. For M63-3, the choice of $K = 6$ is larger than the number of parts in the annotation. This is a consequence of strong variations between the annotated $\mathcal{A}$ parts, where only the first 12 beats are common to all instances. These variations lead to a separation of the $\mathcal{A}$ parts into different components.

To further alleviate possible errors, we apply temporal smoothing to the localization functions, and select the maximal contributing pattern for each time position. This results in segmentation results comparable to state-of-the-art methods [39.56]. It is important to note that, while the results discussed in this section reveal a sensitivity to the setting of K, our research shows that it is possible to decrease this sensitivity by imposing additional sparsity constraints on the model parameters.

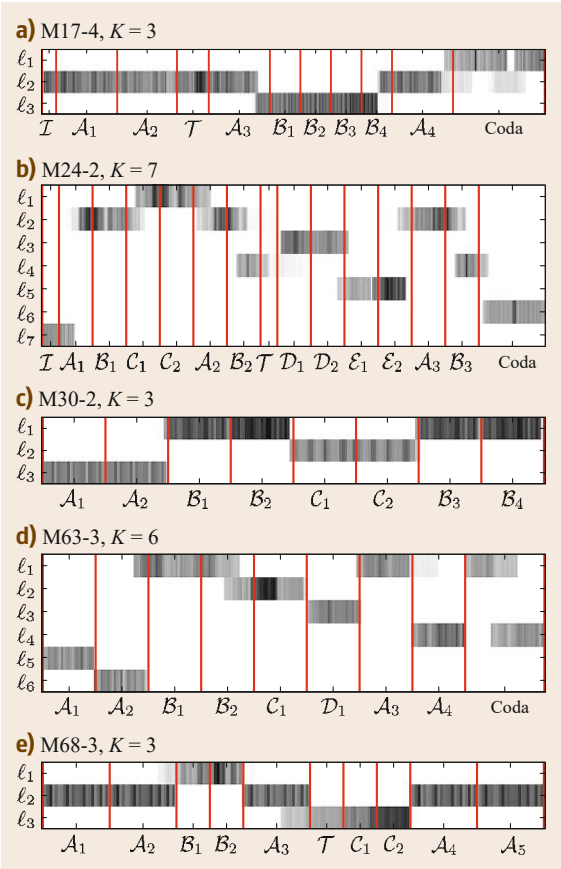


Fig. 39.15a–e Contribution ℓ_k of the basis patterns W_k to the different parts using $L = 24$ and the optimal setting of K for each of the five Mazurkas

It is also important to observe that the above analysis assumes that all patterns are of the same length. This is of course unrealistic for most music. When analyzing popular music this is partly alleviated via beat-synchronous analysis. However, extending this strategy to the analysis of piano music of the classical and romantic canon is problematic, because the variability introduced by expressive tempo changes negatively affects the performance of beat tracking systems. See [39.58] for an in-depth discussion of this problem in the context of the Mazurka dataset. The examples in this section have avoided such complications by making use of beat-synchronous chromagrams based on manually annotated beat positions. In order to analyze music for which such annotations are not available, we could, in the future, extend the SI-PLCA model to be invariant to tempo changes by allowing the basis patterns to be stretched in time.

39.6 Conclusions

In this chapter, we have reviewed the general principles behind several computational approaches to content-based analysis of recorded music and discussed their application to knowledge discovery. For example, we demonstrated how chroma-based audio features can be used as a robust midlevel representation for capturing harmonic information, and to compute similarity matrices that reveal recurrent patterns. By including temporal contextual information, we showed how these matrices can be enhanced to reveal repetitions in the presence of local tempo variations as occurring in expressive music recordings. Furthermore, we discussed how these techniques can be applied to music synchronization, to automate audio annotation and to realize novel user interfaces for inter- and intradocument music navigation. Finally, we demonstrated how structure-based similarity measures can be used to classify music collections based on their global structural form, and how matrix factorization techniques can be used to identify the most representative building blocks of a recorded performance.

Our aim was to highlight the implications of the various approaches for the analysis and understanding of recorded music, and to demonstrate how they can empower the work of music professionals and scholars. In addition we showed how these tools facilitate novel modes of interaction with digital content that could be integrated with music distribution services as ways to enhance the listeners' understanding and appreciation of music. Throughout the text, we illustrated the potential of content-based analysis via many concrete and intuitive examples taken from a collection of recorded piano performances of Chopin's Mazurkas. However, the presented techniques and underlying principles are applicable to the analysis of a wide range of tonal music and, with appropriate modifications of the feature representation, to percussive music or other types of time-dependent structured multimedia data.

Of course, this chapter has only covered a small part of the breadth of techniques and problems in the area of automated music processing. There are numerous challenges and open issues for future research. For instance, many of the examples shown here rely, to some extent, on the use of manually generated expert data such as beat positions or structural annotations. Despite significant research efforts, the automated generation of such annotations using purely content-based analysis techniques still poses challenging and open research problems for large parts of the existing digital music catalog. State-of-the-art algorithms often lack the robustness, accuracy and reliability needed for many music analysis applications. This is especially true for highly expressive music recordings such as the Mazurka performances described in this chapter, which exhibit subtle differences and variations in tempo, articulation, and note execution.

As the amount of digitally available music-related data grows, so does the need for efficient approaches that can scale up to the processing of millions of tracks. As automated procedures become more and more powerful they tend to gain computational complexity, making scalability an increasingly important issue. In addition to improvements in robustness and accuracy, issues related to time and memory efficiency will progressively come into the focus of future research.

Acknowledgments. This material is based upon work supported by the National Science Foundation, under grant IIS-0844654, and the Cluster of Excellence on Multimodal Computing and Interaction at Saarland University. The authors would like to thank Craig Sapp for kindly providing access to the Mazurka dataset and beat annotations.

References

39.1 M.A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, M. Slaney: Content-based music information retrieval: Current directions and future challenges, Proc. IEEE **96**(4), 668–696 (2008)

39.2 M. Slaney: Web-scale multimedia analysis: Does content matter?, Multimed. IEEE **18**(2), 12–15 (2011)

39.3 H. Schenker: *Der freie Satz* (Universal, Vienna 1935)

39.4 A. Ockelford: *Repetition in Music: Theoretical and Metatheoretical Perspectives* (Ashgate, London 2005)

39.5 D. Huron: *Sweet Anticipation: Music and the Psychology of Expectation* (MIT Press, Cambridge 2006)

39.6 M.J. Bruderer, M. McKinney, A. Kohlrausch: Structural boundary perception in popular music. In: *Proc. Int. Conf. Music Inf. Retr. (ISMIR), Victoria* (2006) pp. 198–201

39.7 G. Peeters, E. Deruty: Is music structure annotation multi-dimensional? A proposal for robust local music annotation. In: *Proc. 3rd Workshop Learn. Semant. Audio Signals, Graz* (2009) pp. 75–90

39.8 The AHRC Research Centre for the History and Analysis of Recorded Music: Website of the Mazurka Project, http://www.mazurka.org.uk/

39.9 C.S. Sapp: Comparative analysis of multiple musical performances. In: *Proc. Int. Conf. Music Inf. Retr. (ISMIR), Vienna* (2007) pp. 497–500

39.10 C.S. Sapp: Hybrid numeric/rank similarity metrics. In: *Proc. Int. Conf. Music Inf. Retr. (ISMIR), Philadelphia* (2008) pp. 501–506

39.11 E. Pampalk: *Computational Models of Music Similarity and Their Application to Music Information Retrieval*, Ph.D. Thesis (Vienna University of Technology, Vienna 2006)

39.12 S. Essid: *Classification Automatique des Signaux Audio-Fréquences: Reconnaissance des Instruments de Musique*, Ph.D. Thesis (Université Pierre et Marie Curie, Paris 2005)

39.13 G. Peeters: A large set of audio features for sound description (similarity and classification) in the CUIDADO project, http://recherche.ircam.fr/anasyn/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf (Ircam, Analyis/Synthesis Team, Paris 2004), version 1.0

39.14 A. Sheh, D.P.W. Ellis: Chord segmentation and recognition using EM-trained hidden Markov models. In: *Proc. Int. Conf. Music Inf. Retr. (ISMIR), Baltimore* (2003)

39.15 D.P.W. Ellis, G.E. Poliner: Identifying 'cover songs' with chroma features and dynamic programming beat tracking. In: *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), Honolulu* (2007)

39.16 J. Serrà, E. Gómez, P. Herrera, X. Serra: Chroma binary similarity and local alignment applied to cover song identification, IEEE Trans. Audio Speech Lang. Process. **16**, 1138–1151 (2008)

39.17 E. Gómez: *Tonal Description of Music Audio Signals*, Ph.D. Thesis (Universitat Pompeu Fabra, Barcelona 2006)

39.18 M. Mauch, K. Noland, S. Dixon: Using musical structure to enhance automatic chord transcription. In: *Proc. Int. Conf. Music Inf. Retr. (ISMIR), Kobe* (2009) pp. 231–236

39.19 M. Müller: *Information Retrieval for Music and Motion* (Springer, Berlin, Heidelberg 2007)

39.20 R.N. Shepard: Circularity in judgments of relative pitch, J. Acoust. Soc. Am. **36**(12), 2346–2353 (1964)

39.21 T. Fujishima: Realtime chord recognition of musical sound: A system using common lisp music. In: *Proc. ICMC, Beijing* (1999) pp. 464–467

39.22 M. Mauch, S. Dixon: Approximate note transcription for the improved identification of difficult chords. In: *Proc. 11th Int. Soc. Music Inf. Retr. Conf. (ISMIR), Utrecht* (2010) pp. 135–140

39.23 M. Müller, S. Ewert: Towards timbre-invariant audio features for harmony-based music, IEEE Trans. Audio Speech Lang. Process. **18**(3), 649–662 (2010)

39.24 I.T. Jolliffe: *Principal Component Analysis* (Springer, New York 2002)

39.25 N. Hu, R.B. Dannenberg, G. Tzanetakis: Polyphonic audio matching and alignment for music retrieval. In: *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA), New Paltz* (2003)

39.26 S. Ewert, M. Müller, P. Grosche: High resolution audio synchronization using chroma onset features. In: *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), Taipei* (2009) pp. 1869–1872

39.27 C. Fremerey, F. Kurth, M. Müller, M. Clausen: A demonstration of the SyncPlayer system. In: *Proc. 8th Int. Conf. Music Inf. Retr. (ISMIR), Vienna* (2007) pp. 131–132

39.28 D. Damm, C. Fremerey, F. Kurth, M. Müller, M. Clausen: Multimodal presentation and browsing of music. In: *Proc. 10th Int. Conf. Multimodal Interfaces (ICMI), Chania* (2008) pp. 205–208

39.29 M. Müller, V. Konz, N. Jiang, Z. Zuo: A multi-perspective user interface for music signal analysis. In: *Proc. Int. Computer Music Conf. (ICMC), Huddersfield* (2011)

39.30 M. Goto: A chorus section detection method for musical audio signals and its application to a music listening station, IEEE Trans. Audio Speech Lang. Process. **14**(5), 1783–1794 (2006)

39.31 J. Foote: Visualizing music and audio using self-similarity. In: *Proc. ACM Int. Conf. Multimed., Orlando* (1999) pp. 77–80

39.32 J. Foote: Automatic audio segmentation using a measure of audio novelty. In: *Proc. IEEE Int. Conf. Multimed. Expo (ICME), New York* (2000) pp. 452–455

39.33 G. Peeters: Sequence representation of music structure using higher-order similarity matrix and maximum-likelihood approach. In: *Proc. Int. Conf. Music Inf. Retr. (ISMIR), Vienna* (2007) pp. 35–40

39.34 M. Goto: A chorus-section detecting method for musical audio signals. In: *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), Hong Kong* (2003) pp. 437–440

39.35 M.A. Bartsch, G.H. Wakefield: Audio thumbnailing of popular music using chroma-based representations, IEEE Trans. Multimed. **7**(1), 96–104 (2005)

39.36 J. Paulus, M. Müller, A. Klapuri: Audio-based music structure analysis. In: *Proc. 11th Int. Conf. Music Inf. Retr. (ISMIR), Utrecht* (2010) pp. 625–636

39.37 N. Marwan, M.C. Romano, M. Thiel, J. Kurths: Recurrence plots for the analysis of complex systems, Phys. Rep. **438**(5/6), 237–329 (2007)

39.38 G. Tzanetakis, P. Cook: Musical genre classification of audio signals, IEEE Trans. Speech Audio Process. **10**(5), 293–302 (2002)

39.39 M. Slaney, M. Casey: Locality sensitive hashing for finding nearest neighbours, IEEE Signal Process. Mag. **25**(2), 128–131 (2008)

39.40 J. Serrà, X. Serra, R.G. Andrzejak: Cross recurrence quantification for cover song identification, New J. Phys. **11**(9), 093017 (2009)

39.41 T. Cho, J. Forsyth, L. Kang, J.P. Bello: Time-varying delay effects based on recurrence plots. In: *Proc. 14th Int. Conf. Digit. Audio Eff. (DAFx), Paris* (2011)

39.42 M. Müller, F. Kurth: Enhancing similarity matrices for music audio analysis. In: *Proc. 32nd Int. Conf. Acoust. Speech Signal Process. (ICASSP), Toulouse* (2006) pp. 437–440

39.43 M. Müller, M. Clausen: Transposition-invariant self-similarity matrices. In: *Proc. 8th Int. Conf. Music Inf. Retr. (ISMIR), Vienna* (2007) pp. 47–50

39.44 R.B. Dannenberg, M. Goto: Music structure analysis from acoustic signals. In: *Handbook of Signal*

Processing in Acoustics, Vol. 1, ed. by D. Havelock, S. Kuwano, M. Vorländer (Springer, New York 2008) pp. 305–331
39.45 T. Izumitani, K. Kashino: A robust musical audio search method based on diagonal dynamic programming matching of self-similarity matrices. In: *Proc. 9th Int. Conf. Music Inf. Retr. (ISMIR), Philadelphia* (2008) pp. 609–613
39.46 J.P. Bello: Measuring structural similarity in music, IEEE Trans. Audio Speech Lang. Process. **19**(7), 2013–2025 (2011)
39.47 W. Xie, N.V. Sahinidis: A Branch-and-reduce algorithm for the contact map overlap problem, Res. Comput. Biol. (RECOMB 2006), Lect. Notes Bioinform. **3909**, 516–529 (2006)
39.48 N. Krasnogor, D.A. Pelta: Measuring the similarity of protein structures by means of the universal similarity metric, Bioinformatics **20**(7), 1015–1021 (2004)
39.49 J.P. Bello: Grouping recorded music by structural similarity. In: *Proc. Int. Conf. Music Inf. Retr. (ISMIR), Kobe* (2009)
39.50 I. Borg, P. Groenen: *Modern Multidimensional Scaling* (Springer, New York 1997)
39.51 P. Toiviainen: Visualization of tonal content with self-organizing maps and self-similarity matrices, Comput. Entertain. **3**(4), 1–10 (2005)
39.52 K.W. Church, J.I. Helfman: Dotplot: A program for exploring self-similarity in millions of lines for text and code, J. Am. Stat. Assoc., Inst. Math. Stat. Interface Found. North Am. **2**(2), 153–174 (1993)
39.53 E.L.L. Sonnhammer, J.C. Wootton: Dynamic contact maps of protein structures, J. Mol. Graph. Modell. **16**(33), 1–5 (1998)
39.54 M. Lima: VC blog on *Radial Convergence*, http://www.visualcomplexity.com/vc/blog/?p=876 (2011)
39.55 M.I. Krzywinski, J.E. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones, M.A. Marra: Circos: An information aesthetic for comparative genomics, Genome Res. **19**(9), 1639–1645 (2009)
39.56 R.J. Weiss, J.P. Bello: Identifying repeated patterns in music using sparse convolutive non-negative matrix factorization. In: *Proc. Int. Conf. Music Inf. Retr. (ISMIR), Utrecht* (2010) pp. 123–128
39.57 R.J. Weiss, J.P. Bello: Unsupervised discovery of temporal structure in music, IEEE J. Sel. Top. Signal Process. **5**(6), 1240–1251 (2011)
39.58 P. Grosche, M. Müller, C.S. Sapp: What makes beat tracking difficult? A case study on Chopin Mazurkas. In: *Proc. 11th Int. Conf. Music Inf. Retr. (ISMIR), Utrecht* (2010) pp. 649–654

40. Hearing Aids and Music: Some Theoretical and Practical Issues

Marshall Chasin, Neil S. Hockley

This chapter focuses on the hearing assessment of musicians as well as how to recommend and specify the exact parameters for hearing aid amplification for hard-of-hearing people who either play musical instruments or merely like to listen to music. Much of this is based on the differences between the acoustic features of music and of speech. Music is typically listened to, or played at, a higher sound level than speech and there are some spectral and temporal differences between music and speech that have implications for differing electro-acoustic hearing-aid technologies for the two types of input. This involves a discussion of some hearing aid technologies best suited to amplified music as well as some clinical strategies for the hearing health care professional to optimize hearing aids for music as an input.

The key limitation concerning the capability of current digital hearing aids to accommodate the more intense elements of music is the analog-to-digital (A/D) converter. Expending research and development efforts on the other elements within the hearing aid will not really improve the fidelity of music unless the limitations of the A/D converter are first solved. The topic of music as an input to hearing aids and the technologies that are available is a rapidly changing one. New technologies are on the horizon such as better A/D converters that may be implemented by various manufacturers.

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_40

The study of systematic musicology is concerned with how human beings hear and process complex auditory signals. What happens, however, when there is impairment in the auditory system that disrupts an individual's ability to perceive and potentially process acoustic information? In such cases the individual may need to seek help and this may or may not require the use of hearing aids. This chapter will address the topic of hearing aids and musicians.

Hearing aids are primarily designed to amplify speech to address issues of communication for those individuals whose hearing loss is not of a degree that requires the use of a cochlear implant. Issues of communication after all are why hearing-impaired individuals, or in many cases their families, seek assistance [40.1–3]. The amplification of music is often perceived to be less important by the hearing-aid industry [40.4]. When music is a person's livelihood, lifelong passion, or even a new interest then the priority given to music is increased. This chapter will focus on what can be done with hearing aids to help these individuals. There will be a focus on the assessment of musicians, followed by a brief discussion of the acoustic differences between speech and music. This will be followed by a discussion of the technical solutions that are currently available in hearing aids as well as some practical suggestions about what can be done by the clinician in order to optimize the hearing-aid fitting for music.

40.1 Assessment of Musicians

During a typical diagnostic audiological assessment, a number of tests are performed to examine the capabilities and limitations of an individual's auditory system. These tests are designed to establish the location and type of pathology within the auditory pathways, as well as to provide a reliable measure of the degree of this pathology. These tests, however, are best suited to assess problems arising from the peripheral auditory pathway.

The peripheral auditory pathway consists of the outer ear, the middle ear, and the inner ear. Any subsequent neural involvement is considered to be the central auditory pathway. While there are some tests and assessment procedures that can assess any dysfunction within the central auditory pathway, these tend to be more complex and are not routinely performed in a typical clinical setting. Some of these tests are performed as part of a behavioral central auditory processing (or CAP) assessment [40.5, 6]. It is also possible to evaluate any neural involvement with more objective means that include electrophysiological techniques such as the auditory brainstem response (ABR) [40.7] and imaging assessment procedures such as magnetic resonance imaging (MRI) [40.8].

This chapter will only deal with the assessment of peripheral auditory pathology, specifically conductive and sensory issues. And to be more precise, we are mostly interested in cochlear pathology (the primary peripheral sensory location) as a result of music and/or noise exposure along with presbycusis, which is hearing loss associated with the aging process. Neural issues and other more central problems will not be covered. The reader is referred to [40.9] for an introduction to neural pathology and its effects.

40.2 Peripheral Sensory Hearing Loss

Depending on the clinic, routine audiological testing can include air conducted pure-tone testing under headphones or insert earphones, bone conducted pure-tone testing, speech recognition thresholds, word recognition scores, and immittance testing. With the exception of immittance testing, which is an objective physiological measurement battery that provides information about the healthy functioning of the middle ear [40.10], the other tests involve the active participation of the individual. The classic manifestation of a sensory type of hearing loss includes normal immittance results that indicate normal middle ear function, and pure-tone thresholds that are the same (within ten or 15 dB) regardless of whether they have been presented via air or bone conduction. All of these tests are used to rule out various locations of hearing loss but with musicians, it is the cochlear function that interests us most [40.11].

Musicians are exposed to a sound pressure level (SPL) that can potentially cause long-term hearing loss as part of their profession. This has been shown not only for players of amplified music (e.g., [40.12–15]) but also for those playing classical music (e.g., [40.16–23]). Like industrial noise exposure, music exposure will have its greatest effect in the cochlea, and like noise exposure, the greatest effect will be in the 3000–6000 Hz region [40.24]. In most cases it is quite difficult to de-

termine whether a sensory hearing loss can be attributed to noise or to music. A case history is typically the most important differentiating element where the clinician can document the duration and type of exposure to intense sound pressure levels that the individual has been exposed to. Figure 40.1 shows an audiogram showing the poorest pure-tone sensitivity at 4000 Hz. This is fairly typical of sensory peripheral hearing losses from industrial noise exposure and also from long-term music exposure.

In addition to routine audiological testing, two other types of specialized tests can be used with musicians: the assessment of cochlear dead regions [40.25] and otoacoustic emission (OAE) testing [40.26]; these two techniques will be discussed next.

Dead regions refer to areas of significant cochlear damage to the inner hair cells where amplification of any sort does not provide any benefits for the individual and filtering or transposing amplified acoustic energy away from this location in the cochlea would be strongly advised [40.27]. In cases of normal or mild sensory dysfunction, the assessment of cochlear dead regions is not normally performed. There are a number of assessment paradigms to determine if cochlear dead regions exist but many of these approaches, such as the use of psychophysical tuning curves [40.28], are time consuming and are therefore impractical to use in a clinical setting [40.29]. The threshold equalizing noise test in dB HL (TEN (HL)) [40.30, 31] is a much less time-consuming and easy-to-administer test, which involves the detection of a sinusoidal tone presented within a specially constructed noise – the threshold equalizing noise (TEN). An alternative assessment procedure to the TEN test, which many musicians and other individuals who might have a significant sensory hearing loss seem to prefer, is to have them play the notes on a piano (or electronic keyboard) in a slowly ascending fashion. The task is to determine whether any two adjacent notes on the piano keyboard are considered to be the *same* or *different* in pitch. If two adjacent keys are judged to be the same, then this is potentially an area of significant cochlear pathology i. e., a cochlear dead region, and therefore the use of amplification should be minimized in this frequency region.

Another assessment technique that can be routinely performed on musicians in the clinic is otoacoustic emission (OAE) testing. This procedure assesses the neural feedback route, specifically that which involves the outer hair cells in the cochlea. Changes in the level of the OAE can provide information regarding cochlear (outer hair cell) pathology. There are a number of different types of OAEs that can be measured and are defined based on how they are elicited.

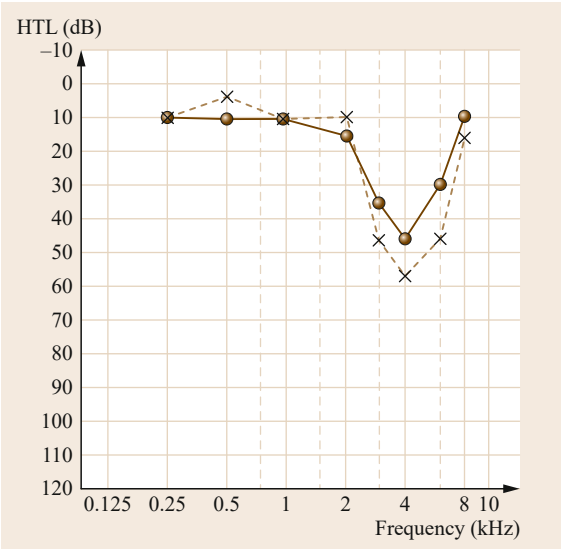


Fig. 40.1 Audiogram showing a typical music (and noise)-induced hearing loss (reprinted with permission from the Musicians' Clinics of Canada)

The first type is spontaneous OAEs (SPOAE), which are not the result of a specific stimuli, but are just a sign of a normal physiological state [40.32]. Transient evoked OAEs (TEOAE) are elicited by brief signals such as a broadband click [40.33]. Distortion product OAEs (DPOAE) are the result of an intermodulation distortion product produced by the cochlea when presented with two simultaneous pure tones that are close in frequency. The distortion product signal produced by the cochlea is a tone that is not present in the eliciting stimulus, hence the use of the term distortion [40.34]. Each OAE measurement type has its own advantages and disadvantages, depending on the desired goal of the assessment, which is either a basic assessment of a normal hearing individual or differential diagnosis with regards to abnormal hearing. In many cases, changes in OAE thresholds are observed prior to changes in pure-tone thresholds. These changes can be viewed as an early warning sign and the results could be used to counsel the musician regarding the use of appropriate hearing protection [40.26]. An inherent feature of OAE testing is the large intersubject variability in responses, which would limit the large-scale use of this type of testing in a large population. The variability could be so large as to obliterate any real changes in the mean results of the OAEs. However, longitudinal changes within an individual i. e., intrasubject variability, in combination with an established history of intense sound exposure, can provide important indications of any early changes in cochlear function.

40.3 Direct Assessment of Music with a Peripheral Hearing Loss

So far in our look at hearing loss and musicians there has been no real discussion about using music as a stimulus to assess the auditory system. Currently, to the best of the authors' knowledge, there are no dedicated assessment tools used widely on a routine basis within the clinic to assess the perceptual needs of hearing-impaired individuals wearing hearing aids to listen to music. There are a number of studies that have looked at the perception of music through cochlear implants [40.35–38] and there have been a number of clinicians who have developed their own methods to program hearing aids based on their own musical experience [40.39]. Some work has been made to develop assessment tools to evaluate music for the hearing-aid user, or potential user [40.40, 41]. *Uys* and *van Dijk* [40.42] developed a comprehensive music perception test (MPT) with a number of subscales that directly address discrimination, and identification of musical elements. The subscale categories include rhythm, timbre, pitch, and melody. Within each subscale there are specific tests to examine the number of instruments heard, the identification of melodies, or pitch discrimination using short melodic sequences. It is hoped that further development of the MPT continues and that the use of this and other tools gain acceptance within clinical settings.

40.4 Acoustic Properties of Music versus Speech

So far in this discussion we have looked at the assessment of the individual musician's auditory system. It is now important to look at how music and speech differ so that we can further discuss the use of hearing aids with musicians.

There are three primary differences between music as an input and speech as an input to a hearing aid. These are the sound level of the music, the shape of the spectra, and the crest factor [40.43]. Each of these three differences have ramifications for the electro-acoustic design of the hearing aid, the selection of the hearing aid for the individual end user, and lastly, the programming of the various software parameters such as compression within the hearing aids.

Before delving into the differences between speech and music, we must first recognize that there are some primary similarities between speech and music. Both are periodic in nature, with the possible exception of percussive music and certain linguistic obstruents such as fricatives, affricates, and stops. There can be a rich harmonic structure where higher frequency energy is found at integer multiples of the fundamental frequency. Speech energy can have a wide bandwidth, with linguistic sonorants (e.g., vowels and nasals) having most of their speech energy in the low- to mid-frequency regions and higher frequency energy deriving from the linguistic obstruents (e.g., fricatives, affricates, and stops). Similarly, music can have significant fundamental energy in the lower frequency region and higher frequency harmonic information in the higher frequency region.

Music can be generated by a number of instruments including the vocal tract. The sound levels of instrumental music are not limited by the same slow-moving neurology and musculature found in and around the human vocal tract. Music can be percussive with a resulting broadly tuned, high-frequency biased spectrum, and music can be quite vocal-like such as a violin. Both the human vocal tract and the violin function acoustically as one-half wavelength resonators and, depending on the note being played, can have similar spectra.

40.4.1 Sound Level Differences

One of the primary differences between speech and music lies in their spectral levels. Music tends to be much more intense than speech and this includes even quieter forms of acoustic folk music. Speech, because of the inherent physical limitations of the human vocal mechanism, is typically limited to sound levels below about 80 dB(A) (with peaks on the order of 90 dB SPL) [40.44]. The dB(A) scale is used to account for the perceptual abilities of the human auditory system. This is in contrast to the physical sound pressure level (SPL). Filters for dB(A) are found on most commercially sound level meters and are defined according to an IEC standard [40.45]. The highest sound level of speech in any language of the world is the low back vowel /a/ as in 'father'. Because the vocal tract is wide open during the articulation of this vowel, there is minimal damping and minimal loss of energy.

In contrast, music can be in excess of 120 dB(A) (with peaks being on the order of 140 dB SPL). Even *quiet* orchestral music can easily reach sound levels in excess of 110 dB(A). Table 40.1 shows some data for sound levels of musical instruments (adapted from [40.46]). These data are based on the assessment of over 1000 musical instruments from a distance of 3 m in the horizontal plane. The top and bottom 25th percentile data has been removed.

Table 40.1 Average sound levels of a number of musical instruments measured from 3 m. Also given is the sound level for the violin measured near the left ear of the players (adapted from [40.46], used with permission)

Musical instrument	dB(A) ranges measured from 3 m
Cello	80–104
Clarinet	68–82
Flute	92–105
Trombone	90–106
Violin	80–90
Violin (near left ear)	85–105
Trumpet	88–108

40.4.2 Crest Factor Differences

Another element that differentiates human speech from many musical instruments is the crest factor. This is the difference in decibels between the instantaneous peak of a signal and its average or long term RMS (root mean square) value.

There is a significant loss of energy during speech production for the highly damped human vocal tract. Acoustic losses typically occur as a result of small openings, or soft highly damped structures such as the tongue, lips, and cheeks [40.47]. Human speech is characterized by all of these things. Additionally, different sounds are articulated with differing positions of the tongue and lips. Opening of the small port at the back of the mouth to the nasal cavity (velopharyngeal port) allows for the articulation of nasal sounds (e.g., /m/ and /n/). In short, human speech has a highly damped energy spectrum such that the levels of the instantaneous peaks are significantly reduced. The resulting crest factor is therefore lower, typically on the order of 12 dB. That is, with speech the difference between the average RMS and any peak is roughly 12 dB.

In contrast, for musical instruments we are dealing with mechanical and acoustical systems that have a relatively low level of damping. A violin or a trumpet does not have soft walls, nor does it have a narrow opening. Subsequently the crest factor of the sound energy emanating from these instruments tends to be higher than that for speech – typically on the order of 18–20 dB. That is, with many musical instruments the difference between the average RMS playing level and any peak is roughly 18–20 dB.

If the playing level is examined in combination with the crest factor, the peaks of music are much higher in level than even shouted speech. As discussed, speech is associated typically with a 12 dB crest factor while instrumental music is associated with an 18–20 dB crest factor.

The crest factor calculation requires some assumptions. The primary one relates to the selection of the bandwidth of the analyzing window. Whenever the instantaneous peak is assessed, which is one part of the crest factor calculation, it is necessary to be able to assess it in a well-defined manner. The hearing-aid industry frequently uses a window of analysis of 125 ms (1/8 of a second). This value was derived initially from work that was undertaken in the 1940s [40.49, 50], and this work has been replicated [40.51, 52]. The choice of a 125 ms window of analysis is related to the temporal integration constraints of the human cochlea, and so it can be referred to as a *time constant* of the human cochlea.

When we discuss the crest factor, with regards to hearing-aid amplification, we are not talking about the characteristics of cochlear time constants. The input to a hearing aid is the microphone and this is before this amplified sound reaches the cochlea. The microphone and the associated digital network that follows it can be thought of as the sound detector. The system, colloquially referred to as the front end of a hearing aid [40.53], in most cases utilizes a 16 bit digital architecture that has a theoretical upper limit to its dynamic range of 96 dB. That is, the range of sounds that can be integrated and transduced through the analog-to-digital (A/D) converter is 96 dB. Because of some engineering decisions, this range can be less than the 96 dB limit and effective dynamic ranges of only 80–85 dB have been found with current technology. This is more than adequate for the processing of speech, however, we need to address the question of what is needed for music?

Of importance is that this front end of a hearing aid is not limited to the 125 ms time constant of the cochlea; it is merely a microphone, A/D converter, and associated components. Table 40.2 shows some data

Table 40.2 With different length windows of analysis and for one music sample, the difference between the RMS of the signal and its instantaneous peak (crest factor) is given. For shorter analysis windows the instantaneous peak is higher than for longer windows of analysis with a resulting higher crest factor. Adobe Audition 1.0 (San Jose California) was used to calculate the difference between the average value of the signal and the instantaneous peak. Used with permission (after [40.48])

Analysis window (ms)	500	400	300	200	125	100	50	25
Crest factor (dB)	16.45	16.43	16.45	16.44	16.45	18.22	21.68	21.68

(from [40.54]) of how the crest factor may change as a function of the analysis time window for the same signal. The only parameter that changes in Table 40.2 is the window of analysis and not the signal.

Table 40.2 clearly demonstrates that for the same signal, the crest factor (specifically the instantaneous peak component) increases significantly as the window of analysis becomes shorter. The assumption of speech having a 12 dB crest factor (at the level of the microphone and input to the A/D converter) is therefore potentially erroneous. Taking these higher crest factor values implies that even the more intense elements of speech, especially the level of an individual's own voice, may have instantaneous peaks that are in excess of what current hearing aids can handle. This is an input-related phenomenon and has nothing to do with the output of the hearing aid or the cochlear function of an individual hearing-aid user.

40.5 Some Strategies to Handle the More Intense Inputs of Music

Given the limitations that most currently available digital hearing aids have in handling the more intense inputs that are characteristic of many forms of music, there are a number of strategies that a clinician can use to alter the electro-acoustic characteristics of hearing aids that otherwise work quite well for speech. That is, given an individual who already has hearing aids that work well for speech, what are some clinical modifications that can be implemented to improve the hearing aids for the listening and playing of music? Four clinical strategies that can improve the usability of hearing aids for music will now be discussed.

40.5.1 Clinical Strategy #1: Reduce the Input to the Hearing Aid, and if Necessary, Increase the Volume

This strategy is based on the idea of bending over slightly to avoid the A/D converter ceiling. Recall that many hearing aids have a front end that has an operating range that cannot handle overly intense inputs. If the input was reduced in level, and then is compensated for later on in the circuitry with an increase made via the hearing-aid volume control, then this may be a more optimal solution than the converse, where the input is left high and the volume control is reduced. In both cases, the output to the individual's ear is the same; it is just that the first case has less front-end distortion associated with it.

There are other clinical approaches that are based on this same approach. For example, an assistive listening device (with its own volume control) can be coupled either wirelessly or via a direct audio input route to the hearing aid. In this scenario, the input to the hearing aid's A/D converter is at a quieter level thereby providing the hearing aid with a signal that is more within its front-end operating range. The assistive listening device can be as simple as a remote microphone with (reduced) volume or a wireless instrument such as a frequency modulation (FM) listening system, again with the volume turned down.

Another scenario to consider is that when riding in a car or other vehicle with an audio system, again, turn down the volume of the audio system, and if necessary, increase the volume of the hearing aid, rather than the converse.

40.5.2 Clinical Strategy #2: Removal of the Hearing Aid for Music

While this may seem counterintuitive, in some situations it may be best to simply remove the hearing aids and listen to, or play the music, unaided. This strategy is based on the finding that loudness growth in our auditory system is not linear and tends to have a more gradual increase with more intense inputs [40.55]. A person who has a moderate sensorineural hearing loss may require 30 dB of gain for soft level inputs (55 dB), 25 dB of gain for medium level inputs (70 dB), and perhaps only 15–20 dB of gain for loud inputs (80 dB). These are the typical ranges of speech. However, music, even quiet music, can have playing levels that are typically much more intense. This same individual, who requires 30 dB of gain for soft speech levels, may require no amplification, or only a couple of decibels of amplification for louder music. Table 40.3 shows the amount of gain prescribed at 1000 Hz by the FIG6 fitting formula [40.56] as a function of how intense the input to the hearing aid is. At high input levels such as with many forms of music very little hearing-aid amplification may be required, if at all. It can therefore be concluded that, in some cases the hearing instrument can and should be removed for music.

40.5.3 Clinical Strategy #3: Use a (Tape) Covering of the Hearing-Aid Microphone

This may seem like an incredibly *low-tech* solution and indeed it is. Placing several layers of household adhesive tape over the hearing-aid microphone(s) will serve to reduce the microphone sensitivity. The exact number

Table 40.3 Calculated amounts of gain required for a given hearing loss at 1000 Hz (column 1) (based on FIG6 after [40.56]). For average levels of music (95 dB(A)) inputs, virtually no amplification may be required even for very significant hearing losses (after [40.48], used with permission)

dB HL at 1000 Hz	65 dB input	80 dB input	95 dB input
15	0	0	0
25	2	1	0
35	8	4	0
45	14	7	0
55	20	10	1
65	28	15	2
75	36	20	3
85	44	24	4

of layers of tape will have to be experimented with because different brands use different thicknesses of tape, however typically three or four layers will suffice. As far as the front end of the hearing aid is concerned, the input is now 10–12 dB less intense and in many cases will be within the optimal operating range of the A/D converter. The user may need to adjust the volume control upwards slightly, depending on the level of the modified music input. This strategy provides for a flat attenuation of the input by about 10–12 dB up to 4000 Hz. There is a slight roll-off in the response above that. The exact nature of the effects of the tape on the frequency response can be measured by the clinician using probe microphone equipment.

40.5.4 Clinical Strategy #4: Change the Musical Instrument

This is typically one of the last suggestions that a clinician may make to a musician but it can be quite useful, especially in the case of the original instrument generating significant output in a frequency region of significant inner hair cell damage, such as a dead region in the cochlea. Changing to a similar but more bass-oriented instrument may be quite useful. A violinist can change to a viola, which is about a fifth lower. And an E flat saxophonist may change to a tenor B flat saxophone. For those readers who would like more information on this strategy, they are referred to an excellent resource put out by the Association of Adult Musicians with Hearing Loss (www.AAMHL.org) called *Making Music with a Hearing Loss* edited by *Cherisse W. Miller* [40.57]

40.6 Some Hearing-Aid Technologies to Handle the More Intense Inputs of Music

If an individual is seeking new hearing aids or have yet to obtain their first set of hearing aids, there are a number of technical innovations that are hearing-aid-hardware based and should be considered. That is, these are not software adjustment that can be made but are inherent in the hearing-aid design. Over the past several years, the hearing aid manufacturing field has transitioned to a post-16 bit architecture that has allowed a larger input dynamic range. In order to handle some of the higher level inputs associated with music, an 18 or 19 bit system should be sufficient. Many of these manufacturers also use one or more of the following digital techniques mentioned below in their algorithm and/or hardware design. Some of these technical innovations are mentioned below and have been on the market for quite a while.

40.6.1 Technical Strategy #1: K-AMP Analog Hearing Aid

Analog hearing aids are no longer widely available today but an advantage of this approach to amplification is that analog hearing aids do not require an A/D converter and may therefore not be limited with regard to the input as discussed earlier. One analog hearing aid, the K-AMP [40.58–60], has shown itself to be a good choice for music and speech for over 25 years. This technology has a specially designed front end that can handle inputs of up to 115 dB SPL, which is the limit of the hearing-aid microphone. Hearing aids with the K-AMP circuitry were the mainstay of working with hearing-impaired musicians in the 1990s and early 2000s.

40.6.2 Technical strategy #2: Changing Where the Dynamic Range of the A/D Converter Operates

There are now a number of technologies in the hearing-aid industry that either automatically adjust to the input signal to ensure that the input level is within the capability of the A/D converter, or increase the dynamic range of the A/D converter to be optimized for the more intense inputs characteristics of music. The first technological approach is analogous to the strategy of bending over slightly under a low bridge. The input level is reduced or compressed after the microphone but before the A/D converter, and then amplified or expanded digitally after the A/D converter. The final digitized input

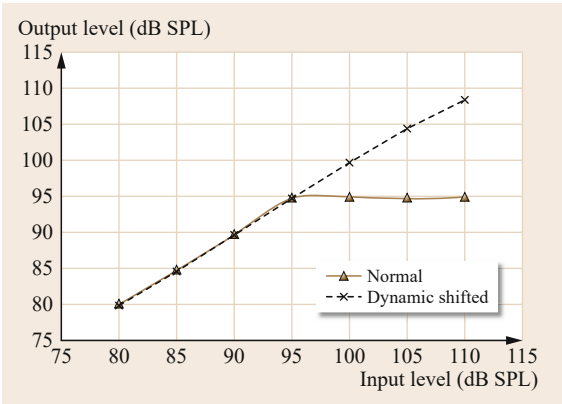


Fig. 40.2 Input/output function with compression settings of 110 dB (dynamic shifted) and 95 dB (normal) before the A/D converter for a 1 kHz sinusoidal signal (reprinted from [40.61], with permission from Elsevier)

is essentially identical to what would have been received had it been generated at a lower level initially. The second technological approach is derived from the definition of *dynamic range*. As the name suggests, it is not 96 dB SPL that is the maximum input to a hearing aid before distortion occurs, but that this is the (maximum) range that is characteristic of the 16-bit architecture in current digital hearing aids. This range can go from 0 dB SPL to 96 dB SPL, or from 15 dB SPL to 111 dB SPL. In both cases the dynamic range is 96 dB but the second one is more appropriate for listening to and playing of music. Both of the above technological solutions are currently in the marketplace and new ones, based on similar approaches, are being developed. For more information please see [40.61–65].

If the level that is permitted to enter the hearing aid is adjusted before the A/D converter, then this can have large effects on the input to the hearing aid. The differences in behavior of an input compression system when the levels are changed are shown with an input/output graph in Fig. 40.2. In the reference situation, seen as a gray line, the threshold of the input compressor becomes active at 95 dB SPL, whereas the black line shows the condition where the level has been shifted to have an effect at 110 dB SPL. This measurement was made with a 1000 Hz pure-tone sinusoidal signal swept in level.

We should next examine what happens with music, using an excerpt from *Eine kleine Nachtmusik* (Serenade No. 13 for strings in G major), K. 525 by Mozart (CD EMI records) played at 110 dB SPL (peak). The first waveform in Fig. 40.3 shows the music with the threshold for compression set to 95 dB SPL. The resulting output, in gray, is compressed in comparison to the original recording, seen in black. In Fig. 40.4 however, the level before compression takes place is shifted to 110 dB and this compressed effect on the waveform is not seen. The dynamic range of the music is therefore preserved and will be converted into the digital domain.

The adjustment to the compressor is being made before the A/D converter and therefore before any amplification is applied [40.63]. The peaks of the music are not increasing the output levels of the hearing aid and therefore the maximum output levels have not changed. The maximum power output (MPO) in the hearing aid is always set based on the real or calculated uncomfortable loudness levels (UCL) to prevent any potential damage to the hearing-impaired listener's residual hearing.

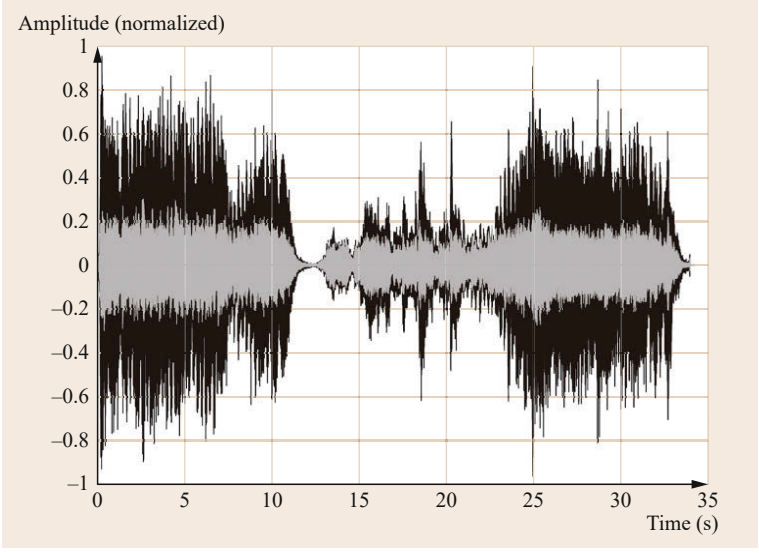


Fig. 40.3 Recording with the threshold of the compressor set to 95 dB before the A/D converter; input level 110 dB SPL. The *black waveform* is the original input file (reprinted from [40.61], with permission from Elsevier)

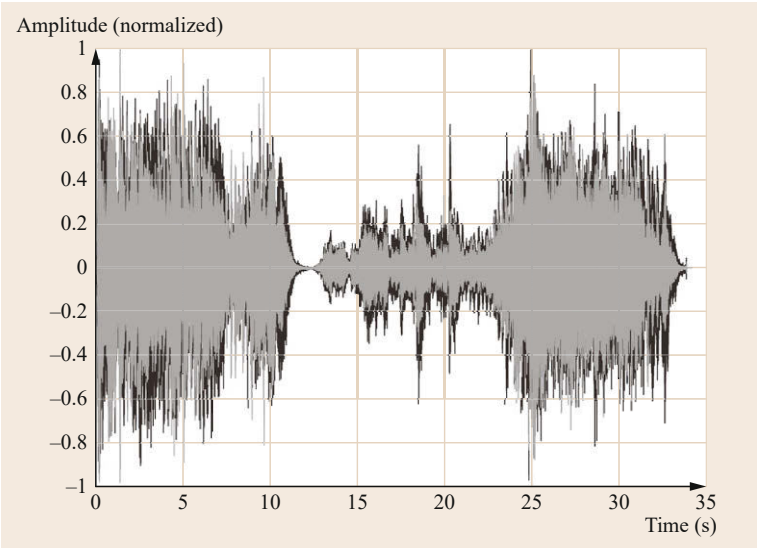


Fig. 40.4 Recording with the threshold of the compressor set to 110 dB before the A/D converter; input level 110 dB SPL. The *black waveform* is the original input file (reprinted from [40.61], with permission from Elsevier)

40.6.3 Technical Strategy #3: Use of a Less Sensitive (or −6 dB/Octave) Microphone

The use of a hearing-aid microphone that is less sensitive to lower frequency sound energy, which can be quite intense, has been shown to be quite useful for those musicians who do not require significant low-frequency amplification. In this approach, a −6 dB/octave low cut (i. e., high pass) hearing-aid microphone system is used in place of the broadband hearing-aid microphone that is typically used. This hearing-aid microphone is less sensitive to 500 Hz energy by 6 dB and to 250 Hz energy by 12 dB (i. e., a 6 dB/octave roll-off). High-level, low-frequency music energy that would otherwise cause the front end to distort will now be reduced, thereby allowing the input to be more within the optimal operating range of the A/D converter and associated circuitry. For those individuals who require a significant amount of low-frequency amplification, this can be compensated for within the fitting software (after the A/D converter) by adding back in the 6 dB of lost input at 500 Hz and 12 dB at 250 Hz. More information on this process can be found in [40.66, 67].

It is important to reiterate that none of these technologies or clinical strategies are simply a matter of adjusting the fitting software provided by the hearing-aid manufacturer. These technologies are built into the hearing-aid hardware and cannot be adjusted within the clinic. Fitting software adjustments are therefore not the primary approach that should be taken whenever music is considered as an input to a hearing aid.

40.7 General Recommendations for an Optimal Hearing Aid for Music

If the clinician has been able to select or modify the hearing aid appropriately to ensure that the more intense components of music are transduced through the front end with minimal distortion, then what types of software manipulations would be useful for a *music program*? We will now review four general clinical recommendations.

40.7.1 Recommendation #1: Similar WDRC Parameters for Speech and for Music

When looking at the compressions settings in a hearing aid, it is quite possible that no changes would need to be implemented for a *music program* that would be different from a *speech in quiet program* [40.43, 68–70]. This could also be attributed to the fact that it is quite difficult to make generalized statements concerning the many differences in the implementation of compression architecture used by the different manufacturers of hearings aids. The use of the wide dynamic range compression (WDRC) circuitry is primarily an attempt to reestablish normal loudness growth due to outer hair cell damage. WDRC can apply more amplification to softer signals than to louder signals so that the dynamic range at the input can be fitted into each individual's personal dynamic range [40.71]. Therefore, audibility and comfort

are maintained. WDRC addresses the damage to the auditory system rather than the nature of the input stimuli per se. *Davies-Venn* et al. [40.68, p. 696] supported this notion and noted that

> *WDRC* [...] *may be just as good for music as for speech* [... and] *that* [this] *hypothesis was supported by the present data.*

40.7.2 Recommendation #2: Gain Settings Within a Dedicated Music Program – the – 6 dB rule

Assuming that both of the *speech in quiet* and *music* programs utilize similar WDRC parameters then the *music program* should be set with about a 6 dB lower OSPL90 (OSPL90) and a 6 dB lower gain than the individual's *speech in quiet* program. Since the OSPL90 is a measure of the maximum amplification that can be applied via a hearing aid within the limitations of using only pure-tone stimuli, then we want to ensure that the peaks of the signal do not exceed the tolerance level of the individual. If the crest factor for music is roughly 6 dB higher than that of speech (18 dB versus 12 dB), then the OSPL90 for music should account for this. As can be seen from Table 40.2 this can vary significantly depending on how one analyzes the crest factor, but generally in order to prevent the peaks of music from causing discomfort, the OSPL90 and the gain should be 6 dB less intense than for the *speech in quiet* program setting.

40.7.3 Recommendation #3: Bandwidth for a Music Program

There is no definitive research that has been performed specifically concerning the bandwidth within the hearing aid that is necessary for music, but here is a summary of what the most relevant studies indicate. Essentially, given the limitations of an individual's hearing loss there is no reason to set the bandwidth differently for a *speech in quiet* program as there is for a *music* program. Any change in bandwidth has more to do with the degree and configuration of the audiometric hearing loss. Examining the work of *Moore* et al. [40.72], *Moore* [40.70] and the work of *Ricketts* et al. [40.73] several general recommendations can be made. A broad bandwidth should be specified (to the limits of the capability of the hearing aid) if there is only a mild to moderate sensorineural hearing loss, and if the configuration is relatively flat. However if the hearing loss is greater and/or the configuration of the audiogram is steeply sloping, that is, the hearing loss is milder in the low frequencies and is much worse in the high frequencies, then a narrower bandwidth may be desirable.

40.7.4 Recommendation #4: Disable the Feedback Cancellation and Noise Reduction Systems

Noise reduction systems and feedback management systems are designed to remove acoustic information that is judged to be undesirable, or even uncomfortable for the hearing-aid user. For this reason a general recommendation can be made to disable the feedback cancellation system and noise reduction system, to the extent that they can be disabled via the hearing-aid fitting software, so they do not affect the musical signal. Noise reduction systems are designed to classify noise on the basis that signals with low modulation are generally less interesting (e.g., steady state noise), than signals with high modulation such as human speech [40.74]. With some genres of music the noise reduction system may actually remove part of a musical signal that could be of interest to the listener based on this classification scheme. A subset of noise reduction systems may also reduce the intrusive nature of transient signals such as cutlery sounds in a restaurant. Again this could affect music depending on the genre and the instrument played. In some cases, feedback cancellation systems can remove periodic signals similar to that produced by some musical instruments such as a flute or a piccolo [40.27]. They may also produce other audible artifacts such as *chirping* in the case of phase cancellation technology for feedback management [40.27, 75]. Although this is becoming less of a problem as these algorithms are improved, it is still possible that a hearing aid might confuse a feedback signal with a harmonic of the music. Some manufacturers have limited the feedback management algorithms to only function in the higher frequency region (e.g., above 2000 Hz) although this, to date, has only been a partial solution. Given the wide variety of implementations of feedback reduction cancellation and noise reduction algorithms that are currently available, it is generally desirable that they should be disabled if at all possible.

40.8 Conclusions and Recommendations for Further Research

Amplification with hearing aids has been shown to reduce the negative effects of hearing loss faced by individuals who have a hearing impairment but who do not require a cochlear implant [40.76–78]. In this chapter we have explored the topic of enhancing the listening experience for music with hearing aids for musicians and enthusiastic concert goers. We first looked at the basic assessment of the auditory system, primarily the periphery where damage is clearly identifiable due to prolonged exposure to noise or loud music. We then explored some key acoustical differences between speech and music and how these differences need to be taken into account when looking at hearing aids for music. Hearing-aid technology was then examined and solutions to make music not only sound better but also be more useful for the musicians were discussed. The key limitation concerning the capability of current digital hearing aids to accommodate the more intense elements of music is the A/D converter. Expending research and development efforts on the other elements within the hearing aid will not really improve the fidelity of music.

As in most areas within the field of audiology the realm of music as an input to hearing aids and the technologies that are available is a rapidly changing one. New technologies are on the horizon, such as better A/D converters that may be implemented by various manufacturers. New assessment techniques such as the MPT discussed earlier could potentially lead to better and more personalized evaluation of the hearing-impaired musician. The authors both hope that music becomes more of an issue within audiology and also that hearing impairment becomes more of an issue within the field of musicology. Researchers and clinicians in two fields could potentially create common solutions to overcome the setbacks that musicians face when confronted with a hearing loss.

References

40.1 S. Kochkin: MarkeTrak VIII: The key influencing factors in hearing aid purchase intent, Hear. Rev. **19**(3), 12–25 (2012)

40.2 A. Laplante-Lévesque, L. Hickson, L. Worrall: Predictors of rehabilitation intervention decisions in adults with acquired hearing impairment, J. Speech Lang. Hear. Res. **54**, 1385–1399 (2011)

40.3 A. Laplante-Lévesque, L. Hickson, L. Worrall: What makes adults with hearing impairment take up hearing aids or communication programs and achieve successful outcomes?, Ear Hear. **33**, 79–93 (2012)

40.4 L.J. Revit: What's so special about music?, Hear. Rev. **16**(2), 12–19 (2009)

40.5 G.D. Chermak: Deciphering auditory processing disorders in children, Otolaryngol. Clin. North Am. **35**(2), 733–749 (2002)

40.6 B. Stach: Diagnoising central auditory processing disorders in adults. In: *Audiology Diagnosis*, ed. by R.J. Roeser, M. Valente, H. Hosford-Dunn (Thieme, New York 2000) pp. 355–379

40.7 J.W. Hall: *New Handbook for Auditory Evoked Responses* (Pearson, Upper Saddle River 2006)

40.8 L.S. Alcord, R.B. Burr, C.A. McCormick: Functional brain imaging in audiology. In: *Audiology Diagnosis*, ed. by R.J. Roeser, M. Valente, H. Hosford-Dunn (Thieme, New York 2000) pp. 121–137

40.9 F.E. Musiek, J.A. Baran, M.L. Pinheiro: *Neuroaudiology Case Studies* (Singular, San Diego 1993)

40.10 R.H. Margolis, L.L. Hunter: Acoustic imittance measurements. In: *Audiology Diagnosis*, ed. by R.J. Roeser, M. Valente, H. Hosford-Dunn (Thieme, New York 2000) pp. 381–423

40.11 Y. Cai, F. Zhao, Y. Zheng: Mechanisms of music perception and its changes in hearing impaired people, Hear. Balance Commun. **11**(4), 168–175 (2013)

40.12 A. Axelsson, F. Lindgren: Hearing in pop musicians, Acta Oto-Laryngol. **85**, 225–231 (1978)

40.13 W.W. Clark: Noise exposure from leisure activities: A review, J. Acoust. Soc. Am. **90**(1), 175–181 (1991)

40.14 J.M. Flugrath: Modern-day rock-and-roll music and damage risk criteria, J. Acoust. Soc. Am. **45**(3), 704–711 (1969)

40.15 C.W. Hart, C.L. Geltman, J. Schupbach, M. Santucci: The musician and occupational sound hazards, Med. Probl. Perform. Artists **2**(3), 22–25 (1987)

40.16 A. Axelsson, F. Lindgren: Hearing in classical musicians, Acta Oto-Laryngol. **91**(Suppl. 377), 1–100 (1981), https://doi.org/10.3109/00016488109108191

40.17 A. Behar, W. Wong, H. Kunov: Risk of hearing loss in orchestra musicians: Review of the literature, Med. Probl. Perform. Artists **21**(4), 164–168 (2006)

40.18 J.E. Camp, S.W. Horstman: Musician sound exposure during performance of Wagner's Ring Cycle, Med. Probl. Perform. Artists **7**(2), 37–39 (1991)

40.19 E.N. MacDonald, A. Behar, W. Wong, H. Kunov: Noise exposure of opera musicians, Canad. Acoust. **36**(4), 11–16 (2008)

40.20 S.L. Phillips, S. Mace: Sound level measurements in music practice rooms, Music Perform. Res. **2**, 36–47 (2008)

40.21 S.F. Poissant, R.L. Freyman, A.J. MacDonald, H.A. Nunes: Characteristics of noise exposure during solitary trumpet playing: Immediate impact on distortion-product otoacoustic emissions and long-term implications for hearing, Ear Hear. **33**(4), 543–553 (2012)

40.22 J.D. Royster, L.H. Royster, M.C. Killion: Sound exposures and hearing thresholds of symphony orchestra musicians, J. Acoust. Soc. Am. **89**(6), 2793–2803 (1991)

40.23 J.H. Schmidt, E.R. Pedersen, P.M. Juhl, J. Christensen-Dalsgaard, T.D. Andersen, T. Poulsen, J. Bælum: Sound exposure of symphony orchestra musicians, Ann. Occupat. Hyg. **55**(8), 893–905 (2011)

40.24 J.J. May: Occupational hearing loss, Am. J. Ind. Med. **37**, 112–120 (2000)

40.25 B.C.J. Moore: *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues*, 2nd edn. (Wiley, Chichester 2007)

40.26 J.A. Lapsley Miller, L. Marshall: Otoacoustic emissions as a preclinical measure of noise-induced hearing loss and susceptibility to noise-inducded hearing loss. In: *Otoacoustic Emissions: Clinical Applications*, ed. by M.S. Robinette, T.J. Glattke (Thieme, New York 2007) pp. 321–341

40.27 H. Dillon: *Hearing Aids*, 2nd edn. (Boomerang, Turramurra 2012)

40.28 C. Halprin: The tuning curve in clinical audiology, Am. J. Audiol. **11**, 56–54 (2002)

40.29 B.C.J. Moore, A.N. Malicka: Cochlear dead regions in adults and children: Diagnosis and clinical implications, Semin. Hear. **34**(1), 37–50 (2013)

40.30 B.C.J. Moore, M. Huss, D.A. Vickers, B.R. Glasberg, J.I. Alcántara: A test for the diagnosis of dead regions in the cochlea, Br. J. Audiol. **34**, 205–224 (2000)

40.31 B.C.J. Moore, B.R. Glasberg, M.A. Stone: New version of the TEN test with calibrations in dB HL, Ear Hear. **25**(5), 478–487 (2004)

40.32 K.E. Bright: Spontaneous Otoacoustic emissions in populations with normal hearing sensitivity. In: *Otoacoustic Emissions: Clinical Applications*, ed. by M.S. Robinette, T.J. Glattke (Thieme, New York 2007) pp. 69–86

40.33 T.J. Glattke, M.S. Robinette: Transient evoked otoacoustic emissions in populations with normal hearing sensitivity. In: *Otoacoustic Emissions: Clinical Applications*, ed. by M.S. Robinette, T.J. Glattke (Thieme, New York 2007) pp. 87–105

40.34 B.L. Lonsbury-Martin, G.L. Martin: Distortion-product otoacoustic emissions in populations with normal hearing sensitivity. In: *Otoacoustic Emissions: Clinical Applications*, ed. by M.S. Robinette, T.J. Glattke (Thieme, New York 2007) pp. 107–130

40.35 V. Looi, P. Winter, I. Anderson, C. Sucher: A music quality rating test battery for cochlear implant users to compare the FSP and HDCIS strategies for music appreciation, Int. J. Audiol. **50**(8), 503–518 (2011)

40.36 V. Looi, K. Gfeller, V. Driscoll: Music appreciation and training for cochlear implant recipients: A review, Semin. Hear. **33**(4), 307–334 (2012)

40.37 H.J. McDermott: Music perception with cochlear implants: A review, Trends Amplif. **8**(2), 49–82 (2004)

40.38 G.L. Nimmons, R.S. Kang, W.R. Drennan, J. Longnion, C. Ruffin, B. Yueh, J.T. Rubinstein: Clinical assessment of music perception in cochlear implant listeners, NIH Public Access **29**(2), 149–155 (2009)

40.39 D.A. Colucci: Hearing matters: Aided mapping for music lovers: Addressing the basic issues. Hear. J. **66**(10) (2013)

40.40 K.L. Rutledge: *A Music Listening Questionnaire for Hearing Aid Users*, Unpublished Master Thesis (The University of Canterbury, Christchurch 2009)

40.41 Y. Cai, F. Zhao, Y. Zheng: Development and validation of a Chinese music quality rating test, Int. J. Audiol. **52**(9), 587–595 (2013)

40.42 M. Uys, C. van Dijk: Development of a music perception test for adult hearing-aid users, South Afr. J. Commun. Disord. **58**, 19–47 (2011)

40.43 M. Chasin, F.A. Russo: Hearing aids and music, Trends Amplif. **8**(2), 35–47 (2004)

40.44 W.A. Olsen: Average speech levels and spectra in various speaking/listening conditions: A summary of the Pearson, Bennett, and Fidell (1977) report, Am. J. Audiol. **7**, 1–5 (1998)

40.45 IEC 61671-1: *Sound Level Meters – Part 1: Specifications* (International Electrotechnical Commission, Geneva 2003)

40.46 M. Chasin: Hearing aids for Musicians, Hear. Rev. **13**(3), 11–16 (2006)

40.47 K. Johnson: *Acoustic and Auditory Phonetics*, 2nd edn. (Blackwell, Oxford 2003)

40.48 M. Chasin: Hear the Music blog, http://www.hearinghealthmatters.org/hearthemusic/ (2011)

40.49 H.K. Dunn, S.D. White: Statistical measurements on conversational speech, J. Acoust. Soc. Am. **11**(3), 278–288 (1940)

40.50 L.J. Sivian, S.D. White: On minimum audible sound fields, J. Acoust. Soc. Am. **4**(4), 288–321 (1933)

40.51 R.M. Cox, J.S. Mateisch, J.N. Moore: Distribution of short-term RMS levels in conversational speech, J. Acoust. Soc. Am. **84**, 1100–1104 (1988)

40.52 R.M. Cox, J.N. Moore: Composite speech spectrum for hearing-aid gain prescriptions, J. Speech Hear. Res. **31**, 102–107 (1988)

40.53 J. Agnew: Amplifiers and circuit algorithms for contemporary hearing aids. In: *Hearing Aids: Standards. Options and Limitations*, ed. by M. Valente (Thieme, New York 2002)

40.54 M. Chasin: Music and the problem with hearing aids. The American Speech Hearing Association, Perspectives on Hearing and Hearing Disorders: Research and Diagnostics, Special Interest Group 6 (2012)

40.55 B.C.J. Moore: *An Introduction to the Psychology of Hearing*, 6th edn. (Emerald, Bingley 2012)

40.56 T.C. Gitles, P.T. Niquette: FIG6 in ten, Hear. Rev. **20**(1), 28–30 (1995)

40.57 C.W. Miller: *Making Music with a Hearing Loss: Strategies and Stories*, http://www.musicianswithhearingloss.org (2011)

40.58 M.C. Killion: An "acoustically invisible" hearing aid, Hear. Instrum. **39**(10), 39–44 (1988)

40.59 M.C. Killion: A high fidelity hearing aid, Hear. Instrum. **41**(8), 38–39 (1990)

40.60 M.C. Killion: The K-AMP hearing aid: An attempt to present high fidelity for the hearing-impaired. In: *Recent Developments in Hearing Instrument Technology: 15th Danavox Symposium*, ed. by J. Beilin, G.R. Jensen (Stougaard Jensen, Copenhagen 1993) pp. 167–229

40.61 M. Chasin, N.S. Hockley: Some characteristics of amplified music through hearing aids, Hear. Res. **308**, 2–12 (2014)

40.62 N.S. Hockley, F. Bahlmann, M. Chasin: Programming hearing instruments to make live music more enjoyable, Hear. J. **63**(9), 30–38 (2010)

40.63 N.S. Hockley, F. Bahlmann, B. Fulton: Analog to digital conversion to accommodate the dynamics of live music in hearing Instruments, Trends Amplif. **16**(3), 146–158 (2012)

40.64 American National Standards Institute: American National Standard: Testing hearing aids with a broad-band noise signal. ANSI S3.42-1992. New York: American National Standards Institute. (1992)

40.65 M. Chasin: A hearing aid solution for music, Hear. Rev. **21**(1), 28–31 (2014)

40.66 M. Chasin, M. Schmidt: The use of a high frequency emphasis microphone for musicians, Hear. Rev. **16**(2), 32–37 (2009)

40.67 M. Schmidt: Musicians and hearing-aid design – Is your hearing instrument being overworked?, Trends Amplif. **16**(3), 140–145 (2012)

40.68 E. Davies-Venn, P. Souza, D. Fabry: Speech and music quality ratings for linear and nonlinear hearing-aid circuitry, J. Am. Acad. Audiol. **18**(8), 688–699 (2007)

40.69 B.C.J. Moore: The choice of compression speed in hearing aids: Theoretical and practical considerations and the role of individual differences, Trends Amplif. **12**(2), 103–112 (2008)

40.70 B.C.J. Moore: Effects of bandwidth, compression speed, and gain at high frequencies on preferences for amplified music, Trends Amplif. **16**(3), 159–172 (2012)

40.71 B.C.J. Moore: Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids, Ear Hear. **17**(2), 133–161 (1996)

40.72 B.C.J. Moore, C. Füllgrabe, M.A. Stone: Determination of preferred parameters for multichannel compression using individually fitted simulated hearing aids and paired comparisons, Ear Hear. **32**(5), 556–568 (2011)

40.73 T.A. Ricketts, A.B. Dittberner, E.E. Johnson: High frequency amplification and sound quality in listeners with normal through moderate hearing loss, J. Speech-Lang.-Hear. Res. **51**, 160–172 (2008)

40.74 A. Schaub: *Digital Hearing Aids* (Thieme, New York 2008)

40.75 J.M. Kates: *Digital Hearing Aids* (Plural Publishing, San Diego 2008)

40.76 National Council on the Aging: *The Consequences of Untreated Hearing Loss in Older Persons* (NCOA, Washington D.C. 1999)

40.77 T.H. Chisolm, C.E. Johnson, J.L. Danhauer, L.J.P. Portz, H.B. Abrams, S. Lesner, P.A. McCarthy, C.W. Newman: A systematic review of health-related quality of life and hearing aids: Final report of the American Academy of Audiology task force on the health-related quality of life benefits of amplification in adults, J. Am. Acad. Audiol. **18**(2), 151–183 (2007)

40.78 S. Kochkin: MarkeTrak VIII: Patients report improved quality of life with hearing-aid usage, Hear. J. **64**(6), 25–32 (2011)

41. Music Technology and Education

Estefanía Cano, Christian Dittmar, Jakob Abeßer, Christian Kehling, Sascha Grollmisch

In this chapter, the application of music information retrieval (MIR) technologies in the development of music education tools is addressed. First, the relationship between technology and music education is described from a historical point of view, starting with the earliest attempts to use audio technology for education and ending with the latest developments and current research conducted in the field. Second, three MIR technologies used within a music education context are presented:

1. The use of pitch-informed solo and accompaniment separation as a tool for the creation of practice content
2. Drum transcription for real-time music practice
3. Guitar transcription with plucking style and expression style detection.

In each case, proposed methods are clearly described and evaluated. Objective perceptual quality metrics were used to evaluate the proposed method for solo/accompaniment separation. Mean overall perceptual scores (OPS) of 24.68 and 34.68 were obtained for the solo and accompaniment tracks respectively. These scores are on par with the state-of-the-art methods with respect to perceptual quality of separated music signals. A dataset of 17 real-world multitrack recordings was used for evaluation. In the drum sound detection task, an F-measure of 0.96 was obtained for snare drum, kick drum, and hi-hat detection. For this evaluation, a dataset of 30 manually annotated real-world drum loops with an onset tolerance of 50 ms was used. For the guitar plucking style and guitar expression style detection tasks, F-measures of 0.93 and 0.83 were obtained respectively. For this evaluation, a dataset containing 261 recordings of both isolated notes as well as monophonic and polyphonic melodies with note-wise annotations was used. To conclude the chapter, the remaining challenges that need to be addressed to more effectively use MIR technologies in the development of music education applications are described.

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_41

41.1 Background

The rapid development of music technology in the past few decades has inevitably changed the way people interact with music today. The means of music consumption have changed, most notably due to the constant availability of digital music via the Internet and mobile applications. Additionally, the technology for music production has moved almost entirely to the digital domain. This evolution made music production technology affordable for individuals and put the ability to create and distribute music in the hands of the average music lover. It was only to be expected that these technology changes would also influence other fields of research, and the application of music technologies in music education began to draw further attention. However, the development of new applications for music education faces many challenges:

1. Development of robust and efficient algorithms that can help in practicing music
2. Bridging the methodical gap between music education and music technology
3. Design of appealing and entertaining applications that motivate the user while developing real musical skills.

When talking about music technology in music education, two important points that need to be addressed are:

1. The development of mechanisms to give users automatic feedback about their performances
2. The possibility to automatically generate relevant practice content.

Several technologies, mainly developed within the music information retrieval (MIR) community, are relevant here. Automatic pitch detection, for example, deals with the extraction of fundamental frequency (f_0) sequences from a given music recording. In a music education context, pitch detection algorithms take an audio signal of the user's performance, and determine which notes the user plays over time. In this way, systems can compare the user rendition to a reference (ground-truth). In terms of automatic content creation, music transcription technologies attempt to automatically extract time and pitch information necessary to recreate the underlying musical score from recorded music. Also relevant for content creation are sound separation technologies that attempt to extract the sound of a particular musical instrument from an audio mixture. In a music education context, music transcription would allow the automatic creation of music scores for pieces whose score is not available, such as improvisations, live performances, etc. Sound separation would allow, for example, the creation of backing tracks that the user can use during practice time or concert preparation.

In Fig. 41.1 an example of music technology applied in a music education context is displayed. Figure 41.1a shows a spectrogram with the results of a pitch detection algorithm. Figure 41.1b shows how this algorithm can be used inside a music education application to to give the user real-time feedback on their performance.

This chapter presents a general overview of the use of music technology in music education. Three specific cases of the use of music information retrieval (MIR) technologies in a music education context are presented:

- Sound source separation
- Drum transcription
- Guitar transcription.

The chapter is organized as follows: Sect. 41.2 gives an overview of existing applications and projects and Sects. 41.3–41.5 present the three aforementioned cases of MIR research and their application to music education. Finally, Sect. 41.6 presents a discussion and outlines future challenges in the field.

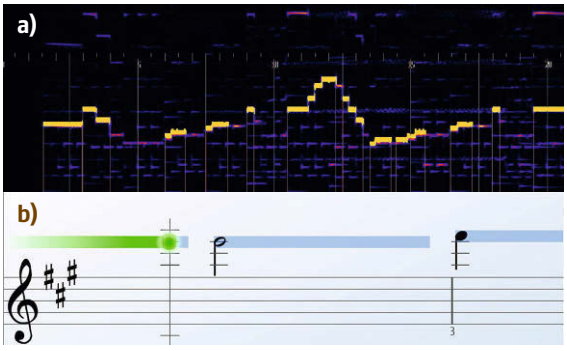


Fig. 41.1a,b From music technology to music education. **(a)** Spectrogram showing the results (in *yellow*) of a pitch detection algorithm. **(b)** Example of a music game with real-time pitch detection functionalities

41.2 Music Education Tools

This section provides a general overview of commercially available music education applications. These applications can be broadly classified into four categories:

- Published music education material
- Music video games
- Software for music education
- Music-learning mobile applications (apps).

For a thorough list of corresponding online references, please refer to [41.1].

41.2.1 Published Music Education Material

Play-along CDs became very popular in the 1980s. They allow the user to practice popular musical pieces in sync with specially recorded (e.g., piano or orchestral) accompaniments. In most cases, the accompaniment and the solo instrument are panned to the left and right stereo channels, so the user can directly choose the mixing ratio. In this way, users can familiarize themselves with the musical piece while developing musical skills. The main drawback of such practicing tools is the limited amount of available content: each track needs to be produced and recorded, leading to high production costs. Therefore, play-along CDs are mainly limited to some representative concerts of the instrumental repertoire and very famous songs. Popular jazz play-alongs have been published by *Jamey Aebersold* [41.2] in the jazz series with over 100 items, and by the Hal Leonard Corporation [41.3] featuring different artists, instruments, playing techniques, jazz standards, and thematic editions. A larger catalog of play-alongs for different genres is offered by Music Minus One [41.4] including different instruments and ensembles.

Instructional videos started as an educational tool, where particular topics like instrument playing techniques are shown and explained by renowned professional musicians, covering important points like proper warm-up exercises, improvisation, and ways to play certain sections of a musical piece. Later on, the popularity of some musicians was used as a marketing tool, moving the focus of the videos from the musical content itself to the featured artist giving instructions. This appealing idea led to a vast and growing catalog of instructional videos. Initially released on VHS tapes in the 1980s and 1990s, modern videos are published on digital media formats like DVD or online resources like YouTube. Educational videos have been published amongst others by Alfred Music Publishing [41.5], Berklee Press [41.6], Icons of Rock [41.7], and Homespun [41.8].

Even though published materials are very popular for practicing at home, they all share the same disadvantage: lack of direct feedback of the users' performance and progress. Therefore, users have to be completely self-motivated and rely on their own perception. This can be very challenging especially for beginners.

Moving away from traditional content distribution, some websites such as Drumeo [41.9] or Get2Play [41.10] focus on offering online instructional videos as their main service. Compared to physically distributed instructional videos, Drumeo tries to avoid the missing feedback by offering community features, custom lesson plans, and feedback by instructors for which students have to submit a video of their performance.

41.2.2 Music Video Games

Music video games were first recognized as a genre in 1996 with the release of the music rhythm game PaRappa the Rapper for Sony PlayStation 1. Rhythm games can be seen as a subgenre of music video games, which had their commercial breakthrough with the release of Guitar Hero in 2005 for Sony PlayStation 2 [41.11]. Its specialized controller simulates a real guitar in shape, replacing strings and frets with five buttons and a strum bar. Later installments of the series extended the game by supporting simplified drums, bass guitar, and vocals. A comparable popular video game is Rock Band, which covered guitar, drums, bass and vocals already in its first release. The third installment added a three-part harmony recognition feature for vocals and a keyboard-shaped controller. Both game series offer new songs as downloadable content and are available on various gaming platforms.

While being entertaining and successful in fostering interest in music, rhythm games fail to develop musical skills transferable to real musical instruments due to their overly simplified instrument controllers.

Karaoke music video games focus on vocal performance, displaying lyrics and the reference melody (often displayed in piano-roll notation) while rating the users' vocal performance. One popular example in this subgenre is SingStar [41.12], released for Sony PlayStation 2 and 3.

The first commercially viable music video game designed to be controlled with a real musical instrument was Rocksmith, released in the United States in September 2011 for Microsoft Xbox 360, Windows, and Sony PlayStation 3 by Ubisoft [41.13]. It supports electric guitar and electric bass guitar. The audio signal is captured with an USB audio interface connected

to the output plug of the instrument. The performance of the user is rated in real-time by means of pitch detection and compared to the reference notation of the backing track. The difficulty adapts to the skill level of the player, adding and removing notes depending on performance. Additional mini games with scales or chord exercises are available to further improve the playing skills of the user. This feature moved the game away from mere entertainment towards music education. The latest installment, Rocksmith 2014, released in November 2013, moves even more in an educational direction by adding a session mode where players can jam with a virtual band. The virtual band members, musical genre, tempo and key of the jam session can be selected and the band adjusts to the user's performance.

While most of the mentioned games have a large database of available content already delivered or available for download, the possibility of including original content is not supported and mostly it is popular rock/pop songs that are available. Therefore, pleasing all users' personal tastes is not possible.

41.2.3 Music Education Software

While music video games focus mainly on entertaining the user, music education software aims at an interactive educational approach.

An example web-based system developed by Grieg Music Education is Music Delta [41.14]. It is available in two different versions:

1. Music Delta Planet, developed for teaching music history, artists, and composers to elementary-school children in an entertaining way
2. Music Delta Master, consisting of modules for creating a deeper understanding of music theory, composition, history, and performance.

When it comes to musical skills, sight reading and dexterity, Synthesia [41.15] and Gigajam [41.16] are software applications that allow users to play along to exercise pieces with Musical Instrument Digital Interface (MIDI) instruments, such as keyboards and drum sets.

SmartMusic is aimed at orchestras, woodwind, brass, percussion, and string musicians as well as vocalists of all levels [41.17]. While students can access the SmartMusic database that contains numerous musical pieces, educators can assign tasks to their students for a more guided learning approach. SmartMusic includes several practice tools, e.g., for displaying fingerings, reducing tempo and recording for self-assessment.

Apple's interactive learning software GarageBand teaches piano and guitar with specially designed content [41.18]. The system uses the computer microphone or USB audio devices to provide direct feedback. Even though it is released for both Mac and iPad, the mobile version lacks the practicing content included in the desktop version.

Songs2See is an application that supports both real-time feedback during practice, as well as importing of custom exercises [41.19]. It was originally developed in an MIR research project [41.20]. In contrast to the above-mentioned music video games, it has a more thorough approach to instrumental learning and lets users choose between common music notation, piano-roll, and tablature.

41.2.4 Music-Learning Mobile Apps

The market for tablets and smartphone applications is growing steadily, with a considerable number of music learning apps published so far. With the rapid and vast development of apps nowadays, keeping an updated list of music-related apps can be very challenging. Here, a series of representative apps have been described that have presented important developments in the music app context. However, this list is by no means extensive and a great number of relevant and valuable apps, which have not been described in this chapter, are also available for the user.

An application targeting guitar players is Rock Prodigy, which started as a pure download product for iPad, iPhone and iPod touch [41.21]. It was later released for Windows and Mac as well. The user is recorded while playing and immediate performance feedback is given. It offers courses about chords, rhythm, scales, technique, and music theory. The courses can be purchased as downloadable content depending on the user's needs.

A more classical approach is taken by the interactive score sheet app Tonara available for iPad [41.22]. The main feature is automatic score page turning by means of score-following, which adjusts to the user's progress by analyzing the microphone input. Content must be purchased in the app store and offers mostly popular and classical music tracks.

The company Yousician – previously called Ovelin – initially brought to the market two games particularly aimed at beginners: Wild Chords and Guitar Bots. While Wild Chords tried to familiarize the user with guitar chords using and animal-themed interface, Guitar Bots came as a followup product also targeting more advanced performers. The company has now completely focused on their latest app Yousician, marketed as a personalized music teacher and first released in 2014. Yousician currently supports guitar, ukulele, bass, and piano.

41.2.5 Research Projects

A few research projects have dealt with the development of electronic learning systems for music education in the past years. Projects funded by the European Commission include the Interactive Multimedia Environment for Technology Enhanced Music Education and Creative Collaborative Composition and Performance (i-Maestro), the Virtual European Music School (VEMUS), the Interactive Music Tuition System (IMUTUS), the aforementioned Songs2See [41.19, 23–25], and Emused.

Focusing on the violin family, i-Maestro offered enhanced and collaborative practice tools by analyzing gesture and posture based on audiovisual systems and sensors attached to the performer's body. IMUTUS' main goal was to develop a practice environment for recorder where students would get immediate feedback about their performance. The follow-up project VEMUS extended the approach with the inclusion of more musical instruments and tools for self-practicing, music teaching, and remote learning.

The Music Representation Research Group at IRCAM developed Antescofo [41.26], a score-following system and a language for musical composition that automatically recognizes the player's position and tempo in a musical score. Therefore, it can be used as a tool for tempo and performance analysis by researchers and as a practicing tool with interactive accompaniment by music students. Furthermore, it assists the user during the composition process by synchronizing computer-generated sounds with instrumental performances.

Moving the focus from music learning to automatically generated computer accompaniments, Music Plus One [41.27] intended to increase the aesthetic and perceptual quality of the generated music. This is achieved by listening to the users and following their timing and expression. This system, developed by the School of Informatics and Computing in Indiana University, consists of three main steps:

1. *Listen*: It identifies the note onsets and matches them to the reference via a hidden Markov model (HMM).
2. *Play*: The audio output is generated by phase vocoding a pre-existing playback.
3. *Predict*: It tries to predict the future timing by using a Kalman filter-like model.

KOPRA-M (Entwicklung und empirische Validierung eines Modells musikpraktischer Kompetenzen) [41.28] focused on the measurement of music competencies of German secondary-school students. The students have to solve a set of tasks that can be grouped into singing, melody and rhythm tasks. The melody and rhythm tasks have to be performed on a tablet with the purpose-built color music grid app, for the purpose of avoiding differences in rating due to prior skills on certain instruments like piano. The audio input of all tasks is recorded on the system's server for automated ratings. The system consists of a main server that communicates with the client software on the students' machines. The system attempts to model ratings from human experts by employing MIR methods.

41.3 Sound Source Separation for the Creation of Music Practice Material

As explained in Sect. 41.2.1, play-along versions are very popular for music practicing. Due to their expensive production costs, only very popular musical pieces are available and finding play-along versions for individual tastes can be very hard. An alternative way to obtain play-along versions is sound source separation of original music recordings. Sound source separation is an umbrella term for diverse signal processing methods used to extract *source signals* from an audio *mixture*. In the music context, sound source separation refers to the extraction of a given musical instrument from a polyphonic, multitimbral music recording. This section focuses on a particular case of sound source separation called *solo* and *accompaniment* separation. For this specific task, the goal is to separate the audio mix into two sources only: the main (or solo) instrument and the accompaniment. Ideally, this process yields a play-along version of any music recording [41.29].

41.3.1 State of the Art

As of today, the use of prior information about the sources has proven to be advantageous for the separation quality, which is especially true for music signals. The inclusion of known information about the sources in the separation scheme is referred to as *informed source separation* (ISS). ISS comprises, among others, the use of MIDI-like musical scores, the use of pitch tracks of one or several sources, and the extraction of model parameters from training data of a particular sound source [41.30].

In a very general sense, a sound source separation process can be divided into three main processing stages. This process is depicted in Fig. 41.2:

1. *Source parameter estimation*: Before any separation can be performed, all methods need to estimate the parameters corresponding to the desired source. Depending on the method used, different parameters might be required, such as magnitude envelopes, frequency locations of harmonic components, activation coefficients, etc.
2. *Prior information*: The estimation stage often makes use of prior information about the sources to guide and make the estimation more robust.
3. *Separation procedure*: After having estimated the source parameters, this stage refers to the actual separation of the spectral content from the different sources.

In Fig. 41.2, a general block diagram of a sound separation process is illustrated. For each of the processing stages of a common sound source separation process, different methods have been applied in the literature.

Prior information of different types has been used for sound separation. Some approaches make use of pitch sequences of the sound sources to perform separation. In some cases, automatic methods for pitch detection such as the ones presented in [41.31, 32] have been used to extract the f_0 sequence of the main instrument [41.33]. Other approaches have proposed pitch-informed separation methods that extract pitch information directly during the parameter estimation stage (Fig. 41.2) of the separation process [41.34, 35]. Other methods have used available symbolic scores (e.g., MIDI-files) of the music piece as prior information [41.36, 37]. Some systems make use of instrument-specific prior information to perform separation [41.38–40]. These approaches naturally lean towards instrument-specific separation and attempt to model a given musical instrument as accurately as possible.

In the parameter estimation stage, a great diversity of models and techniques have been used to characterize the target source before the final separation procedure. Two signal models, for example, have been frequently used in the separation context. Some systems have used a *sinusoidal model* to characterize sound sources as a sum of sinusoids with varying amplitudes and frequencies [41.41, 42]. Other systems have used the *source/filter model* for separation [41.43]. Some systems have taken advantage of signal sparsity to perform separation [41.44], or have used the repetitive structure of some music signals to differentiate them from other sources [41.45].

A very commonly used technique applied in sound source separation is *nonnegative matrix factorization* (NMF). NMF takes advantage of the nonnegativity of the magnitude spectrogram, and attempts to represent the spectrogram audio mixture as a time-varying weighted sum of spectral bases. Some systems that have used NMF in the separation scheme are [41.46–48]. *Nonnegative tensor factorization* (NTF), which can be understood as a generalization of NMF, has been applied in separation of stereo and multichannel signals [41.49, 50].

After the parameters of the sources have been estimated, the goal of the last stage of a separation process is to obtain a spectral representation of each of the sources to separate. This is the separation procedure block in Fig. 41.2. In this stage the most commonly used procedure is generalized Wiener filtering, which can be applied both in a single-channel scenario [41.51] and as a multichannel approach [41.52].

41.3.2 Proposed Method

The proposed method extracts pitch sequences of the solo instrument and uses them to guide parameter estimation. The main characteristic of the proposed method is that it gathers information of tone objects and uses known characteristics of musical instrument tones to improve estimation. The proposed solo/accompaniment separation method is composed of five processing stages:

1. Pitch extraction
2. Tone formation
3. Harmonic series refinement
4. Spectral masking
5. Postprocessing.

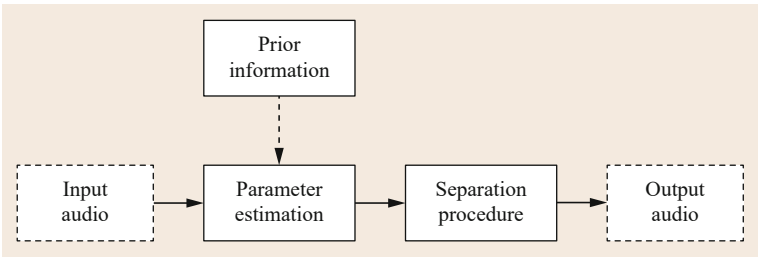


Fig. 41.2 Block diagram of a sound source separation process where prior information about the sources is available

The method operates on spectrograms of the music recordings that are computed via *short-time Fourier transform* (STFT) and its inverse.

Pitch Detection

The pitch detection algorithm described in [41.31] is used to extract frame-wise fundamental frequency sequences $f_0(n)$ of the solo instrument. The algorithm is based on a pair-wise evaluation of spectral peaks that detects partials with successive harmonic numbers. Voices are created from detected pitch candidates and are characterized by a magnitude threshold and frequency range. The most salient voice is selected as the main melody. During pitch extraction, an analysis frame of 46 ms is used in conjunction with a hop size of 5.8 ms. Unpitched frames are marked with $f_0(n) = 0$ Hz.

Tone Formation

The raw f_0 estimates from the pitch detection stage are analyzed over time to create tone objects. A new tone is started when an f_0 value in the range between 65–2000 Hz is found. After the start of a tone has been detected, a moving average filter of length $L = 3$ frames is used to calculate the mean frequency value $\hat{f}_0(n)$ in the time interval defined by the filter length L. The end of a tone is defined either by a new $f_0 = 0$ Hz (no tone was detected) or by a mean frequency variation larger than a semitone (a new tone has started). A minimum tone length of 100 ms is defined. Each tone object is defined by its initial frame n_i, final frame n_f, and its frame-wise *instantaneous frequency* (IF).

Harmonic Series Refinement

This stage estimates the harmonic series of the solo instrument for each tone. Let $k_p(n)$ be the frequency bin of the ideal location of partial p calculated as an integer multiple of the fundamental frequency. The maximum frequency deviation of each partial p from its ideal harmonic location is defined as δ_max. For each time frame n in the range defined by $[n_\mathrm{i}, n_\mathrm{f}]$, a frequency band given by $[k_p(n) - \delta_\mathrm{max}, k_p(n) + \delta_\mathrm{max}]$ is defined where the observed partial location $\hat{k}_p(n)$ is searched. For each partial index $p = 2, \ldots, p_\mathrm{max}$, the search returns the frequency bin $\hat{k}_p(n)$ where the observed harmonic with the largest amplitude is observed. Each partial is allowed to have individual deviation from the ideal harmonic location.

Spectral Masking

After the complete harmonic series has been estimated, initial binary spectral masks for the solo $M_\mathrm{S}(k, n)$ and accompaniment $M_\mathrm{A}(k, n)$ are created. Each time-frequency bin is defined either as part of the solo instrument or part of the accompaniment. To compensate for spectral leakage in the time–frequency transform, a tolerance band $\Delta = 1$ centered at the observed partial location $\hat{k}_p(n)$ is included in the masking procedure.

Postprocessing

This stage refines the initial estimation of the spectral representation of the solo instrument. Three main processing stages are proposed:

1. *Attack refinement*: As clear spectral peaks are mostly observed in the sustained part of the tone, the pitch detection algorithm requires a few processing frames before a valid f_0 value can be detected. To compensate for this inherent delay, the observed harmonic structure $M_\mathrm{S}(k, n_\mathrm{i})$ in frame n_i is replicated in all the frames 70 ms ahead of n_i.
2. *Transient interference reduction*: Transient interferences are assumed to be vertical events in the spectrogram that are common to several (if not all) harmonics and occur in a short time interval [41.53]. In order to detect transients in the solo signal, the temporal envelopes of the partials are smoothed with a median filter of length $L = 3$ and normalized to the [0, 1] range. The lowest partial index considered for the detection of transients $p_\mathrm{low} = 9$. A magnitude threshold $\gamma_L = 0.6$ is defined and frames where the normalized smoothed magnitude envelopes are larger than γ_L are detected as possible transients. The minimum number of partials where a magnitude value larger than γ_L has to be observed for the event to be considered a transient is defined as $\mathrm{min}_p = 6$. To remove a detected transient, the value of $M_\mathrm{S}(k, n)$ in the time frame where the transient was detected is replaced by the mean magnitude value of the normalized smoothed magnitude envelope in the L time frames before the transient was detected. The accompaniment mask is recalculated as $M_\mathrm{A}(k, n) = 1 - M_\mathrm{S}(k, n)$ and the new postprocessed masks are no longer binary.
3. *Common amplitude modulation* (CAM): Partials of a musical instrument tone tend to exhibit highly correlated temporal envelopes. This phenomenon is known as common amplitude modulation [41.54]. Data-driven CAM is enforced in the estimation by weighting the temporal envelopes of the partials with a reference curve. To obtain the reference envelope only the first $p_\mathrm{CAM} = 5$ partials are considered as estimation is always more accurate for lower partials. The reference curve is the normalized temporal envelope of the partial with the highest mean cross-correlation $\bar{r}_p$ with the other $p_\mathrm{CAM} - 1$ en-

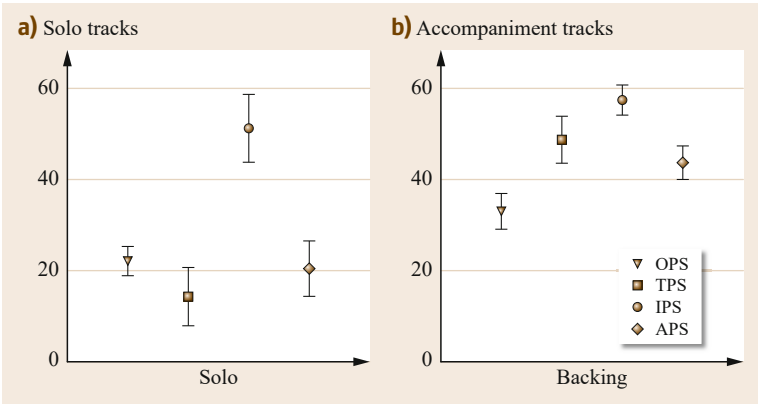


Fig. 41.3a,b Objective perceptual quality measures obtained with the proposed method: overall perceptual score (OPS), target-related perceptual score (TPS), interference-related perceptual score (IPS), artifact-related perceptual score (APS). **(a)** Solo tracks, **(b)** accompaniment tracks

velopes. That is,

$$p_{\text{ref}} = \underset{p}{\text{argmax}}(\bar{r}_p) \,.$$

All the partial envelopes are weighted with the temporal envelope of partial p_{ref}.

41.3.3 Evaluation and Results

The proposed method was evaluated with a dataset composed of 17 multitrack recordings: 10 tracks with vocal solos, and seven tracks with different instrumental solos freely available in [41.29]. The Perceptual Evaluation methods for Audio Source Separation (PEASS) toolbox [41.55] was used to evaluate quality of resulting signals. Resulting quality measures are presented in Fig. 41.3. Mean values with 95% confidence intervals are presented.

The algorithm shows particularly high IPS scores both for the solo and the accompaniment. These results suggest good isolation of the solo instrument. High IPS scores often come hand in hand with lower APS and TPS scores. This can be more clearly observed in the results for the solo tracks. Perceptual quality scores are in general higher for the accompaniment track than for the solo. Larger confidence intervals can be observed for the solo, suggesting that the algorithm can handle some instruments better than others.

The proposed method was submitted to the *Signal Separation Evaluation Campaign* (SiSEC 2013) in the *professionally produced music recordings* task. The algorithm obtained comparable quality ratings to other state-of-the-art approaches with an average processing time of 0.25 s per 1 s of audio on a 2.6 GHz computer. The full table of results can be found on the campaign's website [41.56].

The use of the proposed method in a music education context was also studied in [41.57], by means of a listening test setup specially designed to evaluate the level of comfort of music students when playing to different types of separated signals as audio reference. The main goal was to determine which type of signal distortions (artifact distortions, interference distortions, or target distortions) were most disturbing for musicians during practice. Results from this study can be used as guidelines for optimizing algorithms to better suit the music education context. They indicate that preferences are different for solo and backing tracks. While target distortions were found to be most annoying in backing tracks, artifact distortions proved to be most important for separated solo signals.

41.4 Drum Transcription for Real-Time Music Practice

As shown in Sect. 41.2.3 a small number of music video games and music education software also offer the possibility to play along and practice drums. In all cases, this functionality is enabled by using MIDI-fied drum sets. However, none of the existing applications enable the user to practice using a real-world acoustic drum set. Our goal is to have the drum students play along to a given rhythm pattern or song, while their performances, in terms of hitting the correct drums to the correct metric positions, are assessed in real-time. As a prerequisite, it is necessary to recognize the different drum sounds in a monaural drum recording. In MIR, this task is called automatic drum transcription or drum detection (agnostic to the metric position). Having beginners in mind, we constrained ourselves to detection of onset times for the most basic drums, namely kick drum, snare drum, and hi-hat.

41.4.1 State of the Art

In this section, the most important directions of research for automatic drum transcription are presented. As described in [41.58], the existing approaches can be divided into three different categories depending on the signal model and processing steps taken in the analysis.

Source Separation Methods

The first category is also known as *separate and detect* because the signal is first separated into individual streams and subsequently onset events are detected in each individual stream. The prerequisite is typically a time–frequency transform (e.g., via STFT). The generic signal model decomposes the resulting magnitude spectrogram X into a linear superposition of individual component spectrograms. The components are usually represented by fixed spectral bases B and corresponding time-varying amplitude gains G. The known approaches mostly differ in the decomposition method as well as the constraints and initialization imposed on B and G.

Independent component analysis (ICA) computes a factorization $X = BG$, such that the separated source spectra are maximally independent and non-Gaussian. *Independent subspace analysis* (ISA), first described in [41.59], applies *principal component analysis* (PCA) and ICA in succession for decomposing X. This combination makes it possible to extract multiple sources from fewer observations than normally possible with ICA [41.59]. In order to classify the arbitrarily permuted and scaled components afterwards, feature extraction and classifiers such as *k-nearest-neighbor* (k-NN) or *support vector machines* (SVM) can be used [41.60]. An extension to ICA called *nonnegative independent component analysis* (NICA) has the constraint that the matrix B must be nonnegative [41.61]. In [41.62], how to use NICA for transcription of kick, snare and hi-hat from polyphonic music is shown.

Prior subspace analysis (PSA) was first proposed in [41.63] and utilizes prior knowledge about the sources that should be separated. A starting point for the algorithm is the collection of template spectrum basis functions in a matrix B_p. These consist of the averaged spectra drawn from a large collection of isolated kick and snare sounds. Once the template basis functions are computed, one can get a first approximation of the amplitude gains by

$$\hat{G} = B_\mathrm{p}^{+} X \,,$$

where B_p^{+} denotes the pseudo-inverse. The rows of matrix $\hat{G}$ contain the temporal activations of the template sources in the signal, but are not independent. To make them independent, ICA is applied. This results in an unmixing matrix W transforming $\hat{G}$ into independent amplitude gain functions $G = W\hat{G}$. Subsequently, an improved estimate of the source spectra can be computed by $B = XG^{+}$, which now contains the source spectra adapted to the actual signal. Using this method, [41.64] reports an F-measure of 75% on the detection of kick and snare drums.

An early work applying NMF (Sect. 41.3.1) for the separation of drums from polyphonic music is presented in [41.65]. It uses NMF minimizing the Kullback–Leibler divergence, with random initialization of B and G. From the resulting components, spectral and temporal features are computed and classified by an SVM trained on the classes drums versus harmonic. The reported results show that the NMF and SVM approach performed better than ISA and SVM. Another variant of NMF for drum transcription is described in [41.66]. The NMF is first applied to individual drum spectrograms for kick, snare and hi-hat to get template basis spectra, which are later fixed during the NMF iterations. The method shows good performance on drum loops, yielding an average F-measure of 96% for kick, snare and hi-hat detection. However, one might expect that the quality is largely determined by providing appropriate basis spectra for the initialization. Recently, NMF-based methods have also been applied to real-time drum detection [41.67], where each drum onset is identified with probabilistic spectral clustering based on the Itakura–Saito divergence.

Template Matching

The second category of drum transcription techniques follow a so-called *match and adapt* approach. It relies on temporal or spectral templates for the events that should be detected. In a first approximation the occurrences of events that are similar to the template are detected. Afterwards, the templates are iteratively adapted to the given signal. In turn, this can then be used for a more detailed search. As an example of such an algorithm, the work presented in [41.68] uses template spectrograms for kick and snare, called *seed templates*, which are constructed from a collection of isolated drum sounds. First, onset detection determines possible candidates for drum sounds. At each onset candidate, a spectrogram snippet with the same size as the template is kept and compared with the template. At this stage, the spectrum is filtered according to the typical frequency range of kick, snare or hi-hat and smoothed, so that a rough estimation with the template takes place. The reciprocal of the distance between the observed spectrogram and the template describes the reliability that an onset candidate contains a drum sound. In the adapt stage, the seed templates

are updated by taking the median power over all selected frames. This suppresses more volatile spectral peaks from harmonic instruments. The process of template adaption is used iteratively, so that the output of this median filtering is used as the next seed template. The final stage determines whether the drum sound actually occurs at the onset candidate. As a result of applying the template matching in conjunction with harmonic structure suppression, an F-measure of 82% for kick and 58.3% for snare was reported. A combination of template matching and sound separation is described in [41.69], where the candidates for template extraction are first detected using NMF decomposition. Another example of template matching is given in [41.70], where characteristic band pass filter parameters are learned. The training process is realized through the evolution of the characteristic filters with the *differential evolution* (DE) algorithm and fitness evaluation measures for determining each filter's ability to correctly detect the onset of the respective drum. The output of each filter represents the activations of the single drums and can be processed by means of peak picking.

Machine Learning

The last category of drum transcription algorithms is referred to as *segment and classify*. It first uses temporal segmentation of the audio track into musically relevant parts. Usually, a certain number of frames after each detected onsets is kept or a temporal grid of fixed periodicity is aligned to the audio track. Subsequently, each temporal event is identified by a classifier. Often, well-known machine learning methods such as SVM are used in conjunction with features extracted from each segment. The method in [41.71] uses a set of features comprising averaged Mel-frequency cepstral coefficients (MFCCs), various spectral shape parameters and the log-energy in six frequency bands chosen to mimic the spectral content of different drum instruments. The features are classified by a set of eight binary SVMs that have been trained on the classes kick, snare, hi-hat, clap, cymbal, rim shot, tom and percussion. Evaluated on a dataset of drum loops, the best configuration gave a recognition rate of 83.9%. The method proposed in [41.72] uses a similar approach, but applies it to drum transcription from polyphonic music. The algorithm achieved an average F-measure of 61.1% for the classes kick, snare and hi-hat. Finally, *hidden Markov models* (HMM) are a machine learning method that can be used to model drum sequences. Although they are often counted as part of the *segment and classify* approach, they stand out as they are able to perform the segmentation and detection jointly. HMMs model temporal sequences by computing the probability that a given sequence of observed states were generated by hidden random variables, i. e., the activations of the drum classes. In [41.58], HMMs are used to model MFCCs and their temporal derivatives. The method achieves an F-measure of 81.5% in the recognition of kick, snare and hi-hat for pure drum loops and 74.7% for polyphonic music.

41.4.2 Proposed Method

A block diagram describing the proposed method is presented in Fig. 41.4. A similar approach as the one described in [41.67] is taken, assuming that an initial training phase is possible where the specific drum sounds are presented to the detection system. Instead of using elaborate *gamma mixture models*, the spectral bases B are simply initialized with the average over all spectra of the isolated drum sounds. The number of components per drum is denoted c_d. Each drum gets at least one basis vector assigned, but this number can be increased. In order to achieve real-time capability, the NMF decomposition is applied to each spectral frame

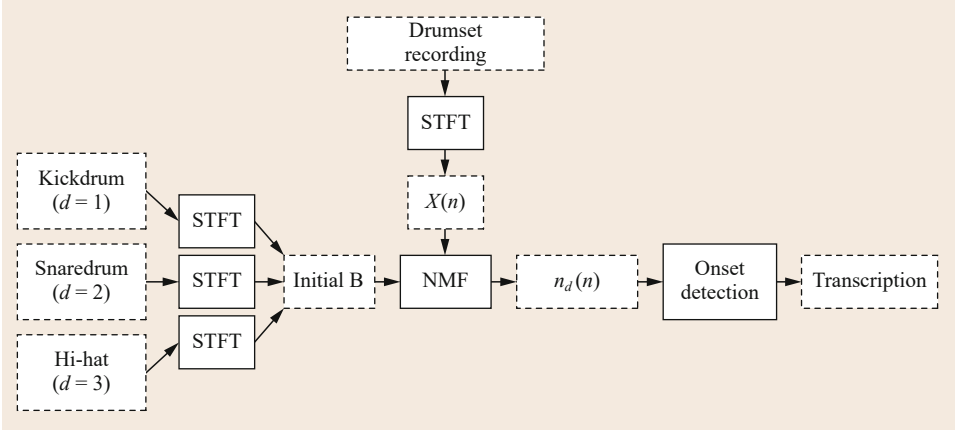


Fig. 41.4 Overview of the proposed method. Initial basis vectors B are trained on isolated drum sounds. Drum set recordings are split into spectral frames $X(n)$ and subjected to NMF. Individual novelty curves $n_d(n)$ are extracted and subjected to onset detection

individually, thus generating a succession of activations for kick, snare and hi-hat in G approximately every 10 ms. The *Kullback–Leiber divergence* (KL) is used as an update rule. As a special trick, the spectral basis functions B are kept semifixed, i. e., they are allowed to deviate more from the initial value the closer the decomposition is to the iteration limit. This is achieved by a linear blending between the initial values and the new ones. This way, the NMF is initially pushed into the direction of the desired drums sounds. Later on, actual deviations of the incoming spectra from the prior spectral templates are accounted for by the update rule.

For drum loops, multiplying each of the activations in G corresponding to a single drum with the corresponding columns in B yields well separated individual spectrograms for kick, snare and hi-hat. On visual inspection, they exhibit very small cross-talk between the different drums. Based on these, onset detection is performed in a straightforward manner. Following the approach proposed in [41.73], a novelty curve $n_d(n)$ is extracted from the successive spectrogram frames for each drum d by differentiating along time in each bin of the logarithmic spectrogram. All bins per frame are summed up and half-wave rectification is applied – only salient positive peaks corresponding to onsets are relevant here. Inevitably, cross-talk artifacts will lead to erroneous spikes and thus, an adaptive threshold on the novelty curve is applied. The threshold is derived by nonlinear compression $n_d(n)^{0.5}$, application of an exponential moving average filter and nonlinear expansion of the result $n_d(n)^{2.0}$. A variable boost factor b can be used to adjust the threshold manually. If the novelty curve is above the adaptive threshold and fulfills other plausibility criteria [41.67], it is marked as an onset. Thus, the whole procedure generates a list of onset times per drum.

41.4.3 Evaluation and Results

In order to assess the transcription performance, experiments with manually annotated real-world drum loops were conducted. The F-measure with an onset tolerance of 50 ms was used as the evaluation metric. A training set was created for initialization of single drums (kick, snare, hi-hat). In order to capture the individual characteristics, the drums were hit separately with varying velocity. For recording, an overhead microphone at a fixed height of 1 m was used. The recordings were made with 10 different drum kits, consisting of different drum sizes and a broad range of materials. The size of the kick drum ranges between 18−24 in in diameter and 16−22 in in depth. Materials were birch, mahogany or maple. The snare drums all had the same size of 14 in in diameter and 6.5 in in depth but differed in their materials (such as metal, wood or acrylic). The sizes for hi-hats differed from 13 to 15 in.

The test set consisted of 30 drum sequences, which were fairly simple groove patterns of kick, snare and closed hi-hat. The tempo of the performed drum grooves varies between 100−140 BPM. Overall, 33 min of audio were recorded (in 44.1 kHz, mono, 16 bit) resulting in 3219 annotated onsets. The shortest annotated interval between consecutive onsets is 107 ms (16th notes at 140 BPM).

With the described data, a grid search to come up with the optimal set of parameters was run. The best average F-measure of ($F = 0.98$) across all drums was obtained with $c_d = 2$ basis vectors per drum, threshold boost $b = 1.5$ and 20 NMF iterations during decomposition. Striving for a more efficient computation, a favorable working point was found around ($F = 0.96$) F-measure when using only $c_d = 1$, $b = 1$ and five iterations.

41.5 Guitar Transcription Beyond Score Notation

As reported in Sect. 41.2.2, the guitar is very popular among musical beginners since it allows relatively quick progress. Hence, algorithms for the automatic transcription of guitar recordings can be used in various music education applications. For fretted string instruments such as the acoustic guitar and electric guitar, a popular way of writing music is called *tablature*. It encodes the notes to be played based on the geometric position on the fretboard, i. e., the string number and fret number. Due to the string tuning of guitars, certain notes with the same pitch can be played on different positions of the fretboard. The tablature notation resolves this ambiguity and provides clear playing instruction to the guitar player. An example of tablature notation is shown in Fig. 41.5. The traditional music score is also shown for reference.

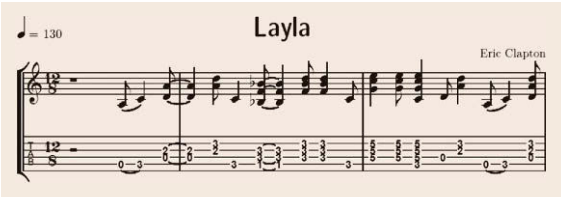


Fig. 41.5 Example of tablature notation and its corresponding score. This figure was originally published in [41.74]

Tablatures gained an especially high popularity among music amateurs since they allow users to circumvent the process of learning to read common music notation. Large amounts of tablatures are freely available online. However, since they are not standardized, they are often incomplete and erroneous. Thus, the goal of this study is the automatic transcription of guitar recordings into tablature notation.

41.5.1 State of the Art

This section describes related works that have dealt with guitar transcription, particularly those concerning tablature extraction.

Sensory Modalities for Data Acquisition

Different *sensory modalities* are used in the literature for analyzing guitar performances or recordings. As discussed in [41.75], the acquisition of physical and perceptual parameters from musical instruments can be categorized into *direct acquisition methods*, which are based on *sensors* that are attached to the instrument, and *indirect acquisition methods*, which are based on *audio* and *visual* analysis of recorded musical performances on the instrument. In addition, some authors propose *multimodal approaches* that combine different sensory modalities in a complementary way.

Audio analysis methods only require an instrument pickup or a microphone for data acquisition. In addition to regular electromagnetic pickups, *hexaphonic pickups* are often used to analyze electric guitar recordings. They allow the capture of the vibration of each individual string, which reduces the problem of polyphonic transcription to a number of parallel monophonic transcription tasks [41.76].

Visual analysis often relies on cameras that are attached to the instrument or positioned close to the performing musician. If the fretboard is recorded, the detection performance can be impaired by bad lighting, masking of the fretboard by the playing hand, or the musician's movement itself.

Mechanically enhanced instruments are extended by sensors that allow a very precise measurement of the spatial hand position. The main disadvantage is that *most of these methods, while accurate, are obtrusive* to the musicians [41.77] since they constrain the natural hand movement on the instrument.

Sensory analysis provides very accurate time-continuous measurements. Also, the movement data measured by motion capturing is closely related to the musician's playing gestures. Signals from capacitive sensors that are used to measure the hand pressure on the instrument fretboard are often noisy and exhibit crosstalk between spatially adjacent sensors.

Guitar Transcription

In order to transcribe a guitar recording and to generate a tablature notation, both the score-level parameters (note pitch, onset and offset) as well as the geometric position (string number and fret number) must be estimated for every note. The first group of publications analyze *monophonic guitar recordings*, i. e., melodies and single notes [41.78, 79]. In order to classify the *string number*, *Barbancho* et al. [41.78] compute various timbre-related spectral audio features such as the inharmonicity, relative magnitude of the overtones, and the temporal decay factor of harmonics. The classification of the *string number* is performed using machine learning algorithms based on the extracted features. The second group of publications focus on *polyphonic guitar recordings*. The estimation of the guitar voicing, i. e., the fretboard position of each finger, is done using *audio analysis* [41.76, 80–82], *visual analysis* [41.83–85], or by combining both modalities in a *multimodal approach* [41.77]. *O'Grady* and *Rickard* perform music transcription on the individual output signals of the hexaphonic guitar pickup [41.76].

Barbancho et al. [41.80] distinguish 330 different fingering configurations for the most common three-voiced and four-voiced guitar chords. Using a multipitch estimation algorithm and an HMM, the most likely chord sequence in a guitar recording is obtained. In *Fiss* and *Kwasinski* [41.81], a multipitch estimation algorithm tailored towards the guitar is presented. Two metrics based on relationships between harmonic frequencies are used to assign the most likely spectral peaks to potential fundamental frequency candidates. *Hrybyk* and *Kim* first estimate the pitch values of all notes in a guitar chord using audio analysis and then estimate the chord voicing using computer vision techniques by tracking the guitar player's hand [41.77]. Different *constraints* are applied in the literature to improve the estimation of spatial parameters. *Yazawa* et al. use the results of the *latent harmonic allocation* (LHA) multipitch estimation and apply three constraints that are tailored to the geometry of the guitar to extract tablatures [41.82]. For instance, the metrics used in [41.81] are based on specific knowledge about the instrument such as the highest possible degree of polyphony (six simultaneous notes) as well as the maximum stretch span of the playing hand within a fixed fretboard position. *Barbancho* et al. use two additional models to constrain the transitions between different HMM states: a musicological model, which captures the likelihood of different chord changes, and an acoustic model, which measures the physical difficulty of changing the chord fingerings [41.80]. In *Dittmar* et al. [41.74], an algorithm capable of real-time guitar string de-

tection is presented, which is also the base for our work.

Estimation of Playing Techniques

The sound production of the guitar can be separated into two *physical gestures* – a *plucking gesture* and an *expressive gesture*. These gestures affect the sonic properties of the recorded instrument notes in a unique way, which allows human listeners to recognize and distinguish different playing techniques. *Frequency modulation techniques* allow the musician to change the note pitch continuously [41.86]. Various publications analyze the estimated fundamental frequency course of single notes to detect the playing techniques *vibrato* [41.87, 88], *bending* [41.89], or *slides* [41.89–91]. In between consecutive note events, different *note transition techniques* can be applied. Common techniques are for instance the *slide* or the *hammer-on* and *pull-off* techniques as investigated in [41.88–90]. *Erkut* et al. analyze repeated string plucks in [41.87] and *Guaus* et al. studied the hand movement if *grace notes* are played [41.88].

41.5.2 Proposed Method

In continuation of the work in [41.74], a guitar transcription algorithm was developed that combines the conventional score-level transcription with the note-wise estimation of the fretboard positions and the playing techniques. The algorithm can be used to transcribe isolated monophonic and polyphonic guitar recordings. A block diagram of the proposed method is shown in Fig. 41.6.

The note onset detection is based on a combined analysis of the spectral flux, the rectified complex domain, as well as a novel onset detection function, which detects beginning overtones [41.92]. Based on a reassigned spectrogram (using the IF spectrogram), the Blind Harmonic Adaptive Decomposition (BHAD) multipitch estimation algorithm [41.34] is applied to get a multipitch estimate for each note event. In order to cope with frequency modulation techniques such as vibrato, bending, and slides, the fundamental frequency is tracked frame-wise over the note duration. The note offset is obtained by tracking the decaying note magnitude envelope.

For each note, the plucking style, i. e., the playing technique used to pluck the guitar string, is separately estimated from the the expression style, i. e., the way in which the playing hand is used to manipulate the string vibration. Based on various audio features that describe the inharmonicity and the magnitude and frequency relationship between overtones, the string number is automatically classified using a SVM classifier. Due to

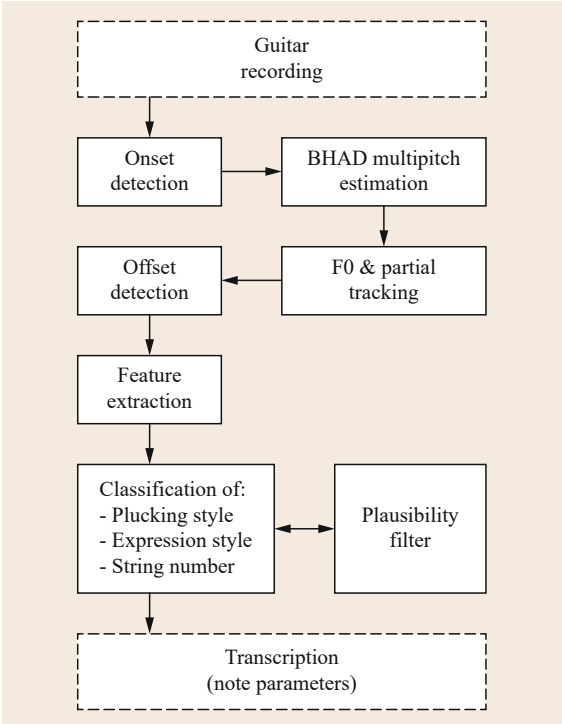


Fig. 41.6 Block diagram of the proposed method

the instrument geometry and the string tuning, different constraints can be applied if multiple notes sound at the same time. This allows corrections to be performed for both the pitch and string estimation. Similarly to the string number, the plucking style and the expression style are also automatically classified using a large set of note-wise audio features. A plausibility filter is applied to correct meaningless classification results based on the estimated note parameters of the previously detected notes.

41.5.3 Evaluation and Results

For the purpose of evaluation, a novel database of guitar recordings was created. It contains 261 audio recordings of both isolated notes as well as monophonic and polyphonic melodies, licks, and chord sequences with extensive note-wise annotations of all discussed parameters. Three different plucking styles (finger-style, picked, palm muted) as well as six different expression styles (normal, vibrato, slide, bending, harmonics, dead-notes) are included. Three different electric guitars were used in the recording session. In total, around 4700 guitar note events are contained in the database.

The score-level transcription steps perform very well as indicated by very high F-measure values for onset detection ($F = 0.985$), pitch detection ($F = 0.96$),

and offset detection ($F = 0.977$) on both monophonic and polyphonic guitar licks in the dataset. High accuracy values were achieved in cross-validation experiments for the classification of the plucking style ($A = 0.93$), expression style ($A = 0.83$), and sting number ($A = 0.82$) [41.93].

41.6 Discussion and Future Challenges

A general overview of the use of music technology in music education has been presented. Additionally, three special cases of technologies developed within the MIR community, successfully applied to music education, have been described. Even when research in the past years has given a tremendous push to the development of music application and education tools, there are still many challenges faced by the research community that need to be overcome before a seamless conjunction between music education and technology can be achieved.

The field of sound source separation has produced very promising results in the last years, especially with different types of informed source separation. Some systems are commercially available and specialized sound separation services are offered by some individuals and audio processing companies. However, the performance of the technologies also very closely depends on the type of audio recording to be processed and on the separation task itself. In most cases, even when solutions for certain types of signals and tasks have been developed, a great amount of manual refinement and detailed postprocessing is still needed to obtain good quality of separated signals. To date, a general solution to the sound separation problem is not available.

In drum transcription research, methods still need to be enhanced to be able to handle a full drum set with all its components. Even though the most popular drum choices have been addressed (kick drum, snare drum, and hi-hat), other commonly used drums like ride cymbals, toms, crash cymbals, and splash cymbals have not been studied. Additionally, other commonly used percussion instruments such as the cowbell, woodblock, tambourine, gong, and triangle also present their own spectral and acoustical characteristics and need to be studied before a robust transcription can be achieved. Furthermore, extensive research efforts need to be directed towards guaranteeing performance robustness when the complexity of the drum patterns is higher and more challenging tempi are used. Lastly, the difficulties of performing the same type of processing in polyphonic audio, where several instruments simultaneously play, greatly increase the complexity of the task.

Guitar transcription research also presents its own particular challenges. The differences between acoustic guitars and electric guitars need to be addressed to guarantee robustness of results for all users. Not only do the audio capturing methods for the two types of guitars usually differ, but the performance style, fingering, and expression style can also be considerably different. Additionally, similarly to drum transcription research, more studies have to be conducted to extend these technologies to polyphonic audio where more complex scenarios occur.

Generally speaking, there are broader challenges that both the music technology and music education communities need to address in order to have technologies and applications that can truly and effectively benefit music education. A critical point of current research is the fact that for the great majority of technologies developed, performance is greatly dependent on the type of audio file processed. Different instrumentations, musical genres, recording conditions, and musical styles still represent a major challenge for MIR research. Even when good results can be achieved under a determined set of conditions and with clear constraints, general solutions capable of handling audio files of various characteristics are still far from being available. After many years of research in the field, results and the current state-of-the-art methods tend to suggest that more instrument-specific algorithms and studies are needed to properly capture the specific characteristics, not only acoustical but also stylistic, of the different musical instruments. Performance possibilities of musical instrument are so diverse and rich that a general algorithm that can handle all instruments under all conditions is hard to conceive.

When it comes to the usability of music technologies in music education, more user studies and listening tests need to be conducted that can properly evaluate users' responses to the technologies. Additionally, the effectiveness of the technologies to achieve a learning goal also needs to be evaluated. For this matter, a very close collaboration between music education, MIR, psychology, and musicology is required.

A final note has to be made regarding the importance of algorithm optimization in the MIR field. Even though technologies have progressed tremendously in the last years, it is still fairly common for algorithms to be too time consuming and computationally demanding

for them to be applicable and helpful in real-world applications such as music education. There is still a lot of work ahead of us in the endeavor of optimizing algorithms and reaching real-time functionalities necessary for technologies to be truly effective in many fields.

References

41.1 C. Dittmar, E. Cano, J. Abeßer, S. Grollmisch: Music information retrieval meets music education. In: *Multimodal Music Process. Dagstuhl Follow-Ups*, ed. by M. Müller, M. Goto, M. Schedl (Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Wadern 2012) pp. 95–120

41.2 Jamey Aebersold Jazz: The original jazz play-a-longs, http://www.jazzbooks.com/jazz/category/AEBPLA (2014)

41.3 Hal Leonard Corporation: Jazz play-along, http://www.halleonard.com/promo/promo.do?promotion=590001&subsiteid=1 (2014)

41.4 Music Minus One: http://www.musicminusone.com (2014)

41.5 Alfred Music Publishing: Alfred music DVD, http://www.alfred.com/Browse/Formats/DVD.aspx (2014)

41.6 Berklee Press: Berklee press DVD titles, http://www.berkleepress.com/catalog/product-type-browse?product_type_id=10 (2014)

41.7 Icons of Rock: http://www.iconsofrock.com/ (2014)

41.8 Homespun: Homespun music instruction, http://www.homespuntapes.com/home.html (2014)

41.9 Drumeo: The ultimate online drum lesson experience, http://drumeo.com/ (2014)

41.10 Get2Play: Learn instruments easily online, http://www.get2play.com/ (2014)

41.11 GuitarHero: http://www.guitarhero.com (2014)

41.12 Singstar: Singstar, http://www.singstar.com (2014)

41.13 Ubisoft: Rocksmith, http://rocksmith.ubi.com/rocksmith/en-US/home/ (2014)

41.14 Music Delta: http://www.musicdelta.com (2014)

41.15 Synthesia: A fun way lo learn how to play the piano, http://www.synthesiagame.com/ (2014)

41.16 Gigajam: Creating musicians, http://gigajam.com (2014)

41.17 Makemusic: Smart music, http://www.smartmusic.com (2014)

41.18 Apple Inc: Garage band, http://www.apple.com/ilife/garageband/ (2014)

41.19 Fraunhofer IDMT: Songs2See: Learn to play by playing, http://songs2see.com (2014)

41.20 E. Cano, C. Dittmar, S. Grollmisch: Songs2See: Learn to play by playing. In: *12th Int. Soc. Music Inf. Retr. Conf. (ISMIR)* (2011)

41.21 Rock Prodigy: Rock prodigy, http://www.rockprodigy.com/ (2014)

41.22 Tonara: http://tonara.com/ (2014)

41.23 i-Maestro: Interactive multimedia environment for technology enhanced music education and creative collaborative composition and performance, http://cordis.europa.eu/projects/rcn/80567_en.html (2014)

41.24 Vemus: Virtual european music school, http://cordis.europa.eu/projects/rcn/80596_en.html (2014)

41.25 IMUTUS: Interactive music tuition system, http://www.ilsp.gr/en/infoprojects/meta?view=project&task=show&id=120 (2014)

41.26 A. Cont: ANTESCOFO: Anticipatory synchronization and control of interactive parameters in computer music. In: *Proc. Int. Comput. Music Conf. (ICMC)* (2008)

41.27 R. Christopher: Music plus one and machine learning. In: *27th Int. Conf. Mach. Learn.* (2010)

41.28 J. Abeßer, J. Hasselhorn, C. Dittmar, A. Lehmann, S. Grollmisch: Automatic quality assessment of vocal and instrumental performances of 9th-grade and 10th-grade pupils. In: *Proc. 10th Int. Symp. Comput. Music Multidiscip. Res. (CMMR)* (2013)

41.29 E. Cano: Solo and accompaniment separation: Towards its use in music education applications, http://www.idmt.fraunhofer.de/en/Departments_and_Groups/smt/solo_and_accompaniment_separation.html (2013)

41.30 A. Liutkus, J. Durrieu, L. Daudet, G. Richard: An overview of informed audio source separation. In: *Proc. 14th Int. Workshop Image Audio Anal. Multimed. Interact. Serv.* (2013) pp. 3–6

41.31 K. Dressler: An auditory streaming approach for melody extraction from polyphonic music. In: *Proc. 12th Int. Soc. Music Inf. Retr. Conf. (ISMIR)* (2011) pp. 19–24

41.32 J. Salamon, E. Gómez, D.P. Ellis, G. Richard: Melody extraction from polyphonic music signals: Approaches, applications and challenges, IEEE Signal Process. Mag. **31**(2), 118–134 (2014)

41.33 E. Cano, C. Dittmar, G. Schuller: Efficient implementation of a system for solo and accompaniment separation in polyphonic music. In: *Proc. 10th Eur. Signal Process. Conf. (EUSIPCO)* (2012) pp. 285–289

41.34 B. Fuentes, R. Badeau, G. Richard: Blind harmonic adaptive decomposition applied to supervised source separation. In: *Proc. 20th Eur. Signal Process. Conf. (EUSIPCO)* (2012) pp. 2654–2658

41.35 R. Marxer, J. Janer, J. Bonada: Low-latency instrument separation in polyphonic audio using timbre models, Latent Var. Anal. Signal Sep. **7191**, 314–321 (2012)

41.36 J. Fritsch, M.D. Plumbley: Score informed audio source separation using constrained non-negative matrix factorization and score synthesis. In: *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)* (2013) pp. 888–891

41.37 J. Ganseman, P. Scheunders, G.J. Mysore, J.S. Abel: Evaluation of a score-informed source separation system. In: *11th Int. Soc. Music Inf. Retr. Conf. (ISMIR)* (2010)

41.38 C. Joder, B. Schuller: Score-informed leading voice separation from monaural audio. In: *13th Int. Soc. Music Inf. Retr. Conf. (ISMIR)* (2012) pp. 277–282

41.39 M. Coic, J.J. Burred: Bayesian non-negative matrix factorization with learned temporal smoothness priors. In: *Int. Conf. Latent Var. Anal. Signal Sep. (LVA/ICA)* (2012) pp. 280–287

41.40 J.J. Burred, A. Röbel: A segmental spectro-temporal model of musical timbre. In: *13th Int. Conf. Dig. Audio Eff. (DAFx-10)* (2010) pp. 1–7

41.41 Y. Li, J. Woodruff, D. Wang: Monaural musical sound separation based on pitch and common amplitude modulation, IEEE Trans. Acoust. Speech Signal Process. **17**(7), 1361–1371 (2009)

41.42 J. Bosch, K. Kondo, R. Marxer, J. Janer: Score-informed and timbre independent lead instrument separation in real-world scenarios. In: *Proc. 20th Eur. Signal Process. Conf. (EUSIPCO)* (2012) pp. 2417–2421

41.43 J. Durrieu, B. David, G. Richard: A musically motivated mid-level representation for pitch estimation and musical audio source separation, IEEE J. Sel. Top. Signal Process. **5**(6), 1180–1191 (2011)

41.44 P. Huang, S. Chen, P. Smaragdis, M. Hasegawa-Johnson: Singing-voice separation from monaural recordings using robust principal component analysis. In: *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)* (2012) pp. 57–60

41.45 Z. Rafii, B. Pardo: Repeating pattern extraction technique (REPET): A simple method for music/voice separation, IEEE Trans. Audio Speech Lang. Process. **21**, 73–84 (2013)

41.46 J. Janer, R. Marxer: Separation of unvoiced fricatives in singing voice mixtures with music semi-supervised NMF. In: *Proc. 16th Int. Conf. Dig. Audio Eff. (DAFx-13)* (2013) pp. 1–4

41.47 C. Févotte, N. Bertin, J.L. Durrieu: Nonnegative matrix factorization with the Itakura–Saito divergence: With application to music analysis, Neural Comput. **21**(3), 793–830 (2009)

41.48 J.-L. Durrieu, J.-P. Thiran: Musical audio source separation based on user-selected F0 track, Latent Var. Anal. Signal Sep. **7191**, 438–445 (2012)

41.49 U. Simsekli, A. Cemgil: Score guided musical source separation using generalized coupled tensor factorization. In: *Proc. 20th Eur. Signal Process. Conf. (EUSIPCO)* (2012) pp. 2639–2643

41.50 D. FitzGerald: User assisted separation using tensor factorisations. In: *20th Eur. Signal Process. Conf. (EUSIPCO)* (2012) pp. 2412–2416

41.51 L. Benaroya, L. Donagh, F. Bimbot, R. Gribonval: Non-negative sparse representation for Wiener based source separation with single sensor. In: *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)* (2003) pp. 613–616

41.52 J. Le Roux, E. Vincent, Y. Mizuno, K. Hirokazu, N. Ono, S. Sagayama: Consistent Wiener filtering: Generalized time-frequency masking respecting spectrogram consistency. In: *Int. Conf. Latent Var. Anal. Signal Sep. (LVA/ICA)* (2010)

41.53 D. Fitzgerald: Harmonic/percussive separation using median filtering. In: *13th Int. Conf. Dig. Audio Eff. (DAFx-10)* (2010) p. 10

41.54 A. Bregman: *Auditory Scene Analysis* (MIT Press, Cambridge 1990)

41.55 V. Emiya, E. Vincent, N. Harlander, V. Hohmann: Subjective and objective quality assessment of audio source separation, IEEE Trans. Audio Speech Lang. Process. **19**(7), 2046–2057 (2011)

41.56 Signal Separation Evaluation Campaign (SiSEC): https://sisec.wiki.irisa.fr/tiki-index.html (2013)

41.57 E. Cano, G. Schuller, C. Dittmar: Pitch-informed solo and accompaniment separation towards its use in music education applications, EURASIP J. Adv. Signal Process. **23**, 1–19 (2014)

41.58 J. Paulus: *Signal Processing Methods for Drum Transciption and Music Structure Analysis*, Ph.D. Thesis (Tampere Univ. Technol., Tampere 2009)

41.59 M. Casey: Separation of mixed audio sources by independent subspace analysis. In: *Proc. Int. Comput. Music Conf.* (2000)

41.60 C. Uhle, C. Dittmar, T. Sporer: Extraction of drum tracks from polyphonic music using independent subspace analysis. In: *Proc. 4th Int. Symp. Indep. Compon. Anal. Blind Signal Sep.* (2003)

41.61 M. Plumbley: Algorithms for non-negative independent component analysis, IEEE Trans. Neural Netw. **14**, 30–37 (2003)

41.62 C. Dittmar, C. Uhle: Further steps towards drum transcription of polyphonic music. In: *Proc. AES 116th Conv.* (2004)

41.63 D. FitzGerald, B. Lawlor, E. Coyle: Prior subspace analysis for drum transcription. In: *Proc. AES 114th Conv.* (2003)

41.64 A. Spich, M. Zanoni, A. Sarti, S. Tubaro: Drum music transcription using prior subspace analysis and pattern recognition. In: *Proc. 13th Int. Conf. Dig. Audio Eff. (DAFx)* (2010)

41.65 M. Helén, T. Virtanen: Separation of drums from polyphonic music using non-negative matrix factorization and support vector machine. In: *Proc. 13th Eur. Signal Process. Conf. (EUSIPCO)* (2005)

41.66 J. Paulus, A. Klapuri: Drum transcription with non-negative spectrogram factorisation. In: *Proc. 13th Eur. Signal Process. Conf. (EUSIPCO)* (2005)

41.67 E. Battenberg, V. Huang, D. Wessel: Live drum separation using probabilistic spectral clustering based on the Itakura-Saito divergence. In: *Proc. AES 45th Conf. Time-Freq. Process. Audio* (2012)

41.68 K. Yoshii, M. Goto, H. Okuno: Automatic drum sound description for real-world music using template adaption and matching methods. In: *Proc. 5th Int. Conf. Music Inf. Retr. (ISMIR)* (2004)

41.69 C. Dittmar, D. Wagner, D. Gärtner: Drumloop separation using adaptive spectrogram templates. In: *Proc. 36th Jahrestag. Akust. (DAGA)* (2010)

41.70 A. Maximos, A. Floros, M. Vrahatis, N. Kanellopoulos: Real-time drums transcription with characteristic bandpass filtering. In: *Proc. 7th Audio Mostly Conf.* (2012)

41.71 O. Gillet, G. Richard: Automatic transcription of drum loops. In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)* (2004)

41.72 K. Tanghe, S. Degroeve, B. De Baets: An algorithm for detecting and labeling drum events in polyphonic music. In: *Proc. 1st Ann. Music Inf. Retr. Eval. eXchange (MIREX '05)* (2005)

41.73 P. Grosche, M. Müller: Extracting predominant local pulse information from music recordings, IEEE Trans. Audio Speech Lang. Process. **19**(6), 1688–1701 (2011)

41.74 C. Dittmar, A. Männchen, J. Abeßer: Real-time guitar string detection for music education software. In: *Proc. 14th Int. Workshop Image Anal. Multimed. Interact. Serv. (WIAMIS)* (2013) pp. 1–4

41.75 A. Carrillo, M. Wanderley: Learning and extraction of violin instrumental controls from audio signal. In: *Proc. 2nd Int. ACM Workshop Music Inf. Retr. User-Centered Multimodal Strateg. (MIRUM)* (2012) pp. 25–30

41.76 P. O'Grady, S. Rickard: Automatic hexaphonic guitar transcription using non-negative constraints. In: *Proc. IET Ir. Signals Syst. Conf. (ISSC)* (2009)

41.77 A. Hrybyk, Y. Kim: Combined audio and video for guitar chord identification. In: *Proc. 11th Int. Soc. Music Inf. Retr. Conf. (ISMIR)* (2010) pp. 159–164

41.78 I. Barbancho, A. Barbancho, L. Tardón, S. Sammartino, L. Tardón: Pitch and played string estimation in classic and acoustic guitars. In: *Proc. 126th Audio Eng. Soc. (AES) Conv.* (2009)

41.79 J. Abeßer: Automatic string detection for bass guitar and electric guitar. In: *Sounds Music Emot. – 9th Int. Symp., CMMR, London 2012* (2013) pp. 333–352

41.80 A. Barbancho, A. Klapuri, L. Tardòn, I. Barbancho: Automatic transcription of guitar chords and fingering from audio, IEEE Trans. Audio Speech Lang. Process. **20**, 915–921 (2011)

41.81 X. Fiss, A. Kwasinski: Automatic real-time electric guitar audio transcription. In: *Proc. IEEE Conf. Acoust. Speech Signal Process. (ICASSP)* (2011) pp. 373–376

41.82 K. Yazawa, D. Sakaue, K. Nagira, K. Itoyama, H. Okuno: Audio-based guitar tablature transcription using multipitch analysis and playability constraints. In: *Proc. 38th IEEE Conf. Acoust. Speech Signal Process. (ICASSP)* (2013) pp. 196–200

41.83 A. Burns, M. Wanderley: Visual methods for the retrieval of guitarist fingering. In: *Proc. 2006 Int. Conf. New Interfaces Music. Expr. (NIME06)* (2006) pp. 196–199

41.84 C. Kerdvibulvech, H. Saito: Vision-based guitarist fingering tracking using a Bayesian classifier and particle filters, Lect. Notes Comput. Sci. **4872**, 625–638 (2007)

41.85 M. Paleari, B. Huet, A. Schutz, D. Slock: A multimodal approach to music transcription. In: *Proc. 15th IEEE Int. Conf. Image Process. (ICIP)* (2008) pp. 93–96

41.86 J. Abeßer, C. Dittmar, G. Schuller: Automatic Recognition and parametrization of frequency modulation techniques in bass guitar recordings. In: *Proc. 42nd Audio Eng. Soc. (AES) Conf. Semant. Audio* (2011)

41.87 C. Erkut, M. Karjalainen, M. Laurson: Extraction of physical and expressive parameters for model-based sound synthesis of the classical guitar. In: *Proc. 108th Audio Eng. Soc. (AES) Conv.* (2000) pp. 19–22

41.88 E. Guaus, T. Özaslan, E. Palacios, J. Arcos: A left hand gesture caption system for guitar based on capacitive sensors. In: *Proc. 10th Int. Conf. New Interfaces Music. Expr. (NIME)* (2010) pp. 238–243

41.89 L. Reboursière, 0. Lähdeoja, T. Drugman, S. Dupont, C. Picard-Limpens, N. Riche: Left and right-hand guitar playing techniques detection. In: *Proc. Int. Conf. New Interfaces Music. Expr. (NIME)* (2012) pp. 1–4

41.90 T. Özaslan, E. Guaus, E. Palacios, J. Arcos: Attack based articulation analysis of nylon string guitar. In: *Proc. 7th Int. Symp. Comput. Music Model. Retr. (CMMR)* (2010) pp. 285–298

41.91 T. Özaslan, J. Arcos: Legato and glissando identification in classical guitar. In: *Proc. Sound Music Comput. Conf. (SMC), Barcelona* (2010) pp. 457–463

41.92 J. Abeßer, G. Schuller: Instrument-centered music transcription of bass guitar tracks. In: *Proc. AES 53rd Conf. Semant. Audio* (2014)

41.93 C. Kehling: *Entwicklung eines parametrischen Instrumentencoders basierend auf Analyse und Re-Synthese von Gitarrenaufnahmen*, Diploma Thesis (Technische Universität Ilmenau, Ilmenau 2013)

42. Music Learning: Automatic Music Composition and Singing Voice Assessment

Lorenzo J. Tardón, Isabel Barbancho, Carles Roig, Emilio Molina, Ana M. Barbancho

Traditionally, singing skills are learned and improved by means of the supervised rehearsal of a set of selected exercises. A music teacher evaluates the user's performance and recommends new exercises according to the user's evolution.

In this chapter, the goal is to describe a virtual environment that partially resembles the traditional music learning process and the music teacher's role, allowing for a complete interactive self-learning process.

An overview of the complete chain of an interactive singing-learning system including tools and concrete techniques will be presented. In brief, first, the system should provide a set of training exercises. Then, it should assess the user's performance. Finally, the system should be able to provide the user with new exercises selected or created according to the results of the evaluation.

Following this scheme, methods for the creation of user-adapted exercises and the automatic evaluation of singing skills will be presented. A technique for the dynamical generation of musically meaningful singing exercises, adapted to the user's level, will be shown. It will be based on the proper repetition of musical structures, while assuring the correctness of harmony and rhythm. Additionally, a module for singing assessment of the user's performance, in terms of intonation and rhythm, will be shown.

In this chapter, we present several methods and techniques to implement a complete educational tool for learning to sing. Typically, singing skills are improved by rehearsing a set of appropriate exercises under the supervision of a music teacher. The role of this music teacher is to evaluate the user's performance, and to recommend new exercises according to the user's evolution. Therefore, the presented methods allow the creation of user-adapted exercises and evaluation of the singing skills of the user automatically right after the performance. The goal of this combined approach is to create a virtual environment that partially resembles the music teacher's role, leading to a faster self-learning process.

Two main submodules are presented in this chapter: (1) an automatic generator of singing exercises, and (2) a singing assessment module, which analyses the user's voice in order to rate the quality of the singing performance. In the first, the generated singing exercises are musically meaningful, based on repeated structures, and can be adapted to suit the level of the user. In the second, the module for singing assessment compares the user's performance with respect to the automatically generated singing exercise, and rates the user's performance with two criteria: intonation and rhythm. In Fig. 42.1, a block diagram of the complete system is shown.

Using these methods, the singing learning process becomes an iterative self-guided process. First, the system provides a set of exercises. Second, the user sings the suggested practices. Third, the system asseses the user's performance. And finally, it suggests new scores according to the grade obtained. Note that the scores generated are not precomposed but dynamically generated according to the current level of the user. In this way, a gradual and fully adapted learning process is assured.

This chapter is organized as follows. In Sect. 42.1 and 42.2, we present the related work on melody com-

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_42

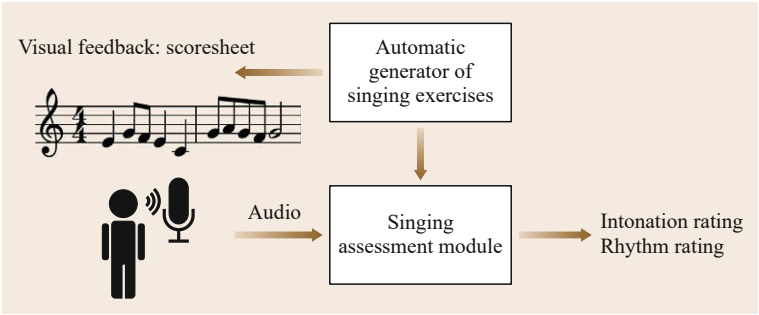


Fig. 42.1 Block diagram of the complete system

position and automatic singing assessment respectively. Then, in Sect. 42.3 an automatic generator of singing exercises is described. A scheme for automatic singing assessment is described in Sect. 42.4. Finally, Sect. 42.5 draws some conclusions about the tool for singing learning presented in this chapter.

42.1 Related Work on Melody Composition

The incorporation of different skills in the field of music information retrieval related to the computational analysis and description of musical pieces allows us to face different tasks like automatic music transcription [42.1], or the identification of relations between songs [42.2], among others. Furthermore, the computational model of the human experience in the musical field and the human brain processes in this field are of great interest for psychology and musicology [42.3].

In this context, the automatic generation of musical content is the topic considered. Often, music is defined as *organized sound* [42.4] with order and structure. Thus, systems to generate music must be trained beforehand to learn a logic composition style as stated in [42.5]. For this reason, an algorithm for learning composition rules and patterns is outlined.

A set of descriptors must be analyzed in order to model the style of the melodies. Concerning the temporal descriptors, tempo and time signature, previous works can be found. The work presented in [42.6] is focused on the onset estimation based on the spectral analysis. In [42.7], histograms to find the most repeated interonset values are employed. Recently, methods for structural analysis such as [42.8] based on a timespan tree for the structural similarity detection, which can be addressed making use of the autosimilarity matrix [42.9], were found.

In this chapter we consider an innovative approach for tempo estimation based on the interonset interval (IOI) histogram inspired by [42.6], followed by a fine adjustment stage.

Regarding music composition, apart from composition schemes based on pattern reallocation and variations, other methods are described in the bibliography. In [42.10–12] Markov models are used for the modeling and composition processes. The use of genetic algorithms such as *Biles*' GenJam system [42.13] are also present in the field of automatic music composition. Methods based on probabilistic approaches such as *Cope*'s experiments in musical intelligence (EMI) [42.14, 15] also focus on the creation of an automatic music composition framework. Inmamusys [42.16] is another composition scheme based on probabilistic structures. This method is very similar to the one considered here, since both use previously learned patterns to generate a new composition from the reallocation of them. However, we considered the presence in the composition system of a postprocessing stage, intended to make all the motives in the database learned fit in the composition. Note that Inmamusys restricts the combination of motives to subsets previously tagged as compatible.

42.2 Related Work on Voice Analysis for Assessment

Regarding the evaluation of singing voice, the literature reports a number of schemes for automatic singing assessment [42.17–27]. These schemes are able to provide feedback about the user's singing performance.

Commonly, in order to attain the desired objectives, the audio is processed according to the following steps. First, a low-level feature extraction process is performed to find a set of frame-level vectors with

meaningful information about the input. In the case of singing analysis, the most important feature is fundamental frequency F_0, although most of the approaches also use other features such as energy, aperiodicity, zero-crossing rate or certain auditory-based features. In the literature, a wide set of approaches for F_0 estimation have been proposed, some of which are based on the time domain, whereas others are based on the frequency domain (see [42.28] for a comprehensive review). One of the most-used approaches is the Yin algorithm [42.29], since it is simple, effective and easily accessible. The Yin algorithm was developed by de *Cheveigné* and *Kawahara* in 2002 [42.29], and it has been found to be effective in many monophonic music transcription systems [42.26, 30–32].

Then, the feature(s) extracted is postprocessed in order to identify voiced regions (the *voicing* process) and, in many cases, a later note-level segmentation is also performed. The estimation of voiced sounds can be performed using a wide variety of descriptors at frame-level: F_0 stability [42.33], root-mean square (RMS) [42.34], aperiodicity [42.35], or zero-crossing rate [42.36], etc.

Additionally, a note-level segmentation process of the singing voice (also called singing transcription) must be performed. To this end, some systems analyze the low-level feature(s) using heuristic rules and a set of thresholds [42.34, 37], whereas other approaches are based on probabilistic models, especially hidden Markov models [42.35, 38].

Finally, the assessment of singing skill is performed by analyzing the postprocessed low-level features and/or the note-level segmentation of the audio input. Prior works have led to various solutions for automatic singing rating. In general, all these systems focus on intonation assessment with visually attractive real-time feedback. Some of these systems use a reference melody (considered the target performance) in order to assess the user's performance, whereas other approaches are melody independent.

42.3 Music Composition for Singing Assessment

In order to be able to accurately design automatic music composition methods, it is necessary to know the parameters involved in the composition. In this section, we present both the parameters used by a novel autonomous music compositor that generates new melodies using a statistical model and the composition scheme itself. Different aspects related to the traditional way in which music is composed by humans such as harmony and structure repetitions will be considered.

The approach is focused on an educational context. The student should be able to automatically generate reinforcement melodies according to a particular musical level enlarging the number of available training exercises.

42.3.1 Learning Musical Parameters

The approach designed for the generation of contents is based on the music theory method called *ostinato* [42.39]. This method considers the composition of music on the basis of pattern repetition with harmonic variations in such a way that the repetition of the motives creates the melody structure.

Thus, rhythm patterns, pitch contours, harmonic progressions and tempo structures must be learned [42.40].

Thus, a database of musical parameters can be used to model the training level of certain musical pieces, as in [42.41]. Since the main objective is to develop a music model for the automatic creation of compositions with style replication, the discovery of this type of information and the development of specific procedures to make use of the different pieces of information to model music corresponding to different training levels are considered. This can be done on the basis of a probabilistic analysis of rhythm and pitch patterns stored in a database filled with music samples of different complexity levels. In Fig. 42.2, a diagram of a suitable analysis system is presented.

According to the characterization parameters required, the database can be divided into three levels

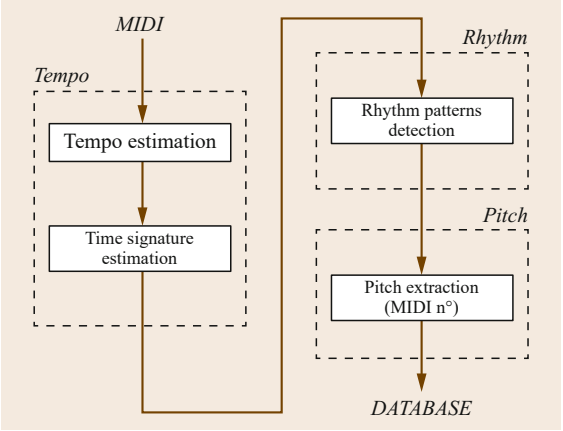


Fig. 42.2 Illustrative scheme of the music analysis system

hierarchically organized corresponding to: (L-1) time signatures, (L-2) rhythm patterns and (L-3) pitch contours (Fig. 42.3).

The measures the training samples will be split into will be the elements stored in the database. These elements will be used for the composition of novel music scores. In order to achieve this goal, the bar length has to be established for proper measure splitting. The bar length can be manually set but also a system to perform this task automatically can be devised. Thus, tempo estimation is required in first place.

The tempo can be extracted easily by analyzing Musical Instrument Digital Interface (MIDI) metadata messages [42.42], if present. But this information can also be incorrectly stored. In order to develop a robust estimation scheme, an algorithm to estimate the tempo and time signature from MIDI files can be used. Note that the availability of correct tempo information is critical in order to relate the duration of the notes obtained by means of the analysis of Note On and Note Off messages [42.42] to musical figures.

Note that according to [42.43], the target parameters in this work are the basic ingredients for the composition of music: rhythm, pitch motives (melody) and harmony (which is considered at the score composition stage).

Now, we consider the specific estimation stages.

Temporal Estimations

The tempo and the time signature have to be estimated in order to correctly perform bar separation and properly split rhythmic patterns. Note that here we consider a rhythmic pattern to be equivalent to a complete measure from the input training data.

Tempo Estimation. The algorithm considered for tempo estimation is inspired by the work presented in [42.7]. However, in our scenario, the analyzed IOIs are directly extracted from the melody. Initially, the most repeated IOI value can be considered a candidate pulse, or tactus [42.44]. This pulse is related to tapping or dancing while listening to a piece of music [42.45]. However, some considerations must be taken into account:

- Resting periods are not explicitly extracted. MIDI files contain information about the notes solely (Note On and Note Off events [42.42]). However, resting periods can be indirectly extracted and must be used to properly estimate the tempo.
- The tactus extracted can be a multiple or a divisor of the actual tactus. By setting a valid tempo range, this value can be corrected.
- The tactus estimated will not be the exact one due to the discrete nature of the histogram. The value can be finely corrected in a postprocessing stage.

The objective of the fine adjustment of the tactus is to find the value that causes the lowest displacement from the constant beat and the input file onsets. This can be achieved by defining a specific model to interpolate the histogram of the IOIs.

Then, the tactus has to be associated with a certain rhythmic element to define the duration of the quarter note and for the estimation of the tempo. Since the range of valid tempos in music is large – from *Largo* ($40 \sim 60$ quarters per minute) to *Presto* ($180 \sim 200$ quarters per minute) [42.46] – the tempo range must be manually reduced in order to establish an accurate relation between durations and musical elements. A suitable hypothesis is that the tempo of the training data is *Moderato* ($76 \sim 108$ quarters per minute). Anyway, a mapping can always be defined by calculating the duration of each rhythm figure in the range selected.

As described in [42.47], the tempo estimation algorithm often estimates doubled or halved tempos. This is a normal behavior caused by the tempo concept itself [42.48]. The tempo is actually subjective, which means that some editors may use shorter rhythm figures reducing the selected tempo, while other can do the opposite, to represent the same performance speed.

Finally, observe that the duration of the metric figures is well known after the tempo is estimated. The duration of each measure can be obtained by multiplying the duration of the pulse by the number of pulses that fit in one bar. So, the next goal is estimate the bar duration. The approach for this purpose can be based on the evaluation of different bar split scenarios for several tentative lengths. Then, some features like the number of repeated bars, the number of bars and the number of split notes can be considered.

Time Signature Estimation. The estimation of the time signature can be based on the analysis of bar repetitions, which can be done using a multiresolution analysis scheme [42.49, 50] to obtain the bar length that best suits the input melody among a set of candidates.

The rhythm self-similarity matrix (RSSM), as described in [42.51], is useful for the purpose at hand. The input melodies can be split into the k candidate bar lengths in order to build the RSSM using the tactus as a unit. Note that the analysis system will create k RSSM matrices, one for each of the candidates. Using those

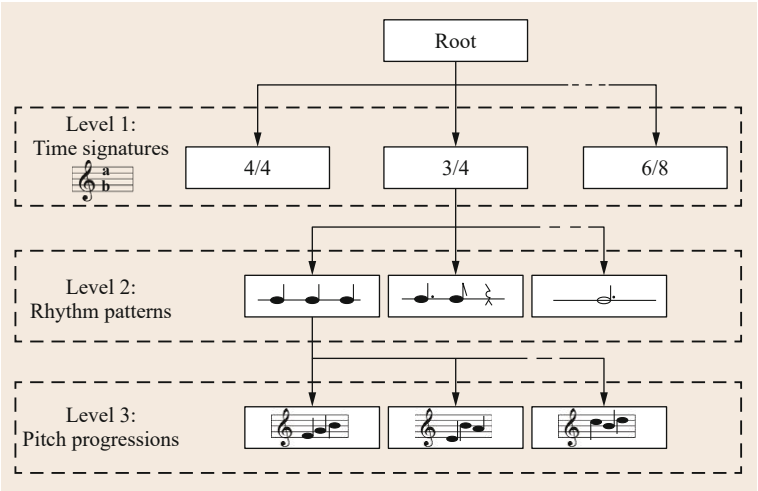


Fig. 42.3 Representation of the hierarchical database considered to store and organize music parameters for music composition based on pattern repetitions

RSSMs, the following descriptors can be extracted and considered in the estimation process:

- Number of repeated bars: the amount of different bars repeated along the split performed for a certain length candidate. The most repeated bar will, more probably, perform a proper separation.
- Number of repeated bar instances: the average number of instances per repeated bar. The larger this number, the more probable the separation will be.
- Number of ties between bars: the number of notes divided between two bars for the candidate length. The lower the number of ties between bars, the more probable the separation will be.
- Number of detected bars: the number of bar splits in the original stream. The number of bars tends to be a power of two.

Note that the number of repeated bars, bar instances and ties must be normalized by the total number of bars detected to give rise to comparable measures. In order to classify the time signature of the input melodies using these descriptors, different classification schemes can be considered, such as the J48 decision tree classifier [42.52], or another one based on sequential minimal optimization (SMO) [42.53], which are available in the Weka machine learning software suite [42.54].

Rhythm Patterns

Rhythm is probably the musical feature more closely related to the structure of a musical composition. The parameters obtained by the tempo estimation stage (see previous section) can be used to quantize the duration information extracted from the input MIDI file and relate the intervals to the corresponding figure duration.

Observe that splitting into measures is accomplished by applying thresholds to the accumulative sum of measure durations of the input.

If the accumulation of durations equals the threshold, the measure splitter gets the measure and stores it in the database since the measure is complete. If the accumulation of durations overpasses the threshold, then a tie between bars exists and the estimated time signature at the current point is assumed to be correct, although a note is between two bars. Also, the note that overpasses the measure duration must be split into two notes: one with the proper duration to complete the previous bar, and another one with the remaining duration that will be part of the following bar, to remove the tie.

The pitch contour of the rhythmic patterns obtained by the splitting scheme are stored (Fig. 42.3). Later, patterns with more contour versions will be selected with higher probability than others by the composition system. This choice is oriented to the replication of the probabilistic model of the rhythmic patterns in the melodies composed.

Pitch Progression

The pitch contour [42.46] is more important for the generation scheme than the notes themselves since, in order to maintain the personality and the style of the reused motives, the pitch contour must be preserved [42.55]. Note that the notes are specified by the MIDI messages.

Summing up, the actual notes are not necessary if a harmony corrector is used to adapt the melody to the chord progressions. Also, the use of variations instead of the unmodified pitch patterns provides flexibility so that the patterns can be adapted to the harmonies and, additionally, the output melody can be set up to any desired key signature.

42.3.2 Melody Generator

The melody generator will use the rhythmic and pitch information and the predefined chord progression stored in the database (Fig. 42.3) to create new melodies that replicate the style or complexity of the songs previously analyzed. The melody generation can be performed by means of the concatenation of rhythmic patterns according to composition rules defined by the analyzed music samples and by previously selected musical parameters (time signature, tempo for a chosen complexity level).

First of all, the initial tonality, the time signature, the number of bars and the dataset of parameters corresponding to a certain training level can be selected beforehand. Then, other specific parameters can be selected or modified: all these are presented in Table 42.1.

Then, some rules have been considered to be automatically applied in order to guide the pattern selection, to ensure the proper harmony adaptation and to guarantee the continuity of the pitch contour. In order to define the rules, *Schellenberg*'s simplification [42.56] of *Narmour*'s Realization-Expectation model [42.57] is perfectly suitable. A specific algorithm based on music theory concepts can be used for harmony adaptation at each measure [42.58]. Figure 42.4 shows a schematic representation of the stages of the melody generation algorithm.

In the next subsections, the steps performed by the melody generator proposed will be described in detail.

Pattern Selection

The items i the dataset that fulfill the time signature requirement (database level 1) chosen by the user beforehand will be selected. Then, among these patterns, stored in the database after the analysis stage, the ones required for the creation of the rhythmic structure will be selected. For example, if the melody structure is defined as A-B-B-A, then two rhythmic patters (from the database, level 2), will be acquired.

Table 42.1 Music generation: selectable parameters

Global level
• Initial tonality
• Time signature
• Number of bars
• Style database
Phrase level
• Predefined rhythmic pattern
• Predefined harmonic pattern
Measure level
• Tonality
• Chord
• Rhythmic pattern

After linking each measure in the structure to a particular rhythm pattern, a pitch contour is selected randomly among all the pitch version for each of the motives (from database, level 3).

Note that at this stage, the pitch progression selected may not be in accordance with the harmony set up. At a later stage, a chord transposition system should adapt the pitch curve to fulfill the given harmony progression keeping the continuity of the melodic curve.

Harmony Progression

A user can design a particular chord progression. However, note that the chord progression is a very important parameter for the musical success: there are combinations of chords that do not sound well together while others do [42.59]. So, in order to guide the selection of the chord progression, a set of harmonic progressions that sound well together can be defined [42.60].

The reason why some harmonic progressions sound well while others do not is related to the listener expectation [42.57], which is linked to the cultural environment and the preference of the listener for some chord transitions rather than others. The predefined progressions considered follow the Western music theory [42.60]. These progressions are I-ii-V-I, I-vii-I-V, I-I-IV-V or I-IV-V-I, among others.

Finally, in order to adapt the patterns selected in the previous stage, a melody transposition scheme based on music theory rules must be employed. This method will be described in next section.

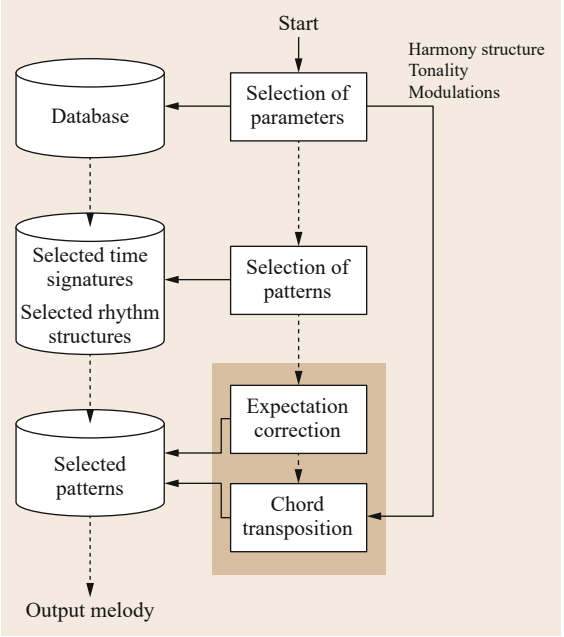


Fig. 42.4 Scheme of the melody generator

Chord Transposition

The chord transposition system must perform the proper changes to the sequence of notes to ensure that the harmony attained is the one selected and to guarantee the continuity of the melodic line according to the expectation model [42.56, 57]. This can be achieved by applying a musical harmony adaptation method based on level changing [42.58] together with additional constrains derived from the expectation model.

As the simple concatenation of patterns causes the appearance of transitions that do not sound natural, the idea to fix this issue is to generate a Narmour candidate [42.57] that properly follows the melodic line in the posterior measure. This candidate should fulfill the rules regarding the musical expectation.

The system analyses the last two notes of each measure to generate a new note. These two notes (called implication) are used to evaluate a third note (called realization), which will be the candidate note [42.56]. The following items are observed for the generation of the candidate notes [42.56]:

- Interval: A small interval [42.61] (less than a tritone) implies that the next note should follow the direction of the pitch progression. Otherwise, it would not achieve the expectation.
- Pitch jump: The pitch jump after a small interval should be similar to the previous one and in the same direction, according to the previous rule.
- Progression of the intervals:
  - If the implication interval is less than two semitones, then the third note should be back closer to the first note of the implication.
  - After a change in the direction or a large interval, the realization interval should be smaller than a tritone.

Recall that the position of the notes in the measures is key for the chord transposition stage. So, first, the chord notes, considered responsible of the harmony definition, and the nonchord notes, commonly called passing notes, are identified [42.59]. This process can be based on the analysis of the position of each note within each measure. The notes in downbeats will be considered chord notes, whilst those placed in upbeats will be considered nonchord notes.

Then, the chord transposition subsystem applies two different procedures to these two types of notes:

- Accented notes must belong to the chord
  - First chord note (or Narmour candidate [42.57]): This note is assigned to the closest pitch of the chord.
  - Secondary chord notes: Following the original pitch contour, secondary accented notes are moved to the closest pitch in the contour direction.
- Unaccented notes: The original interval between the previous note and the current note is replicated.

When the pitch and harmony adaptation processes are finished for every measure, the creation of a new melody is completed. Then, the performance on the melody by the user must be assessed.

42.4 Singing Assessment

In this section, we consider the problem of singing assessment for music learning. The descriptions will be based on the algorithm described in [42.62]. This algorithm evaluates the user's singing performance by comparing the processed audio against a reference melody.

In our case, the reference melody corresponds to the final output of the methods described in previous sections.

The main steps often required for the task at hand are illustrated in Fig. 42.5. These steps include the following global tasks: fundamental frequency (F_0) extraction and singing assessment based on F_0 alignment.

Next, we will briefly describe these steps following the approach selected, although other relevant schemes can be found in the literature [42.19, 25, 63].

42.4.1 F_0 Extraction

The Yin algorithm [42.29] has been found to be a good choice to extract the F_0 vector. This evolves from the idea of the autocorrelation method [42.64] to introduce relevant improvements. The modifications are based on the definition of the so-called cumulative mean normalized difference function $d'_t(\tau)$. This function peaks at the local period with lower deviations than the conventional autocorrelation function [42.29]. The cumulative mean normalized difference function is defined upon the squared difference function $d_t(\tau)$ given by

$$d_t(\tau) = \sum_{j=t}^{t+W} (x(j) - x(j+\tau))^2 \ , \qquad (42.1)$$

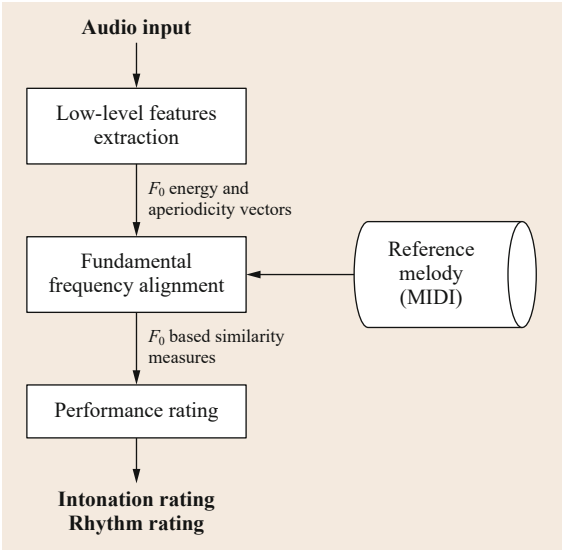


Fig. 42.5 A block diagram of the method for automatic singing assessment proposed in [42.62]

where $\tau \in [0, W)$ is an integer lag variable, W is the window size, $x(\tau)$ is the amplitude of the input signal x at time τ and t is the time index. Then, this function is normalized to give rise to the cumulative mean normalized difference function

$$d'_t(\tau) = \begin{cases} 1 & \tau = 0 \\ \dfrac{d_t(\tau)}{\frac{1}{\tau}\sum_{j=1}^{\tau} d_t(j)} & \text{otherwise} . \end{cases} \quad (42.2)$$

The Yin algorithm finds the local minimum in $d'_t(\tau)$ with the smallest τ'. Afterwards, a parabolic interpolation stage is performed using $d'_t(\tau')$, $d'_t(\tau'-1)$ and $d'_t(\tau'+1)$ to obtain an accurately estimated local minimum at τ_p. This value can be used to calculate the F_0 with $F_0 = f_s/\tau_p$, where f_s stands for the sampling rate.

The aperiodicity measure or voicing parameter is given by $ap = d'_t(\tau_p)$. This parameter is useful to identify voiced/unvoiced frames [42.62].

Note that the original Yin algorithm, implemented in Matlab, can be found in [42.65].

42.4.2 Assessment of Singing Voice

Once the F_0s of the user's performance and the reference melodies are extracted, they must be compared. A suitable method to align the functions for comparison is dynamic time warping (DTW) [42.66, 67]. This technique is useful for finding an optimal match between two sequences under certain restrictions. Note that the definition of the optimality criterion of the match strongly affects the performance of the alignment.

In [42.62], the cost matrix M for the DTW algorithm is defined as (other choices could be considered)

$$M_{ij} = \min\left\{(F_{0T}(i) - F_{0U}(j))^2, \alpha\right\} , \quad (42.3)$$

where $F_{0T}(i)$ is the F_0 of the target melody in the frame i, and $F_{0U}(j)$ represents the F_0 of the user's performance in the frame j. M_{ij} is the cost and α is a constant. Note that using this scheme, when the squared difference between F_0s becomes larger than α, the situation is considered to correspond to a spurious value and its contribution to the cost matrix is bounded.

The DTW algorithm uses the cost matrix to provide an optimal path $[i_k, j_k]$ for $k \in 1 \ldots K$, where K is the length of the path, matching the two input signals. Figure 42.6 illustrates the alignment performance.

In [42.68], a Matlab implementation of the DTW algorithm can be found.

DTW as a Similarity Measure

The path for the alignment between the user's performance and the reference melody conveys relevant information for singing evaluation. Actually, the DTW is suitable for assessing both the intonation and the rhythmic performance [42.62].

DTW to Assess Intonation. The cost matrix M provides information about the instantaneous deviation of the sung note with respect to the reference, as well as information about the overall F_0 deviation. Consequently, the total cost of the optimal alignment path found can be used as the similarity measure for intonation assessment. Then, the total intonation error (TIE) can be computed as

$$TIE = \sum_{k=1}^{K} M_{i_k j_k} , \quad (42.4)$$

where M is the cost matrix previously defined, and $[i_k, j_k]$, with $k \in 1 \ldots K$, represents each of the steps of the optimal path, K being the length of the path.

DTW to Assess Rhythm. DTW is also a powerful procedure for automatic rhythm assessment. The specific shape of the optimal path contains the necessary information about the rhythmic performance.

In the cost matrix of the DTW, a diagonal straight line represents a perfect rhythmic performance (no de-

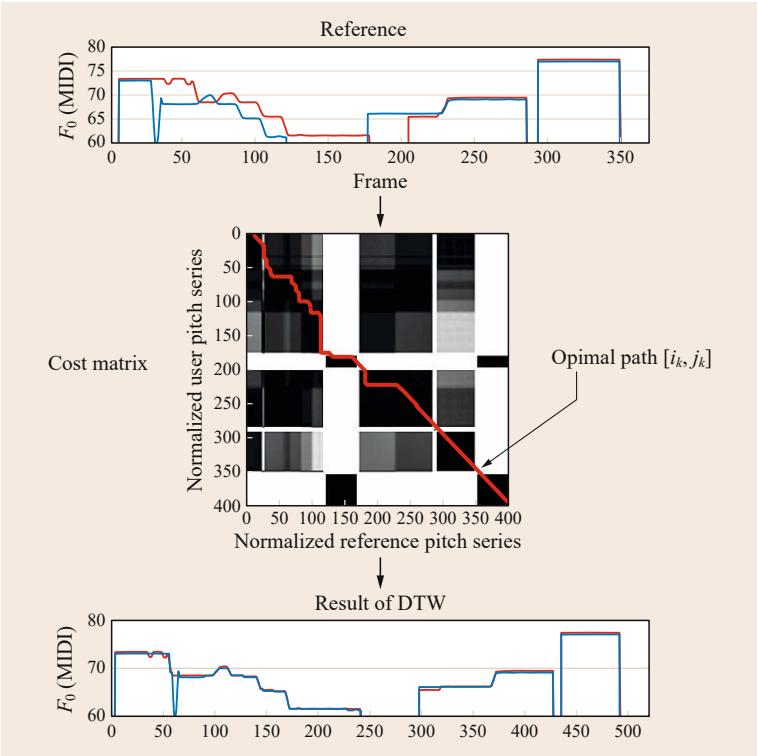


Fig. 42.6 F_0 alignment between a user's performance and the reference melody using dynamic time warping (DTW)

viation with respect to the target melody). A poor rhythmic performance would yield large deviations with respect to such a straight line. Figure 42.7 illustrates this idea.

The analysis of the deviations of the alignment path found with respect to the ideal path provides relevant rhythm assessment information. Specifically, a straight line with a slope different from the ideal one represents good rhythmic performance in a different tempo. On the other hand, the straightness of the path reveals the presence of erratic rhythmic errors. The straightness can be quantified by performing an ad hoc linear approximation to the path found, and then analyzing the error.

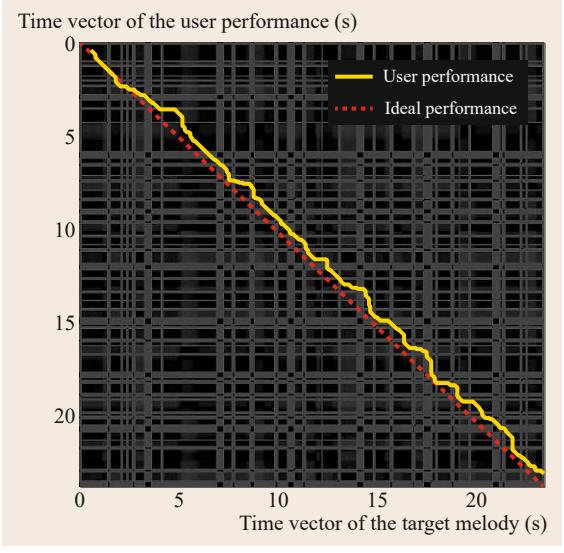


Fig. 42.7 Sample of the usage of DTW with F_0 signal for rhythm assessment. Rhythmically stable user's performance (*solid line*) and ideal rhythm performance (*dotted line*) ►

42.5 Summary

In this chapter, a complete approach to the development of computational tools for singing learning has been proposed.

Two main subsystems are required for the singing learning purpose: a module for the automatic generation of singing exercises with selectable complexity levels,

and a module for the automatic assessment of the user's singing performance.

A complete melody generator scheme, including the required analysis stages, has been presented. The generator described is able to automatically generate new melodies adapted to a certain music level selected beforehand.

An approach for the automatic assessment of singing voice has also been described. The method selected compares the F_0 of the user's performance against the reference F_0 of an automatically generated melody. The scheme provides an evaluation of both intonation and rhythm.

Acknowledgments. This work has been funded by Ministerio de Economía y Competitividad of the Spanish Government under Project No. TIN2016-75866-C3-2-R. This work has been done at Universidad de Málaga, Campus de Excelencia Internacional Andalucía Tech.

References

42.1 M.P. Ryynänen, A.P. Klapuri: Automatic transcription of melody, bass line, and chords in polyphonic music, Comput. Music J. **32**(3), 72–86 (2008)

42.2 J. Serrá, E. Gómez, P. Herrera: Audio cover song identification and similiraty: Background, approaches, evaluation, and beyond. In: *Advances in Music Information Retrieval*, Vol. 274, ed. by Z.W. Ras, A.A. Wieczorkowska (Springer, Berlin, Heidelberg 2010) pp. 307–332

42.3 S. Koelsch, W.A. Siebel: Towards a neural basis of music perception, Proc. TRENDS Cogn. Sci. **9**(12), 578–584 (2005)

42.4 R.F. Goldman: Ionisation; Density, 21.5; Integrales; Octandre; Hyperprism; Poeme Electronique, Musical Q. **47**(1), 133–134 (1961)

42.5 G. Nierhaus: *Algorithmic Composition: Paradigms of Automated Music Generation*, Vol. 34 (Springer, Wien 2010)

42.6 C. Uhle, J. Herre: Estimation of tempo, micro time and time signature from percussive music. In: *Proc. Int. Conf. Digital Audio Effects (DAFx)* (2003)

42.7 F. Gouyon, P. Herrera, P. Cano: Pulse-dependent analyses of percussive music, Proc. ICASSP **4**, 396–401 (2002)

42.8 S. Tojo, K. Hirata: Structural similarity based on time-span tree. In: *Proc. 9th Int. Symp. Comput. Music Model. Retriev. (CMMR)* (2012) pp. 645–660

42.9 M. Müller, D.P.W. Ellis, A. Klapuri, G. Richard: Signal processing for music analysis, IEEE J. Sel. Top. Signal Process. **5**(6), 1088–1110 (2011)

42.10 A. Van Der Merwe, W. Schulze: Music generation with Markov models, IEEE Multimed. **18**(3), 78–85 (2011)

42.11 M. Pearce, G. Wiggins: Towards a framework for the evaluation of machine compositions. In: *Proc. AISB'01 Symp. AI Creat. Arts Sci* (2001) pp. 22–32

42.12 D. Conklin: Music generation from statistical models. In: *Proc. Symp. Artif. Intell. Creat. Arts Sci. (AISB)* (2003) pp. 30–35

42.13 E.R. Miranda, J.A. Biles: *Evolutionary Computer Music* (Springer, London 2007)

42.14 D. Cope: Computer modeling of musical intelligence in EMI, Comput. Music J. **16**(2), 69–83 (1992)

42.15 D. Cope: *Computer Models of Musical Creativity* (MIT Press, Cambridge 2005)

42.16 M. Delgado, W. Fajardo, M. Molina-Solana: Inmamusys: Intelligent multiagent music system, Expert Syst. Appl. **36**(3), 4574–4580 (2009)

42.17 D.M. Howard, G. Welch, J. Brereton, E. Himonides, M. Decosta, J. Williams, A. Howard: WinSingad: A real-time display for the singing studio, Logop. Phoniatr. Vocology **29**(3), 135–144 (2004)

42.18 Barcelona Music and Audio Technologies: *SKORE Performance Rating*, http://skore.bmat.me (2008)

42.19 O. Mayor, J. Bonada, A. Loscos: The singing tutor: Expression categorization and segmentation of the singing voice. In: *Proc. AES 121st Convention* (2006)

42.20 D. Rossiter, D.M. Howard: ALBERT: A real-time visual feedback computer tool for professional vocal development, J. Voice Off. J. Voice Found. **10**(4), 321–336 (1996)

42.21 Sony Computer Entertainment Europe: *Singstar* (SCEE London Studios 2004)

42.22 T. Nakano, M. Goto, Y. Hiraga: An automatic singing skill evaluation method for unknown melodies using pitch interval accuracy and vibrato features. In: *Proc. INTERSPEECH (ICSLP)* (2006) pp. 1706–1709

42.23 J. Callaghan, P. Wilson: *How to Sing and See: Singing Pedagogy in the Digital Era* (Cantare Systems, Surry Hills 2004)

42.24 D. Hoppe, M. Sadakata, P. Desain: Development of real-time visual feedback assistance in singing training: A review, J. Comput. Assist. Learn. **22**(4), 308–316 (2006)

42.25 S. Grollmisch, E. Cano Cerón, C. Dittmar: Songs2see: Learn to play by playing. In: *41st Int. Audio Eng. Soc. Conf. (AES)* (2011)

42.26 Z. Jin, J. Jia, Y. Liu, Y. Wang, L. Cai: An automatic grading method for singing evaluation, Rec. Adv. Comput. Sci. Inf. Eng. **5**, 691–696 (2012)

42.27 C. Dittmar, E. Cano, J. Abeßer, S. Grollmisch: Music information retrieval meets music education, Multimod. Music Process. **3**, 95–120 (2012)

42.28 E. Gómez, A. Klapuri, B. Meudic: Melody description and extraction in the context of music content processing, J. New Music Res. **32**(1), 23–40 (2003)

42.29 A. De Cheveigné, H. Kawahara: YIN, a fundamental frequency estimator for speech and music, J. Acoust. Soc. Am. **111**(4), 1917 (2002)

42.30 T. Viitaniemi, A. Klapuri, A. Eronen: A probabilistic model for the transcription of single-voice

melodies. In: *Proc. 2003 Finn. Signal Process. Symp. FINSIG'03* (2003) pp. 59–63
42.31 M. Ryynänen, A. Klapuri: Modelling of Note Events for Singing Transcription. In: *Proc. ISCA Tutor. Res. Workshop Stat. Percept. Audio Process. (SAPA)* (2004)
42.32 G.E. Poliner, D.P.W. Ellis, A.F. Ehmann, E. Gómez, S. Streich, B. Ong: Melody transcription from music audio: Approaches and evaluation, IEEE Trans. Audio Speech Lang. Process. **15**(4), 1247–1256 (2007)
42.33 E. Molina: *Automatic Scoring of Signing Voice Based on Melodic Similarity Measures* (Universitat Pompeu Fabra, Barcelona 2012)
42.34 R.J. McNab, L.A. Smith, I.H. Witten: Signal processing for melody transcription, Proc. 19th Australas. Comput. Sci. Conf. **18**(4), 301–307 (1996)
42.35 M. Ryynänen: Singing transcription. In: *Signal Processing Methods for Music Transcription*, ed. by A. Klapuri, M. Davy (Springer Science/Business Media LLC, New York 2006) pp. 361–390
42.36 J.J. Mestres, J.B. Sanjaume, M. De Boer, A.L. Mira: *Audio Recording Analysis and Rating*, US Patent 8158871 (2012)
42.37 G. Haus, E. Pollastri: An audio front end for query-by-humming systems. In: *Proc. 2nd Int. Symp. Music Inf. Retriev. (ISMIR)* (2001) pp. 65–72
42.38 W. Krige, T. Herbst, T. Niesler: Explicit transition modelling for automatic singing transcription, J. New Music Res. **37**(4), 311–324 (2008)
42.39 E. Molina: Hacer música... para aprender a componer, Eufonia, Didáct. Músic. **51**, 53–64 (2011)
42.40 M.K. Shan, S.C. Chiu: Algorithmic compositions based on discovered musical patterns, Multimed. Tools Appl. **46**(1), 1–23 (2010)
42.41 P.J. Ponce de León: Statistical description models for melody analysis and characterization. In: *Proc. Int. Comput. Music Conf.*, ed. by J.M. Iñesta (2004) pp. 149–156
42.42 Association MIDI Manufacturers: *The Complete MIDI 1.0 Detailed Specification* (The MIDI Manufacturers Association, Los Angeles 1996)
42.43 R.S. Brindle: *Musical Composition* (Oxford Univ. Press, Oxford 1986)
42.44 F. Lerdahl, R. Jackendoff: *A Generative Theory of Tonal Music* (MIT Press, Cambridge 1983)
42.45 W.T. Fitch, A.J. Rosenfeld: Perception and production of syncopated rhythms, Music Percept. **25**, 43–58 (2007)
42.46 W. Appel: *Harvard Dictionary of Music*, 2nd edn. (The Belknap Press of Harvard Univ., Cambridge, London 2000)
42.47 K. Seyerlehner, G. Widmer, D. Schnitzer: From rhythm patterns to perceived tempo. In: *Int. Soc. Music Inf. Retriev. (ISMIR)* (2007) pp. 519–524
42.48 M.F. McKinney, D. Moelants: *Ambiguity in Tempo Perception: What Draws Listeners to Different Metrical Levels?* (Univ. of California Press, Oakland 2006) pp. 155–166
42.49 M. Gainza, D. Barry, E. Coyle: Automatic bar line segmentation. In: *123rd Convent. Audio Eng. Soc. Convent. Paper* (2007)
42.50 M. Gainza, E. Coyle: Time signature detection by using a multi resolution audio similarity matrix. In: *122nd Convent. Audio Eng. Soc. Convent. Paper* (2007)
42.51 J. Foote, M. Cooper: Visualizing musical structure and rhythm via self-similarity. In: *Proc. 2001 Int. Comput. Music Conf.* (2001) pp. 419–422
42.52 J.R. Quinlan: *C4.5: Programs for Machine Learning* (Morgan Kaufmann, San Francisco 1993)
42.53 J. Platt: *Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines* (Microsoft Research, Redmond 1998)
42.54 M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten: The WEKA data mining software: An update, SIGKDD Explor. **11**(1), 10–18 (2009)
42.55 W.J. Downling, D.S. Fujitani: Contour, interval and pitch recognition in memory for melodies, J. Acoust. Soc. Am. **49**, 524–531 (1971)
42.56 E. Schellenberg: Simplifying the implication-realization model of musical expectancy, Music Percept. **14**(3), 295–318 (1997)
42.57 E. Narmour: *The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model* (Univ. of Chicago Press, Chicago, London 1992)
42.58 D. Roca, E. Molina (Eds.): *Vademecum Musical* (Enclave Creativa, Madrid 2006)
42.59 B. Benward: *Music: In Theory and Practice*, Vol. 1, 7th edn. (McGraw-Hill, New York 2003)
42.60 R.W. Ottman: *Elementary Harmony: Theory and Practice*, 5th edn. (Prentice Hall, Englewood Cliffs 1989)
42.61 A.E. Yilmaz, Z. Telatar: Note-against-note two-voice counterpoint by means of fuzzy logic, Knowl.-Based Syst. **23**(3), 256–266 (2010)
42.62 E. Molina, I. Barbancho, E. Gomez, A.M. Barbancho, L.J. Tardon: Fundamental frequency alignment vs. note-based melodic similarity for singing voice assessment. In: *IEEE Int. Conf. on Acoust. Speech Signal Process. (ICASSP)* (2013) pp. 744–748
42.63 J. Wapnick, E. Ekholm: Expert consensus in solo voice performance evaluation, J. Voice **11**(4), 429–436 (1997)
42.64 L.R. Rabiner, R.W. Schafer: *Digital Processing of Speech Signals*, Prentice-Hall Series in Signal Processing No. 7, Vol. 25 (Prentice Hall, Englewood Cliffs 1978) p. 290
42.65 A. De Cheveigné: *Matlab Implementation of YIN Algorithm*, http://audition.ens.fr/adc/sw/yin.zip (2012)
42.66 H. Sakoe: Dynamic programming algorithm optimization for spoken word recognition, IEEE Trans. Acoust. Speech Signal Process. **26**, 43–49 (1978)
42.67 C.A. Ratanamahatana, E. Keogh: Everything you know about dynamic time warping is wrong. In: *3rd Workshop Min. Tempor. Seq. Data, 10th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. (KDD-2004)* (2004)
42.68 D. Ellis: *Dynamic Time Warp (DTW) in Matlab*, http://labrosa.ee.columbia.edu/matlab/dtw (2003)

43. Computational Ethnomusicology: A Study of Flamenco and Arab-Andalusian Vocal Music

Nadine Kroher, Emilia Gómez, Amin Chaachoo, Mohamed Sordo, José-Miguel Díaz-Báñez, Francisco Gómez, Joaquin Mora

In this chapter we approach flamenco and Arab-Andalusian vocal music through the analysis of two representative pieces. We apply a hybrid methodology consisting of audio-signal processing to describe and contrast their melodic characteristics followed by musicological analysis. The use of such computational analysis tools complements a musicological-historical study with the aim of supporting the discovery and understanding of the specific characteristics of these musical traditions, their similarities and differences, while offering solutions to more general music information retrieval (MIR) research challenges.

43.1 Motivation

Over the last decades, the music information retrieval (MIR) community has earnestly striven towards the development of technologies for the automatic description of signals from musical phenomena such as melody, harmony, dynamics, or instrumentation. These technologies have been exploited in the context of numerous applications such as music identification, retrieval and recommendation. However, while the number of applications has been plentiful and covering a broad spectrum of needs, in the majority of cases they are limited to mainstream popular music from the so-called Western tradition [43.1].

Fortunately, a genuine interest in applying those techniques to varied repertoire, including traditional, folk and ethnic music [43.2] has awakened recently. The task of adapting existing technologies is not clear-cut and many challenges are posed because of the particular musical features of the given tradition, which may be markedly different from the Western one.

The main goal of this chapter is to illustrate the potential of music content description techniques for the analysis of traditional music. For this study, we focused on two well-established music traditions, flamenco and Arab-Andalusian music, which are studied through a computer-assisted analysis of a selected set of pieces in audio format. These traditions are an important part of the musical heritage of Andalusia and the North of Africa. Given the main role of the singing voice in those traditions, our study pays close attention to its melodic aspects.

Through the analysis of two representative pieces, a flamenco *martinete* and an Arab-Andalusian *inshád*, we illustrate some commonalities and divergences in these traditions and present evidence on how existing technologies allow us to formalize expert knowledge and complement traditional analysis methodologies by discovering relationships that might otherwise have been unnoticed. Moreover, this study implements a method for the computer-assisted analysis of traditional music, which can provide means to formulate, test out, and confirm research hypotheses that are not always evident from a perceptual and musicological analysis.

43.1.1 Computational Music Analysis

Regarding the issue of how to computationally model and study the musical phenomenon, two different trends shape the existing research in music information re-

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_43

trieval (MIR) [43.3], namely, the symbolic approach and the signal description approach. On one side, the symbolic description disregards any audio signal analysis in order to concentrate just on abstract representations of musical concepts, such as notes, durations, beats, rhythm, melody patterns, harmony, or structural relationships. This type of analyses use scores, Musical Instrument Digital Interface (MIDI) files or other symbolic representations of music as input data and, in most of the cases, assumes a listener who has been educated on some formal music tradition (by learning to play an instrument and by being exposed to this type of music). On the other side, the audio-content approach, which is the path pursued here, uses the audio signal as the basis for the description and processing of music.

Since the beginning of MIR, most models and technologies have been developed for mainstream popular music in the so-called *Western* tradition. In parallel to that extensive and rich body of research, over the last few years an increasing interest in applying available techniques to the study of traditional, folk or ethnic music has developed. As mentioned in [43.4], the name of *computational ethnomusicology* could be at least as old as 34 years [43.5]. In [43.5], two mathematicians and an engineer provided an interesting discussion about the role of computers in five areas of research in ethnomusicology: collection, administration, notation, selection and systematization, and scientific treatment of music data. They state that computers might be essential tools for managing three groups of data: melodies (the study involves 10 000 melodies), recordings and social context. Although these principles and usages of computers are still valid today, computational models have evolved to the extent that they can be employed to efficiently collect field data or simulate complex social interactions.

The term computational ethnomusicology was recently redefined by *Tzanetakis* et al. [43.6] as the *design, development and usage of computer tools that have the potential to assist in ethnomusicological research*. Under this view there is a main discipline, ethnomusicology, that takes advantage of advances made in supportive disciplines such as computer science, music cognition or cultural studies, and applies them – after the appropriate tuning – to its particular tasks. Even though this could be the most accepted and practiced view, *Gómez* et al. [43.4] argue that this is a restrictive definition of a discipline as a tool provider. According to them,

> *computer models can be theories or hypotheses (not just tools as a spreadsheet or a statistical package can be) about processes and problems studied by traditional ethnomusicologists. When we approach computational ethnomusicology this way, we adopt a new mental framework that helps to restructure problems and perceive the relationships between their constitutive elements under a different perspective.*

With the present study we aspire to help overcome such procedural views of computation and strive for a broad and fertile conception of the discipline.

State-of-the-art techniques for the description of music audio signals allow the automatic computation of features related to different musical facets such as melody, harmony, rhythm, and instrumentation [43.7]. These features have been used to compute similarity distances and classify musical pieces according to, e.g., artist, genre, or mood. In this way, current systems can, for instance, locate different versions of the same song with a high accuracy rate [43.8]. Although computational techniques have been proven to be of great interest when applied to different musical repertoires, it is apparent that we need to develop culture-specific methods to understand and model different music repertoires. Existing literature has addressed repertoires from a variety of music traditions, e.g., Turkish makam music [43.9], Hungarian *siratok*, Torah cantillation, 10th century St Gallen plainchant and Koran recitation [43.10], music from Central Africa [43.11] and Indian classical music [43.12], as well as the automatic detection and classification of non-Western music genres [43.13].

For the case of flamenco, first approaches to computational analysis and description have been mainly carried out by the Computational Analysis of Flamenco Music (COFLA) [43.14] research group, which has focused on rhythmic properties [43.15], melodic similarity [43.16, 17], and pattern matching [43.18, 19]. *Wu* [43.20] approached computational modeling of the complex rhythmical structures of *bulerías*. Considering the extensive use of stylistic ornamentation in the singing interpretation in the music traditions under study, adjustments of existing technologies are necessary. Resulting computational tools can support and complement traditional musicological studies and thus provide a novel approach that permits advancements in rigorous understanding, diffusion and preservation of these musical traditions.

43.2 Background

43.2.1 Flamenco

Flamenco is an oral tradition with roots as diverse as the cultural influences of its area of origin, Andalusia (a region in southern Spain). In the course of history, immigrants with a great variety of cultural backgrounds arrived at its harbors and settled in the surrounding cities. Over the centuries, the area and, of course, its music have been influenced by the ancient Tartessian culture as well as Phoenician and Roman colonizations, notwithstanding later settlements by the Visigoths, the Arabs, the Jews, the Christians, and to a large extent gypsies, who decisively contributed to shape its form as we know it today. The reader is referred to the books of *Blas-Vega* and *Ríos-Ruiz* [43.21], *Navarro* and *Ropero* [43.22], and *Gamboa* [43.23] for a comprehensive study of styles, musical forms, and history of flamenco.

Flamenco music germinated from and was nourished mainly by the singing tradition [43.23]. Accordingly, the role of the singer soon became dominant and fundamental. In the flamenco jargon, singing is called *cante*, and songs are termed *cantes*. The flamenco singing voice can be characterized as unstable in pitch, timbre and dynamics [43.24]. Furthermore, the typical voice quality is described as matte, containing few high-frequency harmonics, and usually lacking the singer's formant. Melodic movements are composed of conjunct degrees, a high degree of complex, microtonal ornamentation and melisma [43.16]. Flamenco singing is usually accompanied by the guitar. Other forms of accompaniment may include clapping, stamping of feet, or percussion instruments. The origin and evolution of the more than fifty different flamenco styles (*palos*) [43.21] and variants have been studied from different disciplines, including ethnomusicology, literature and anthropology, and different theories have been proposed [43.21, 23, 25]. The most widespread [43.25] claims a period of isolation in which Andalusian gypsies kept their musical tradition from the outside world and performed only at intimate family reunions. However, this theory lacks reliable documentation and recent research has suggested that the growing popularity of singing performances put on at theaters in Andalusian cities could be the origin of this music as we know it today.

43.2.2 Arab-Andalusian Music

The Arab-Andalusian music (or simply, Andalusian music) is a musical tradition that can be traced back to the 12th century in Al-Andalus, to the Muslims and Christians living in the Moorish Spain [43.26]. Andalusian music is the result of many influences, including Middle-East Arabic classical music, the Hispanic music traditions of the Iberian peninsula, and other classical traditions such as the Gregorian and Byzantine. The Andalusian tradition is maintained in quite a few north African regions [43.27, 28], mainly in Morocco, Algeria, and Tunisia.

Andalusian music is organized around the concept of *nawba* [43.27, 28], a collection of melodies belonging to the same *tab'* (roughly, a melodic mode). The melodic structure is modal in essence, a factor that is critical for an appropriate description of the pitch relationships as music unfolds over time. Beyond the definition of scale or a set of pitches, a particular *tab'* [43.28] is linked to specific emotions or states of mind and is consequently associated with certain social occasions and circumstances.

Apart from the clear dominance of the singing voice, the *oud* is the most prominent instrument in Andalusian music. The original *oud* is a lute-related instrument, whose four strings are described by ancient music treatises as bringers of temperament (mood), senses (color, touch), spiritual or energetic states, and other metaphorical human constructs [43.28]. These relationships are actually used as methodological tools in current music training, and they presumably play an important role during performance and listening engagement. Other instruments in the Andalusian music tradition include *rbab, derbuga, qanun, tar, kamanya, and nay*. Despite the close connection between the musical instruments and the emotions or spirituality, the singing voice represents the key element. Andalusian music is furthermore characterized by the use of sung poems (*san'as*) [43.29]. The instruments are considered as accompaniments or bridges between sung verses or hemistichs. The poems (*muwashshahas* or *zĂl'jels*) are taken from Arabic classical poetry. However, their meter and rhymes have been evolved and adapted to the local sociocultural background [43.26].

As in the case of flamenco, the Andalusian music has been preserved and kept alive as an oral tradition. The modern study of the Andalusian music theory started in the colonial period (20th century), but most of these studies did not consider the plurality and the diverse influences of Andalusian music.

In this comparative analysis of flamenco and Arab-Andalusian music we consider both as independent music traditions. Even though both share a geographic

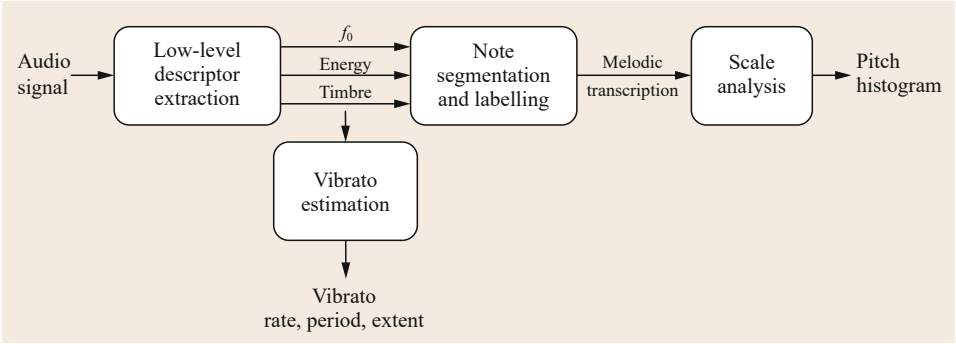


Fig. 43.1 Block diagram for melodic characterization

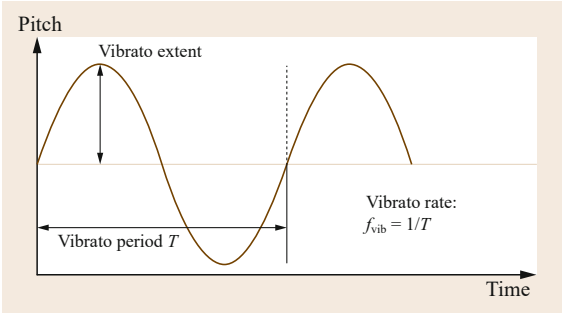


Fig. 43.2 Vibrato rate, period and extent

location of origin, they posses individual characteristics regarding aesthetics, musical structure and communicative intention. Nevertheless, some common variables can be easily found by simply listening to examples of both styles: melismatic ornamentation, importance of vocal vibrato, and even similar melodic patterns.

43.2.3 Music Content Description

In a first step towards retrieving musical content from a piece of audio, physical properties are extracted directly or in a mathematically derived way from the signal without incorporating further high-level musical and contextual knowledge. In order to obtain a higher level of abstraction and prove comprehensive content descriptors (i. e., instrumentation, genre, or melodic transcription), such low-level descriptors need to be combined with content-based generalizations about the data. For the analysis of the pieces under study we mainly focus on properties related to the vocal melody and consider state-of-the-art methods according to the steps shown in Fig. 43.1.

Low-Level Descriptors

The most relevant low-level descriptors required to obtain melodic descriptors are fundamental frequency, energy and timbre-related spectral features [43.30]. The fundamental frequency (f_0) is closely related to its perceptual equivalent, the pitch, and in most scenarios the former corresponds to the latter. Changes in timbre (spectral content) and energy are correlated to note onsets and offsets and therefore give complementary information when segmenting the pitch contour into single note events. A great variety of algorithms for extracting the fundamental frequency from monophonic audio signal have been proposed in the past. For a detailed overview the reader is referred to [43.31]. In the present study we estimated the fundamental frequency by computing the spectrum of the signal and analyzing the magnitude correlation in the frequency domain [43.30], thus exploiting the regular distance between fundamental frequency and overtones.

Vibrato Descriptors

Vocal vibrato can be characterized as an oscillation of the fundamental frequency within a range of 4–8 Hz and an extent of less than one semitone (Fig. 43.2). We estimate the local vibrato rate and extent from the fundamental frequency [43.17]. The contour is treated as a time series and high-pass filtered in order to obtain a zero-centered signal in which only fast pitch fluctuation is preserved. If vibrato is present, the spectrum of the preprocessed contour shows a peak in the considered frequency range. The instantaneous vibrato rate corresponds to the peak frequency and the vibrato extent can be estimated as the magnitude of the pitch fluctuation in the current frame. We furthermore extract the vibrato amount as the percentage of voiced frames in which vibrato is detected.

Melodic Contour Analysis

In order to provide a symbolic representation for melodic analysis, we transform the fundamental frequency contour into a sequence of single notes, described by their energy, pitch and duration, as explained in *Gómez* and *Bonada* [43.30]. After estimating a global tuning frequency, the fundamental frequency contour is

segmented into notes by maximizing a likelihood function along the analyzed excerpt. In this step, the typical pitch range of a singer as well as limitations for note durations are considered when finding the optimal path for the pitch progression. In a later step, successive short notes are consolidated if pitch, energy and timbre descriptors indicate a single note instead of consecutive onsets. From the resulting symbolic note representation we can extract statistical descriptors of the melodic content, as, i. e., pitch range, pitch fluctuation or average note duration.

Scale Analysis

Pitch histograms have been widely used to characterize the pitch content distribution of music in different traditions [43.11, 32]. In order to analyze the tonality of the considered pieces, a pitch histogram is computed from the note transcription, displaying the occurrence of distinct pitches, weighted by their duration. Since tonal centers and pitches belonging to the underlying scale tend to appear more frequently, such pitch histograms are used to extract and analyze the tonality of a piece of music.

43.3 Case Study

43.3.1 The Flamenco Martinete

The *martinete* is a traditional flamenco singing style without guitar accompaniment, characterized by a distinct melodic skeleton. It forms part of the subgenre of the *tonás* (from Spanish *tonadas* – *singable* or *melodic fragment*), a group of a cappella styles, which are thought be the origin of a number of flamenco styles [43.25]. According to [43.33], the *martinete* has its origins in Arab chants that were heard and adapted by gypsies. Flamenco experts distinguish between a number of substyles, corresponding to their creator or geographic origin (i. e., *martinete* from *Triana, Jerez*, and *Los Puertos*), and whether lyrics are repeated (*natural* without repetition and *redoblao* with repetition). Similar styles are *debla*, which shares musical elements with *martinete redoblao*, *carcelera* and *saeta*, which in some cases are interpreted by using *martinete's* distinct melodic progression. In 1952 *Antonio el Bailarín* was the first to dance to a *martinete*, using the characteristic *siguiriya* rhythm. Furthermore, in order to pay tribute to the assertion that the style has been sung by workers in blacksmith's shops, *martinetes* are sometimes accompanied by mallet strokes on metal surfaces.

General Characteristics

Characterized by a slow tempo, solemn performance, free rhythmic interpretation and a large amount of melismatic ornamentation, songs belonging to the style have a common melodic skeleton based on the major scale. While the major mode is dominant throughout, the third degree may be occasionally lowered by a semitone, converting the scale into minor mode. The four eight-syllable verses are arranged in an A-B-A-C form, where the first and the third sections share a similar melodic contour. These two sections are separated into two segments by a rest. The first section contains a characteristic melodic sequence that clearly identifies the style: a jump from the subtonic or the dominant to the tonic followed by a raise to the subdominant by conjunct degrees.

Martinete by Singer Tomás Pavón

In the present study we focused on the example of a *martinete* performance by the renowned singer Tomás Pavón, which belongs to the substyle of *martinetes* by *Juan Pelao*, also referred to as *origin martinete* (*martinete de origen*).

Figure 43.3 shows a manually annotated representation of the melodic contour of the piece, where verses are marked by numbers. Figure 43.4 provides a manual

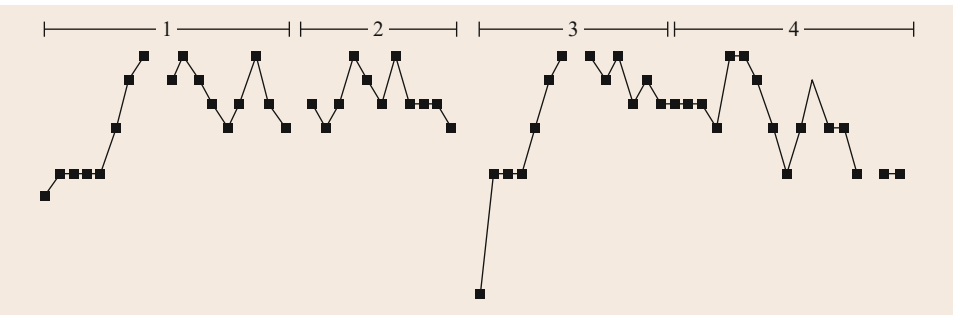


Fig. 43.3 Symbolic representation of the melodic contour of a *martinete* by singer Tomás Pavón

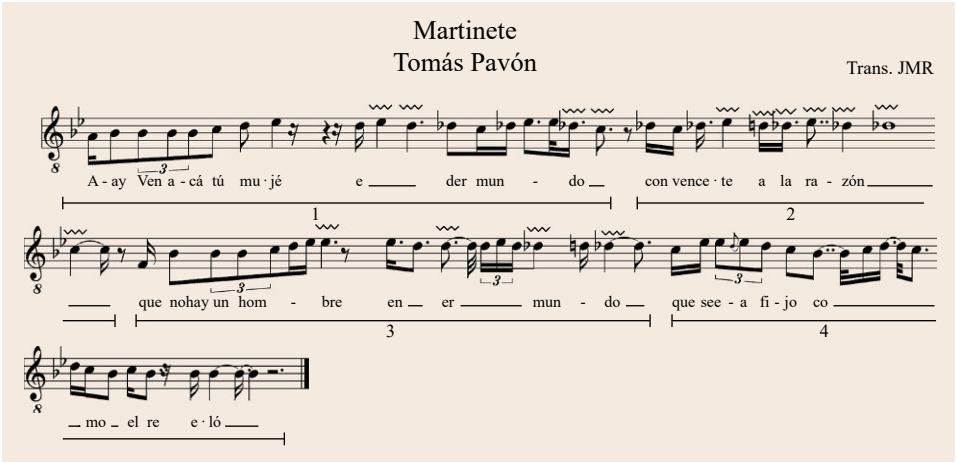


Fig. 43.4 Transcription in Western Music notation of a *martinete* by singer Tomás Pavón

Table 43.1 Relevant pitch values for the four verses of the *martinete* by singer Tomás Pavón

Verse	First note	Recitative note	Last note	Direction
1	A4	Eb5	C5	Rising
2	D5	Eb5	C5	Horizontal
3	F4	Eb5	Db5	Rising
4	C5	C5	Bb4	Falling

transcription of the piece in Western notation and Table 43.1 provides some features of the different verses of the piece. The melodic contour is based on the B-flat major scale and shows the style-specific note sequences described in the previous subsection. In the first half of the first verse, the tonic is reached from the subtonic, followed by three subsequent degree steps to the subdominant. In the first half of the third verse, the melodic contour is repeated, whereas the tonic is reached from the dominant. The recitative note in each movement is prolonged and subject to strong melismatic ornamentation. Furthermore, note that the first and second verses are separated by silences, while the third and fourth are fused together.

Tomás Pavón's interpretation is rich in melismatic ornamentations, which cause a stronger prolongation of some syllables than in most other performances. Trills and double relishes become denser with increasing scale degree, while chromatic ornamentations are bound to the third scale degree. Frequently, such ornamentations take on rhythmic patterns of triplets and quintuplets. The melisma are performed rapidly and with a precise breath management. While the tuning varies during the performance, the transcription in Western music notation (Fig. 43.4) refers to note values quantized to the equal-tempered scale. Since the song is performed in free rhythm, a separation into bars is furthermore omitted, while traditional symbolism is used to describe note durations. In order to simplify reading, some ornamentation details have been taken out. Figure 43.5 shows these details for a small excerpt.

43.3.2 The Arab-Andalusian Inshád

Arab-Andalusian music is strongly modal and, similarly to Gregorian chant, based on particular melodic cells known as *centones*. The use of *centones* is quite characteristic of Christian music in the Middle Ages (in the Byzantine tradition they are sometimes referred to as *schemata*). Rather than creating new *centones*, composers of those times made creative use of the existing ones. Furthermore a mode is defined by its pitch range, a tonal center around which the melody gravitates, important scale degrees that provide the mode's characteristic flavor and a set of *centones* on which melodic phrases end.

The *inshád* is a sung poem of two verses. Forming part of the Arab-Andalusian musical heritage, the melodic progression is based on set motives that are

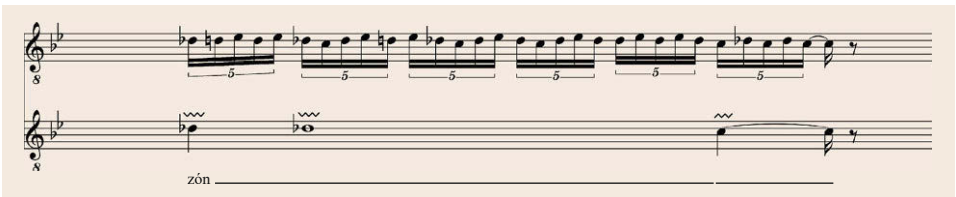


Fig. 43.5 Ornamentation details of an excerpt of a *martinete* by singer Tomás Pavón

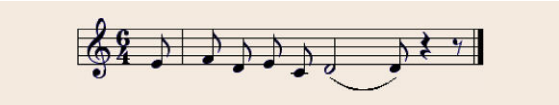


Fig. 43.6 Example melodic phrase of the *San'a "Sallu ya 'ibad"* of *Basit Raml al-Maya*

spontaneously ornamented [43.26]. Below we describe the main characteristics of the analyzed piece.

General Characteristics

The *inshád* has a series of particular musical characteristics [43.26]. Its melodies are integrated in the modal system of Arab-Andalusian music. The most common intervals are a (major or minor) second, followed by minor third and major third. Most of the melodic phrases end with a long note. The latter is a common characteristic found in different forms in Arab-Andalusian music corpora [43.34]. Figure 43.6 shows a melodic phrase of the *San'a "Sallu ya 'ibad"* of *Basit Raml al-Maya* (*Basit* is the name of the rhythmic structure and *Raml al-Maya* is the name of the *nawba*).

Even though the meter of the *inshád* is free, this freedom is not absolute, but rather rational. Indeed, the notes of the *inshád* are not tied to a rhythmic pulse or an exact execution time. Yet they have to be subject to a temporal hierarchy of notes and complete phrases, rationally approximate but intuitively guided: The importance of a note in a phrase implies an extension of its duration compared to the other notes within the phrase. Furthermore, the duration of the rests between (blocks of) melodic phrases depends on the global cadence of the *inshád*. This cadence depends on the general ambiance of the concert as well as the typical tendencies of orchestra and singer.

Melodic Improvisation in the Inshád

The melodic improvisation in Arab-Andalusian music can be of two types: ornamentation of existing melodies or creation of new melodies, such as *mawwals* and *istijbárs*. In a performance of an *inshád*, the improvisation can only occur as an ornamentation of the its base melody.

Ornamentation is a key feature in Arab-Andalusian music, given the limitation on a set of fixed melodies. Hence, it represents an important tool for musicians to express their individual interpretation and performance character. As a result, individual performances of an *inshád* vary regarding the amount and complexity of ornamentation. There are however some constraints: the musical mode, an intuitive phrase duration, the aesthetic traits and the emotional context of the performance need to be preserved (e.g., an ornamentation related to a sad emotion should be avoided in a wedding performance). Even though the melodic content is limited to the mode, singers can include notes from so-called *neighbor modes*: as explained in [43.26], two musical modes are considered to be neighbors if their scale sequences differ by an alteration of a single note.

Al-Sika Mode

The *Al-Sika* mode with its *centones* is shown in Fig. 43.7. In contrast to Western scales, the tonal center does not correspond to the first scale degree, but to the note E, the third degree. The first (C) and fifth degree (G) are fundamental notes with higher melodic importance. Furthermore the scale is subdivided into four units of consecutive notes with a common intervallic structure (*al-Dayl*).

The ethos (the character or emotion) associated with the *Al-Sika* mode is a complaint of a resigned person

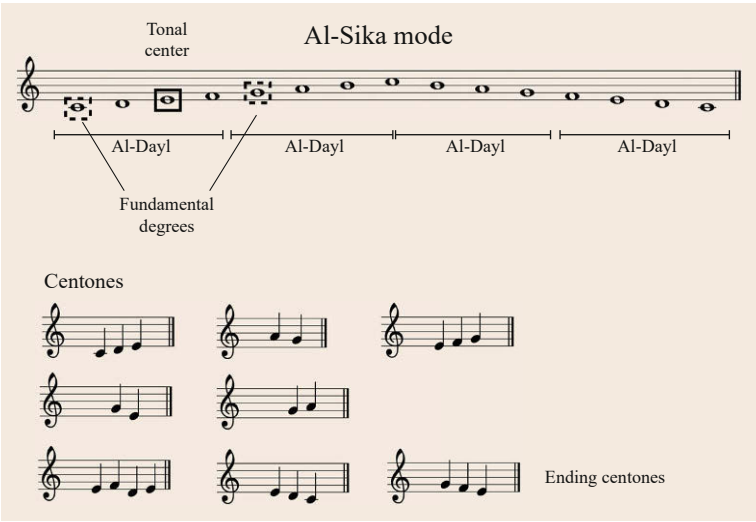


Fig. 43.7 Summary of the main characteristics of the Al-Sika mode

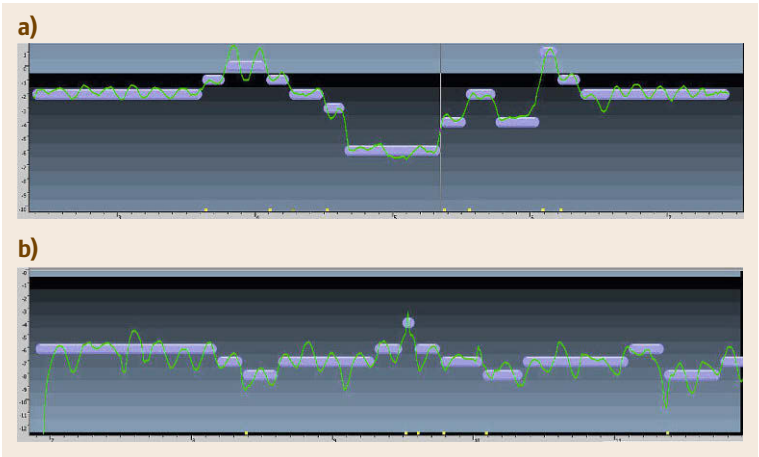


Fig. 43.8a,b Fundamental frequency contour and note transcription of a phrase; **(a)** inshád; **(b)** martinete

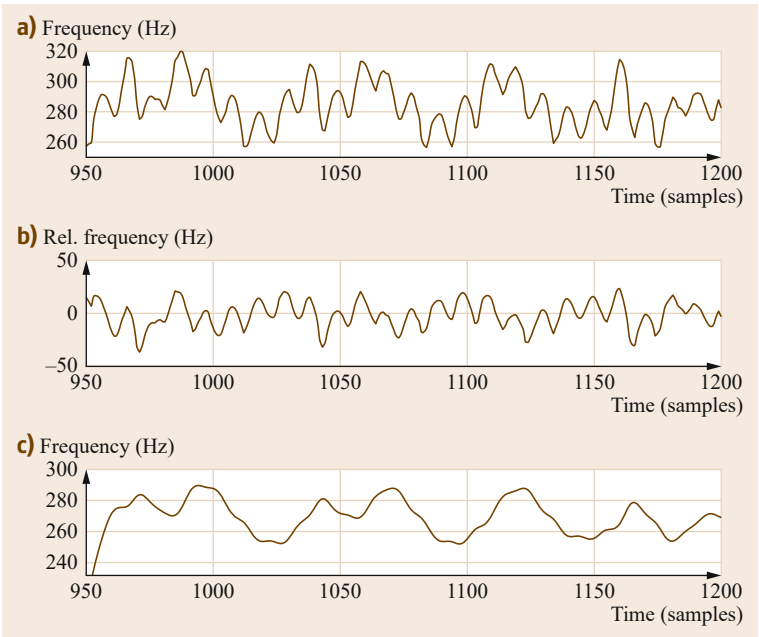


Fig. 43.9a–c Martinete excerpt; fundamental frequency contour; **(a)** unfiltered; **(b)** vibrato; **(c)** slow pitch modulation

who is watching her loved one moving away. The resignation appears in phrases such as E-F-G-E or E-F-D-E, and the distance of the loved one shows up in the alteration of F to F# in the phrase E-F-G (which becomes E-F#-G).

Inshád in Al-Sika Mode by Zohra Abbetiw

The present study focuses on the analysis of an *inshád* performance in *Al-Sika* mode by Zohra Abettiw, a female singer from the school of Tetouan (north Morocco). She is accompanied by the Orchestra of the Tetouan Conservatory, led by Mohammed Ben Larbi Temsamani. The audio piece is a recording of the Moroccan national radio and television and it dates back to the 1960s. The melody ranges from B_3 to B_4, if we consider the ornamentation notes as part of the melody, that is, the interpretation as a whole. The studied *inshád* consists of three blocks. The first and the third blocks contain three phrases. The first phrase is a typical phrase in *Al-Sika* mode and acts as the presentation of the *inshád*. The second phrase ends with on the note C, where the substitution of an F with an F# symbolizes the feeling of distance to the loved one. Finally, the third phrase is similar to the second phrase, but it connects with the last interval, which corresponds to the final *centon* (E-F-D-E). On the other hand, the second block is the shortest, containing only one isolated

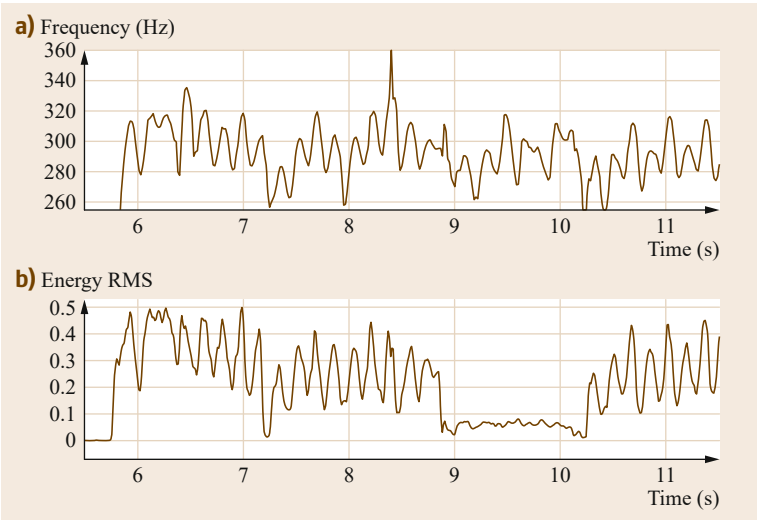


Fig. 43.10a,b Martinete excerpt; **(a)** fundamental frequency contour; **(b)** energy trajectory

Table 43.2 Comparison of statistical note and vibrato descriptors for the considered pieces

Descriptor	Martinete	Inshád
Average vibrato rate (Hz)	6.19	5.91
Average vibrato extent (cents)	75.47	62.02
Vibrato amount (%)	41	64
Lowest pitch (MIDI pitch)	53	59
Highest pitch (MIDI pitch)	64	71
Mean pitch (MIDI pitch)	60.83	65.03
Pitch fluctuation (MIDI pitch)	2.11	3.17
Pitch range (MIDI pitch)	11	12
Average note duration (s)	0.34	0.42
Note duration fluctuation (s)	0.24	0.43
Onset rate (notes/s)	2.52	1.61

phrase, which is again similar to the second phrase of the first block.

43.3.3 Computational Analysis

This section contrasts previous manual analyses with computational description tools. We compare fundamental frequency envelopes and symbolic note representations and compute statistical descriptors related to pitch content and vibrato. We analyze pitch histograms in relation to the tonal concepts introduced in the previous section.

Melodic Contours

We first compare automatically computed fundamental frequency envelopes and melodic contours of both pieces, as illustrated in Fig. 43.8. We observe that notes possess more ornamentation in the *martinete* than in the *inshád* performance. The fundamental frequency contour of the *martinete* excerpt can be modeled by an overlay of two periodic pitch fluctuations: the high frequency voice vibrato and a slow periodic melisma. When filtering the fundamental frequency envelope at suitable cutoff frequencies, both periodic pitch fluctuations can be visualized (Fig. 43.9). By analyzing the energy trajectory of both excerpts, we furthermore observe that the fast pitch fluctuation is accompanied by a simultaneous dynamic modulation (Figs. 43.10 and 43.12). The fundamental frequency contour of the *inshád* piece shows glissandi in note transitions and vibrato during longer notes, but we do not observe melismatic ornamentation (Fig. 43.11).

Statistical Descriptors

From both automatic transcriptions, we extract statistical features related to the pitch content and vibrato, which are presented in Table 43.2. In both pieces we observe an extensive use of vibrato, as shown by the *vibrato amount* descriptor. The corresponding average frequency and extent reflect the singer's individual characteristics. The pitch range of both pieces is comparable. The pitch content of the *inshád*, sung by a female singer, is located in a higher register, with a higher average pitch. The comparatively high onset rate calculated for the *martinete* originates from the large amount of melismatic ornamentation and the low pitch fluctuation relates to the insistence on the recitative notes, while the pitch content of the *inshád* appears to be wider spread.

Pitch Histograms

By analyzing the temporal note predominances in the automatic transcriptions, we obtain pitch class histograms, illustrated in Fig. 43.13. All occurring pitches are wrapped in a single octave and displayed relative to their tonic. The tonal center (E) and the fundamen-

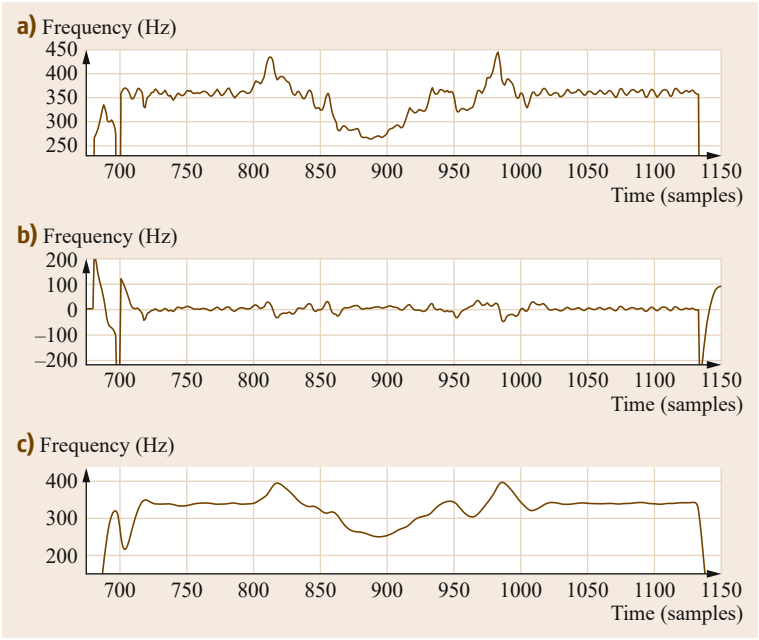


Fig. 43.11a–c Inshád excerpt; fundamental frequency contour; **(a)** unfiltered; **(b)** vibrato; **(c)** slow pitch modulation

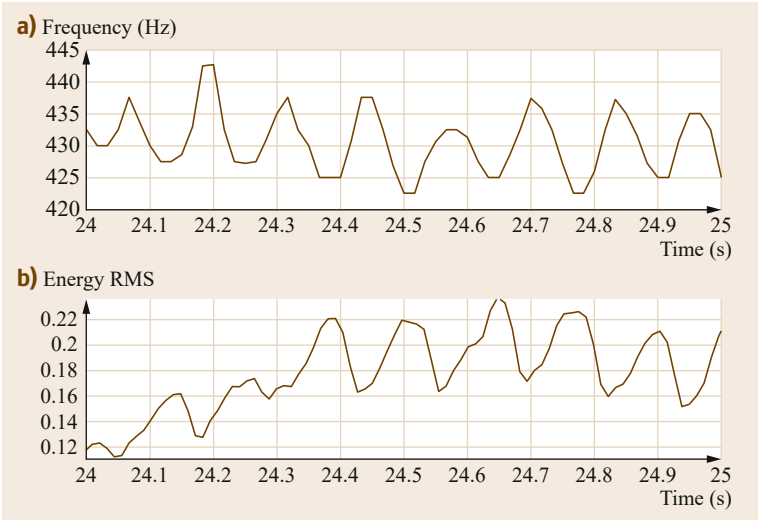


Fig. 43.12a,b Inshád excerpt; **(a)** fundamental frequency contour; **(b)** energy trajectory

tal scale degrees (C and G) of the *Al-Sika* mode in the Andalusian piece appear with a significantly higher frequency throughout the contour. Also the modulation from F to F# can be observed. The pitch content of the *martinete* is concentrated in the first six chromatic scale degrees. It can be seen that the total duration of occurrence of the tonic is lower than for the fourth, fifth and sixth degree, which are extensively used in melismatic ornamentations around the recitative notes. Consequently, the pitch content is mainly concentrated between the tonic and the subdominant and includes an elevated occurrence of notes that do not belong to the underlying scale. In contrast, the *inshád* shows a wider-spread distribution of all pitch classes over the octave, strongly weighted by the importance of the corresponding scale degrees.

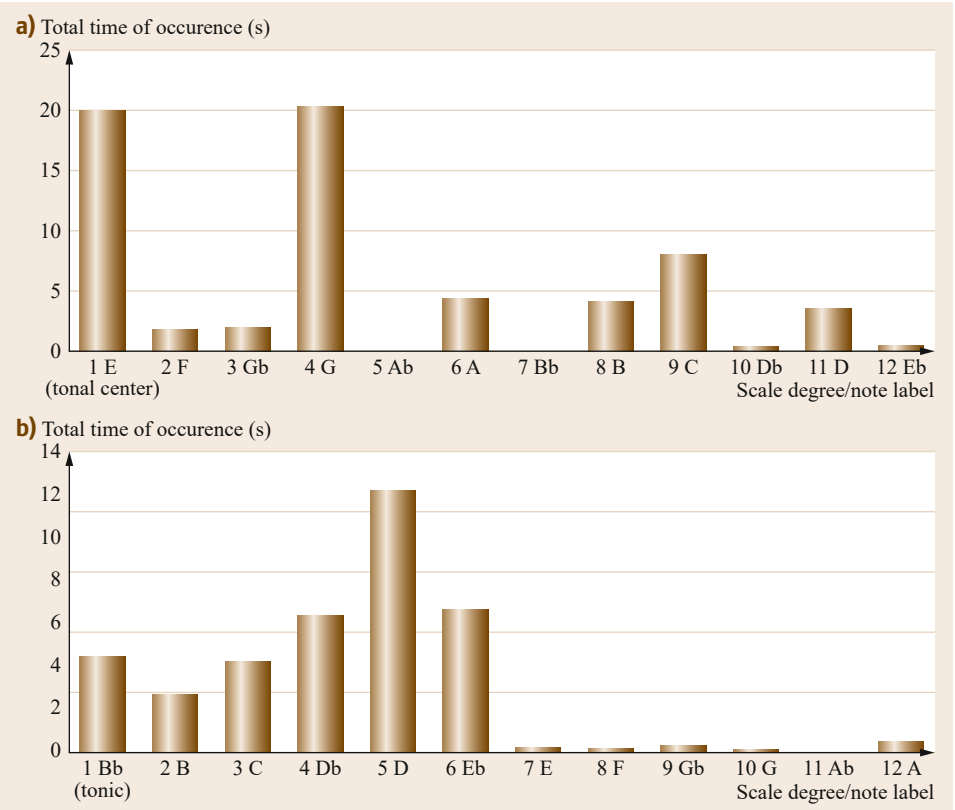


Fig. 43.13a,b Pitch histograms from note transcriptions; **(a)** inshád; **(b)** martinete

43.4 Conclusion and Future Perspectives

As we stated at the outset, our intention in carrying out this piece of research is to show how to use computational tools in a meaningful manner, beyond mere consideration of them as powerful or sophisticated tools. In this study, we have presented a methodology for the analysis of vocal music of two complex, rich music traditions, flamenco and Arab-Andalusian music. We adopted the audio approach because we consider it more appropriate for the repertoires under study, given the complex ornamentations and the absence of score. We have studied vocal music and a symbolic approach would have missed many musical phenomena of importance, such as vibrato or microtonality. Also, this approach allows automatic computational analysis of large corpora, as it can easily be deduced from the description of our methodology.

However, the proposed methodology does not ignore the musicological aspects nor concentrates on physical features of the audio files. To the contrary, this work is the fruit of the collaboration of experts on flamenco and Arab-Andalusian music and computer scientists, thus showing that, because music is a complex and multidimensional phenomenon, its analysis should a be carried out from variety of standpoints and with various tools.

In particular, we have presented a methodology for the analysis of music recordings that combines manual and automatic descriptions of music recordings by using state-of-the-art techniques. Through our computational analyses, we detected similarities and differences on the two analyzed pieces, *martinete* and *inshád*, in terms of ornamentation, melodic line, scale, and vibrato. Although these properties refer to specific pieces, some of the traits are representative of flamenco and Arab-Andalusian music traditions. In the future, we intend to carry out a more extensive analysis that will include a large number of pieces of each music tradition.

Our methodology can be applied to other pieces and traditions, but proper adaptation should be carried out. As a matter of fact, this adaptation should be based on

the theoretical foundations of the tradition under study. Our endeavor is to provide novel tools and methodologies for the analysis and description of traditional music that may complement traditional approaches.

43.5 Complementary Material

The link http://mtg.upf.edu/flamenco-arabandalusian contains additional material to foster reproducibility of the presented study:

- Audio files
- Manual transcriptions
- Automatic transcriptions and descriptions
- Multimedia examples from both music traditions.

Acknowledgments. This research was partly funded by the Spanish Ministry of Economy and Competitiveness (grant TIN2012-36650), the Junta de Andalucía (grant P09-TIC-4840), the FEDER funds of the European Union and the PhD fellowship program of the Department of Information and Communication Technology (DTIC), Universitat Pompeu Fabra.

References

43.1 M.A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, M. Slaney: Content-based music information retrieval: Current directions and future challenges, Proc. IEEE **96**(4), 668–696 (2008)

43.2 O. Cornelis, M. Lesaffre, D. Moelants, M. Leman: Access to ethnic music: Advances and perspectives in content-based music information retrieval, Signal Process. **90**(4), 1008–1031 (2010)

43.3 N. Orio: Music retrieval: A tutorial and review, Found. Trends Inf. Retr. **1**(1), 1–90 (2006)

43.4 E. Gómez, P. Herrera, F. Gómez-Martin: Computational ethnomusicology: Perspectives and challenges, J. New Music Res. **42**(2), 111–112 (2013)

43.5 I. Halmos, G. Köszegi, G. Mandler: *Computational Ethnomusicology in Hungary in 1978* (Univ. of Michigan Library, Ann Arbor 1978)

43.6 G. Tzanetakis, A. Kapur, W.A. Schloss, M. Wright: Computational ethnomusicology, J. Interdiscip. Music Stud. **1**(2), 1–24 (2007)

43.7 F. Gouyon, P. Herrera, E. Gómez, P. Cano, J. Bonada, A. Loscos, X. Amatriain, X. Serra: Content processing of music audio signals. In: *Sound to Sense, Sense to Sound: A State-of-the-Art in Sound and Music Computing*, ed. by P. Polotti, D. Rocchesso (Logos, Berlin 2008) pp. 83–160

43.8 J. Serrà, E. Gómez, P. Herrera: Audio cover song identification and similarity: Background, approaches, evaluation, and beyond. In: *Advances in Music Information Retrieval*, Studies in Computational Intelligence, Vol. 274, ed. by Z. Ras, A.A. Wieczorkowska (Springer, Berlin, Heidelberg 2010) pp. 307–332

43.9 A.C. Gedik, B. Bozkurt: Pitch-frequency histogram-based music information retrieval for Turkish music, Signal Process. **90**(4), 1049–1063 (2010)

43.10 S. Ness, D.R. Biró, G. Tzanetakis: Computer-assisted cantillation and chant research using content-aware web visualization tools, Multimed. Tools Appl. **48**(1), 207–224 (2010)

43.11 J. Six, O. Cornelis, M. Leman: Tarsos, a modular platform for precise pitch analysis of western and non-western music, J. New Music Res. **42**(2), 113–129 (2013)

43.12 P. Chordia, A. Rae: Raag recognition using pitch-class and pitch-class dyad distributions. In: *Proc. ISMIR Conf* (2007) pp. 431–436

43.13 A. Kruspe, H. Lukashevich, J. Abesser, H. Grossmann, C. Dittmar: Automatic classification of musical pieces into global cultural areas. In: *Proc. 42nd AES Int. Conf. Semant. Audio* (2011)

43.14 Universitat Pompeu Fabra: http://mtg.upf.edu/research/projects/cofla

43.15 J.M. Díaz-Báñez, G. Farigu, F. Gómez, D. Rappaport, G.T. Toussaint: Similaridad y evolución en la rítmica del flamenco: Una incursión de la matemática computacional, La Gaceta de la Real Sociedad Matemática Española **8**(2), 489–509 (2005)

43.16 J. Mora, F. Gómez, E. Gómez, F. Escobar-Borrego, J.M. Díaz-Báñez: Characterization and melodic similarity of a cappella flamenco cantes. In: *Proc. 11th Int. Soc. Music Inf. Retr. Conf. (ISMIR)* (2010)

43.17 N. Kroher: *The Flamenco Cante: Automatic Characterization of Flamenco Singing by Analyzing Audio Recordings*, Master Thesis (Universitat Pompeu Fabra, Barcelona 2013)

43.18 F. Gómez, A. Pikrakis, J. Mora, J.M. Díaz-Báñez, E. Gómez: Automatic detection of ornamentation in flamenco. In: *4th Int. Workshop Mach. Learn. Music* (2011)

43.19 A. Pikrakis, F. Gómez, S. Oramas, J.M. Díaz-Báñez, J. Mora, F. Escobar, E. Gómez, J. Salamon: Tracking melodic patterns in flamenco singing by analyzing polyphonic music recordings. In: *13th Int. Soc. Music Inf. Retr. Conf. (ISMIR), Porto* (2012)

43.20 D. Wu: Simultaneous unsupervised learning of flamenco metrical structure, hypermetrical structure, and multipart structural relations. In: *Proc. 14th Int. Soc. Music Inf. Retr. Conf* (2013)

43.21 J. Blas Vega, M. Ríos Ruiz: *Diccionario Enciclopédico Ilustrado del Flamenco* (Cinterco, Madrid 1988)
43.22 J.L. Navarro, M. Ropero: *Historia del Flamenco* (Tartessos, Seville 1995)
43.23 J.M. Gamboa: *Una Historia del Flamenco* (Espasa-Calpe, Barcelona 2005)
43.24 F. Merchán Higuera: *Expressive Characterization of Flamenco Singing*, Master thesis (Universitat Pompeu Fabra, Barcelona 2008)
43.25 R. Molina, A. Mairena: *Mundo y Formas del Cante Flamenco* (Librería Al-Andalus, Seville 1963)
43.26 A. Chaachoo: *La Música Andalusí. Al-Ála. Historia, Conceptos y Teoría Musical* (Almuzara, Córdoba 2011)
43.27 C. Poché: *La Musique Arabo-Andalouse* (Cité de la Musique/Actes Sud, Paris 1995)
43.28 M. Guettat: *La Musique Arabo-Andalouse. L'empreinte du Maghreb* (El Ouns/Fleurs sociales, Paris 2000)
43.29 M. Cortés García: *Kinnas al-Haik* (Centro de Documentación Musical de Andalucía, Granada 2003)
43.30 E. Gómez, J. Bonada: Towards computer-assisted flamenco transcription: An experimental comparison of automatic transcription algorithms as applied to a cappella singing, Comput. Music J. **37**(2), 73–90 (2013)
43.31 A. Klapuri, M. Davy: *Signal Processing Methods for Music Transcription* (Springer, New York 2006)
43.32 G. Tzanetakis, A. Ermolinskyi, P. Cook: Pitch histograms in audio and symbolic music information retrieval, J. New Music Res. **32**(2), 143–152 (2003)
43.33 P. Lefranc: *El Cante Jondo* (Secretariado de Publicaciones de la Universidad de Sevilla, Seville 2000)
43.34 B. Champigneulle: *Histoire de la Musique* (Presses universitaires de France, Paris 1969)

44. The Relation Between Music Technology and Music Industry

Alexander Lerch

The music industry has changed drastically over the last century and most of its changes and transformations have been technology-driven. Music technology – encompassing musical instruments, sound generators, studio equipment and software, perceptual audio coding algorithms, and reproduction software and devices – has shaped the way music is produced, performed, distributed, and consumed. The evolution of music technology enabled studios and hobbyist producers to produce music at a technical quality unthinkable decades ago and have affordable access to new effects as well as production techniques. Artists explore nontraditional ways of sound generation and sound modification to create previously unheard effects, soundscapes, or even to conceive new musical styles. The consumer has immediate access to a vast diversity of songs and styles and is able to listen to individualized playlists virtually everywhere and at any time. The most disruptive technological innovations during the past 130 years have probably been:

1. The possibility to record and distribute recordings on a large scale through the gramophone.
2. The introduction of vinyl disks enabling high-quality sound reproduction.
3. The compact cassette enabling individualized playlists, music sharing with friends and mobile listening.
4. Digital audio technology enabling high quality professional-grade studio equipment at low prices.
5. Perceptual audio coding in combination with online distribution, streaming, and file sharing.

This text will describe these technological innovations and their impact on artists, engineers, and listeners.

The development of the music industry has always been closely related to the creation of new music technology. Many changes in the industry can be directly related to technological innovation. It can be argued that technology shaped the music industry as much as the industry shaped technological evolution.

During the 19th century, the distribution of music was practically limited to sheet-music; music could only be consumed live in concert halls and at private concerts. The invention of recording devices such as the phonograph [44.1], the graphophone [44.2], and the gramophone [44.3] in the late 19th century changed that radically: the possibility to record and reproduce a live performance not only allowed publishers to mass-produce and distribute records, but eventually enabled the consumer to listen to a recording anywhere and anytime. Additional technologies such as radio broadcasting and, much later, mobile playing devices and internet streaming enforced this development: music is now omnipresent in our daily lives.

After the music industry changed from being dominated by sheet-music publishers to a new industry with a focus on recording performances and the mass production and distribution of media, it adapted and reinvented itself several times over the last century. The landscape of today's music industry includes musicians, record labels, rights holders, distributors, recording studios, manufacturers of instruments, and manufacturers of production and reproduction equipment and software. The impact of technological evolution on how

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_44

music is produced, recorded, and distributed is easily identifiable from a modern perspective, but on closer inspection it also becomes apparent how technology changed the way music is consumed and, within certain limits, influenced music performance styles and music composition as well.

This influence is not only identifiable in retrospect – many technological innovations led to a lively discussion in the music world after they were introduced. Just to give two examples, the conductor *Furtwängler* commented on radio broadcasting [44.4]:

> *The sapless and insubstantial rehash that broadcast listeners hear will only be interpreted as an adequate alternative for the concert by those who do not know the real concert anymore (translated from German by the author: Den vitaminlosen, saft- und kraftlosen Aufguss, den die Hörer des Rundfunks von einem Konzert erhalten, können wirklich nur diejenigen für vollwertigen Ersatz des Konzertes halten, die nicht mehr wissen, was ein wirkliches Konzert ist.)*

while the composer Babbitt noted [44.5]:

> *I can't believe that people really prefer to go to the concert hall under intellectually trying, socially trying, physically trying conditions, unable to repeat something they have missed, when they can sit at home under the most comfortable and stimulating circumstances and hear it as they want to hear it.*

Table 44.1 lists the approximate release dates for selected technological innovations as mentioned later in this text. The two columns of the table categorize the technology into equipment for recording, music production, and music performance (left) and technology for music distribution and consumption (right).

The remainder of this text is structured as follows: the following section will describe the interaction of music technology with both the performance and the recording of music, and the impact of the technological evolution on the recording studio sector. Section 44.2 discusses the creative uses of technology. The influence of technology on the distribution and the consumption of music will be reviewed in Sect. 44.3.

Table 44.1 Approximate dates of technological innovation in the music industry

Year	Production	Consumption
1878	Phonograph (Edison)	
1885	Graphophone (Bell and Tainter)	
1887	Gramophone (Berliner)	
1920	Condenser microphone (Wente)	
1923		Radio broadcasting
1928	Neumann CMV3 condenser microphone Theremin	
1930	Trautonium	
1933		Stereophony
1935	Hammond organ	
1941	High-quality tape recording (AEG)	
1948		12″ 33-1/3 rpm LP (Columbia)
1949		7″ 45 rpm single (RCA Victor)
1957	Ampex 8-track recorder	
1959	Fender Rhodes Piano	
1963		Philips compact cassette
1976		Dolby Stereo
1978	Lexicon L224 Reverb 3M digital multitrack	
1979	NED Synclavier	Sony Walkman
1980	Sony DAE-1100	
1981	Sony PCM-3324 digital multitrack	
1983	Yamaha DX-7 MIDI	
1985	CSound	CD (Sony, Philips)
1987	DAT	
1988	Harddisk Editing (Sonic Solution) Akai MPC and S1000	
1990		MP3
1991	Alesis ADAT	Sony MiniDisc

Table 44.1 (continued)

Year	Production	Consumption
1993	TASCAM DA-88	ITU 5.1 Multichannel
1995		Real-Audio Streaming
1997	Antares Auto-tune	DVD
1998		MP3-Player
1999		File Sharing (Napster) SACD
2000		DVD-Audio
2001		iPod (Apple)
2003		iTunes Music Store (Apple)

44.1 Recording and Performance

The recording process attempts to capture a music performance and its characteristics in order to reproduce it later. However, it does not capture all details of a performance; elements such as the communication by body language, gestures, and facial expression will be missing from the recording. Furthermore, the experience of performing for a recording can be significantly different from a live performance. The suspense of performing in front of an audience is missing, and no interaction with the audience is possible. There is also the recording process itself that may impact the performance: the musicians can perform a piece of music, listen to the recorded result, and discuss and possibly adapt their performance. The pianist *Glenn Gould* describes the advantages of this process [44.6]:

> *By taking advantage of the post-taping afterthought, however, one can very often transcend the limitations that performance imposes upon the imagination.*

The production team may also influence the performance by pointing out technical necessities and by making suggestions on how to better convey musical thoughts over the medium.

Since the introduction of multitrack recording machines, each instrument or instrument group can be recorded separately on its own track. After the first analogue multitrack machines were released in the 1950s (e.g., *Ampex* 8-track), the number of tracks increased over the decades; digital multitrack tape machines were released in the late 1970s and early 1980s with a number of tracks between 24 and 32, followed by the cassette-based recorders *Alesis ADAT* and *Tascam DA-88* in the 1990s. Both the ADAT and the DA-88 allowed one to sync several machines up to an overall limit of 128 simultaneous tracks. Modern digital audio workstations (DAWs) have no restriction with respect to the maximum number of tracks. In the context of classical music, multitrack recordings are primarily used to allow for postprocessing of the recording by adjusting the level and equalization of individual tracks. In the case of popular music, however, multitrack recording not only offered enhanced postproduction options but it transformed the way music is recorded: instead of recording all musicians of a band simultaneously, the sound engineers could record individual instruments at different times and possibly different places. This process is not applicable to all styles of music, but its implications for the performance are noteworthy; not only have the musicians lost the communication with the audience, but they are not even able to communicate with each other during the recording session. Musicians who have never met can *perform together* by sending the recording to a different studio and record the next musician at a different time and a different place.

It has also been argued that the process of recording and the public availability of recordings can influence long-term performance strategies and trends. For example, *Mark Katz* points out that the increased use of vibrato at the begin of the twentieth century coincides with the increasing availability of the gramophone. He makes a compelling argument that this change in performance style was motivated and triggered by the growing popularity of the gramophone [44.7]. Katz lists three reasons for the violinists to increase their vibrato usage for a recording:

1. It allowed the violin to be more perceptible on early recordings with limited dynamic range and poor transfer characteristics.
2. It helped to obscure imperfect intonation which is more likely to be noticeable on a recording than in a live setting.
3. It might enhance the artist's presence in order to compensate for the missing visual cues in a concert situation.

The so-called crooning is another example of the inter-relation of technology and performance styles.

Crooning was a popular singing style in the 1930s to the 1950s that created a personal atmosphere by singing at a low volume. The vocals could only carry over the accompanying band or orchestra by means of amplification of the microphone signal. Singers such as Bing Crosby and Frank Sinatra used this technique to create intimacy between audience and singer.

44.1.1 Recording Perfection

Technological innovations and improvements increased the sonic quality of recording and reproduction equipment continuously throughout the 20th century. The most notable quality improvements were probably a) the transition from an acoustic to an electric way of recording by means of the condenser (also: capacitor) microphone (1920s: [44.8]), b) the possibility to make high-quality recordings on tape (1940s), and c) the introduction of digital means for recording, processing, editing, storing, and distributing music (1970s–1990s).

These general improvements in audio quality are an unsurprising and foreseeable technological evolution. However, it was less foreseeable how technological possibilities would influence the musical content itself. It has been observed that the level of perfection of recorded music performances has been continuously increasing over the last century [44.9]. The main reason is probably that a recording, unlike a live performance, can be listened to repeatedly and is preserved for a long time. Even a minor imperfection or performance nuance, barely noteworthy in a live setting, can become an annoying artifact. The technical flawlessness of a studio recording may, in some ways, compensate for the missing live concert experience.

Editing (or *splicing*) a recording is an aspect of the postproduction process that originated in the 1940s and 1950s. While in early recording sessions the artist had to perform the whole piece for the recording process and repeat this until satisfaction was obtained, technological development opened up new ways of dealing with nonoptimal recordings. After the introduction of the first tape machines it was soon discovered that the tape could be split and joined together. This allowed for merging two or more recordings of the same piece of music; a section with a small mistake could easily be replaced with another recording of this section from the same recording session. The advent of digital processing allowed an unprecedented level of editing; *Weinzierl* and *Franke* presented evidence that the amount of edit points increased by a factor of two to three with the introduction of digital editing stations such as the *Sony DAE-1100* and digital audio workstations (DAWs) [44.10]. It is a valid assumption that modern productions make even more extensive use of the possibilities of splicing and joining of two or more recordings at virtually any position; hundreds of edit points can be expected in such a modern production. Note that the editing process not only involves removing errors and mistakes from the recording – the editor essentially has the power to select and merge takes according to his own artistic taste. This choice does not necessarily reflect the musical intentions of the artists, although it is common for the artists to give final approval of an edited recording.

The modern postproduction process offers more options to change a recording. Traditionally, these modifications were restricted to changing the sound quality by using equalization or adding artificial reverberation and (passively or actively) modifying the dynamic variation or the level relationships between instruments. Newer digital tools, however, allow more drastic changes of the music content. For instance, technologies such as time-stretching (changing the tempo without changing the pitch) and pitch-shifting (changing the pitch without changing the tempo) allow detailed modifications of the (micro-)tempo and the pitch in order to correct or modify intonation and timing. The final, published recording thus evolves more and more from the *plain* reproduction of one recording studio performance to a new performance with a level of technical perfection hard to match in a live context.

44.1.2 Recording Authenticity

Throughout the last century the music industry put significant effort into marketing records as a realistic medium which is indistinguishable from the live concert experience [44.11]. However, listeners realized early on that playing recorded music does in fact not result in the same experience as a live concert. In the early days of recording, one important factor was clearly the quality of the recording and reproduction systems. To give an example, the dynamic range of the recording devices was limited which meant that instrumentalists had to minimize the dynamic range of their performance or continuously change the distance from the recording device to adjust for the correct level. The quality-related restrictions such as insufficient dynamic range, added noise, and added distortion disappeared with improved recording technology and better storage media. However, while research in virtual acoustics led to recording and reproduction methods that enabled a listening experience that was practically indistinguishable from a live setting [44.12], the standard consumer reproduction technologies (e.g., stereo, 5.1) are unable to reproduce this concert experience.

Since the consumer's normal listening environment is fundamentally different from a live concert setting, it

is doubtful that even an accurate physical reproduction of the sound field would be perceived by the listener as realistic. Therefore, the sound engineer's goal is usually not to create a realistic reproduction, but a recording that is *perceived* as realistic and thus implies realism. This has been referred to as the *concept of artificial realism* [44.13].

It has also been argued that, because of limitations in the listening environment, the recording is not even supposed to be as realistic as possible. To give an example, the dynamic range of symphonic recordings is usually reduced to allow a comfortable listening experience – to reproduce the whole dynamic range of an orchestra would result in the listener constantly adjusting the playback level. The musically informed adjustment of the level of single instruments and instrument groups by the sound engineer may compensate for the missing visual component and probably allows for a better, more transparent listening experience.

Similar to the discussion of realism in classical music, there has been a debate about authenticity in popular music as well [44.14]. The controversy is about the use of new technologies to produce or postprocess recordings and the utilization of new effects and sounds versus a *plain* recording made with as few technological gadgets as possible. *Alex Ross* comments on this discussion [44.15]:

> *Yet frills-free, low-fi recording has no special claim on musical truth; indeed, it easily becomes another phonograph effect, the effect of no effect.*

44.1.3 Changes in the Recording Studio Sector

The changes in the recording studio sector are well aligned with technological innovation. Leyshon summarizes the different stages of historical development [44.16]: during the 1930s and 1940s, studios were built by the major labels with custom-designed and unique equipment. The introduction of the tape machine in the 1940s made the recording process less expensive and triggered the establishment of new, privately owned recording studios in the 1950s. Digital audio technology disrupted the recording studio market twice: first, with the introduction of digital studio hardware for recording and processing offering a sonic quality and functionality at the same or a higher level than comparable analogue hardware at a fraction of the price. Second, with software applications offering to replace most of the studio hardware (except microphones and speakers) at affordable prices. Essentially, computer software is able to replace the tape machine, patch bay, mixing console, sound generators and synthesizers, and external audio effects. Although analogue studio equipment continues to be used for its special sound characteristics, virtually every modern music production will use software; for some genres, music may be produced exclusively by using software.

There is a general trend for high-quality audio technology, previously only available to specialists at high prices to become more easily available and affordable to nonprofessionals and hobbyists. The price drop is not only observable between hardware and software, but also in the software market itself. To give one example, the sales price of *Apple's* DAW Logic Pro dropped from $999 to $499 and then to today's price of $199 between 2004 and 2011. Similar price developments can be observed for professional-grade audio plugins. This democratizes the recording market by decoupling audio quality from the budget invested into the studio equipment.

For the music (software) industry, however, this presents a problem due to decreasing revenue. As a result, new distribution models are being evaluated. Current trends, possibly influenced by the rise of the mobile app market, include consumer versions of software targeted at professionals at low prices and profit generation through sales of additional content such as audio loops. Software-as-a-Service solutions are not yet popular in the consumer market, but are starting to gain traction in the business-to-business market with companies such as *EchoNest* (now part of Spotify), *Gracenote*, and *sonicAPI.com*.

44.2 Music Creation

The creative interaction of musicians and composers with technology is a fascinating topic. New technologies have always inspired composers and musicians to explore new ways of creating sound and music, either by using new or modified musical instruments, by integrating new technology into their work, or by *misusing* gear to create new sounds and new musical concepts.

44.2.1 Instruments

The development of music instruments is often closely related to technology; to give two historic examples, the mechanical systems of wind instruments allowed a broader pitch range and easier intonation, and improved piano mechanics increased volume and expressive control of the instrument. Both changes were

welcomed by composers and musicians alike and were accepted quickly. In the twentieth century, electric and electronic instruments added to the well-established pool of traditional instruments. Some of them, despite a certain fame, were never really accepted by large numbers of performers; well-known examples from the 1920s are Trautwein's *Trautonium* and Theremin's instrument (later simply referred to as *Theremin*). These two instruments have been used in art music, popular music, and film music, however, they still remain more or less unknown to nonmusicians.

Other instruments, however, have been seminal to certain genres. The genre of rock music is hard to conceive without the (distorted) electric guitar, and synthesizers certainly contributed to shaping the sound of pop music. Instruments such as the *Hammond* organ, the *Rhodes* piano, and synthesizers such as the ones invented by *Moog* expanded the dimensions of sound and enabled musicians to generate new, never-heard-before sounds. With growing popularity, these new instruments created new markets and a new instrument manufacturer landscape. As in other areas, the introduction of digital technology was the next major step in instrument design. The famous *Yamaha DX-7* synthesizer, for instance, allowed exploration of the unknown sound spaces of FM synthesis. In addition to sound synthesis, digital audio allowed sampling (see Sect. 44.2.4); the introduction of *Akai's S-1000* and *MPC* is seen as a milestone in popular music production and sound design.

Meanwhile, performers and composers of electro-acoustic music moved on to other, non-mainstream tools. Examples of these tools are the audio programming languages *CSound* and *SuperCollider* [44.17], and solutions such as *Max/MSP* [44.18] or *PD* [44.19], software that allows the rapid design and prototyping of real-time audio processing and synthesis systems with a visual programming language.

There is academic and commercial interest in developing new software tools and new devices for sound synthesis such as virtual instrument plugins; the interested reader is referred to events such as the *International Conference on New Interfaces for Musical Expression* (NIME) [44.20] and the *Margaret Guthman Musical Instrument Competition* [44.21]. However, new instruments with new interfaces that are used by a larger group of musicians are not easy to find. One prominent example of a more recently introduced musical interface is the *reacTable* [44.22], an instrument which was featured by the artist Björk on her 2007 tour.

Playback media have also been used as instruments or as part of a performance. Wolpe, for instance, experimented with gramophones on the stage. He remembers a concert in Berlin in the early 1920s in a lecture [44.23]:

> *I had eight gramophones, record players, at my disposal. And these were lovely record players because one could regulate their speed. Here you have only certain speeds – seventy-four and so on – but there you could play a Beethoven symphony very, very slow, and very quick at the same time that you could mix it with a popular tune.*

However, despite these efforts and related experiments by Hindemith at the time, the active use of turntables as musical instruments did not establish itself until the rise of the DJ in the 1980s. DJs can use the turntable to create new sounds and sound effects by means of mixing the sound of two vinyl discs dynamically and modifying the playback speed and the playback direction in fast patterns (scratching). Effectively, they use the turntable as a musical instrument. Another recording and playback medium has been more successful in electro-acoustic music: the tape machine. Starting in the early 1950s, composers such as Cage, Varèse, and Stockhausen actively used tapes and tape machines in their compositions and performances. The creative use of the tape machine was not restricted to electro-acoustic music: The Beatles also experimented with tape loops, for example, in their song *Tomorrow Never Knows* on the studio album *Revolver*, enabling them to create a soundscape not reproducible in a live performance.

44.2.2 MIDI

It was not only new instruments or devices reused as musical instruments that had an impact on the creation of music. The MIDI protocol, introduced in the 1980s and eventually supported by a large number of manufacturers, allowed musicians to connect two or more keyboards and sound generators and to slave an arbitrary number of synthesizers in order to play them simultaneously, to switch instantaneously between the instruments without changing the control keyboard itself, or to send or store configuration information from or to devices [44.24]. The introduction of MIDI, combined with the advances in computer hardware and software, resulted in the rapidly growing popularity of the so-called sequencer. A sequencer could record, edit, store, and reproduce a performance and corresponding data in MIDI. It also allowed prototyping and evaluation of the arrangement of complete songs in a small home-recording set-up or playback of an additional synthesizer track during a live performance. A sequencer can be either a hardware or a software product. Over the last two decades, especially soft-

ware sequencers gained new features – last but not least the ability to record and edit multitrack audio data – so rapidly that software-based sequencers became the most common tool to work with MIDI and audio in the studio. Today they are usually referred to as digital audio workstations (DAW). There are variants of DAW software (for example, *Ableton Live*) and DJ software (for example, *Serato Scratch Live* or *Native Instrument's Traktor DJ*) which try to redefine the role of the computer as a mere recording, editing, and playback device to an instrument that can be interacted with in a live performance.

44.2.3 Audio Effects

While in the production of classical music audio effects such as equalization and artificial reverberation are mainly used to increase the realism and the pleasantness of the listening experience, the creative potential of audio effects in popular music has been explored extensively. Dynamics processing, modulated effects, delay and reverberation, distortion, etc. became standard tools to work on the sound of individual tracks or on the complete mix. Digital technology not only made these effects more affordable but also allowed unprecedented quality (artificial reverberation), flexibility (digital amp simulation), and not-heard-before effects (pitch-shifting). Pitch-shifting, for example, can be used to correct the intonation of a (vocal) performance, but it can also be used as a sound effect. One example is the so-called auto-tune effect, initially conceived as a real-time vocal intonation correction tool. By setting the parameters of this auto-tune effect to very aggressive settings instead of the recommended settings, a very audible effect is created. The robotic character of singer Cher in her song *Believe* stems from using an auto-tune effect, and the artist T-Pain is frequently associated with this effect as well. Pitch-shifting can also be used for auto-harmonization (i. e., for the generation of additional voices from a single-voiced input in order to create a choir-effect), or to transform musical phrases into something impossible to perform.

44.2.4 Sampling

With digital recording and storage, the phenomenon called *sampling* appeared. In digital sampling, an artist records a segment of a song or sound, modifies, edits, and transforms it, and then reuses it (and possibly other recordings) by incorporating it into a new work. Musical quotations are well known in the music history, however, sampling allows what *Katz* refers to as *performative quotation* [44.7]; not only does the quote feature the same melodic, harmonic, and rhythmic content, but it also has the same timbre and can be, in fact, a bit-accurate copy of the original recording. Hip hop and electronic music are examples of genres with intensive use of sampling. Sampling is legally controversial, but also enabled an interesting way of producing music; since the composer works directly with sounds to produce the recording, there is no performer as intermediary between the abstract musical ideas and the sound creation. *Tara Rodgers* notes [44.25]:

> *Electronic music culture is at least in conversation, if not inextricably entwined, with the legacy of modernist notions of authorship and authenticity. But electronic musicians destabilize these notions by exposing the fluidity of boundaries between human- and machine-generated music, as well as the cultural angst over the dissolution of these boundaries.*

44.2.5 Inspiration and Restriction

From the examples presented above, it is clear how technology can shape new instruments and tools and how technology can been (mis-)used to explore new realms of musical expression. But technology can also serve as a source of inspiration; the composer Steve Reich, for instance, describes how he discovered the phase-shifting effect for his work *It's Gonna Rain* by accident while experimenting with two tape recorders playing two nearly identical loops, resulting in the loops going in and out of sync slowly [44.26]. He used the same effect later for nontape works as well.

However, apart from the role of music technology as both artistic inspiration and as a provider of new ways to create new works or new sounds, technology can also restrict creative processes. The composer *Stravinsky*, for example, wrote in his autobiography on the topic of making records of some of his music in America in the 1920s [44.27]:

> *This suggested the idea that I should compose something whose length should be determined by the capacity of the record. I should in that way avoid all the trouble of cutting and adapting. And that is how my* Sérénade en LA pour Piano *came to be written.*

When *RCA Victor* introduced the vinyl single format in addition to the vinyl LP, artists were forced to a song length of three minutes for their songs. It is unclear whether the format was introduced because most of the songs at that time had this length, or if the company thought that songs of this length would market better. Either way, the introduction of this format did have an undeniable impact on standard song lengths and possibly also on how music is marketed and distributed.

44.3 Music Distribution and Consumption

In the previous sections we considered the impact of technology on composer and performer, but technology has also transformed our listening habits. In times before recording was possible, we could only listen to live performances and had to share our experience with others in the audience. The gramophone, and later the radio, revolutionized that; without the need for a specific concert venue, recordings could be listened to at home, at convenient times, and both in a group or as an individual. And not only could recordings be listened to anytime, they could also be listened to repeatedly. Being able to reproduce the same recorded performance as often as desired allowed listeners to subject the recording to an analysis of both musical ideas and the performance in unprecedented detail.

With respect to end-user media, technological innovation during the first half of the twentieth century focused on improving the sound quality. Vinyl discs, both the LP and the single, offered a significant quality improvement over shellac, and the introduction of stereophony provided the listener with an unheard-of sound field envelopment compared to monophonic recordings [44.28].

The introduction of the Compact Cassette (*Philips*) initiated another substantial change in the way music is consumed; it not only allowed sharing of music with friends but also compilation of song collections on the medium and creation of customized playlists. *Sony's* Walkman, a portable device for playing Compact Cassettes reached high popularity in the 1980s. It enabled listeners to consume music of their choice with headphones in any environment, for example, while working out or while traveling. All these changes might be interpreted as a logical continuation of the listening habit trend we observed earlier: make music accessible anytime, everywhere, and transform the act of listening to music to a potentially more private, intimate, experience.

The impact of digital (audio) technology on the distribution has been dramatic and not unopposed as can be seen from the formation of groups such as *Musicians Against Digital (MAD)* [44.29]. The Compact Disc (CD), presented by *Sony* and *Philips*, had several advantages over the vinyl disc: it is a more compact, one-sided medium, its quality does not deteriorate with the number of playbacks, and it arguably offers higher general audio quality than vinyl. The CD also allowed the music industry to generate new revenue by re-releasing previous recordings for this new medium. It stores stereo audio signals at a sample rate of 44.1 kHz and with a word length of 16 bit per sample. On the one hand, these specific values were chosen due to technical practicalities. On the other hand, they result in an audio bandwidth of approximately 20 kHz, corresponding to the upper limit of hearing for the average young adult, and a dynamic range of roughly 90 dB, providing sufficient signal-to-noise ratio for most use cases. The resulting audio quality appears to be more than sufficient for the requirements of the average consumer, as later industry attempts to introduce higher quality and multichannel audio media such as the Super Audio CD (SACD) and the DVD-A were not met with success.

The MP3 (short for MPEG-1 Layer 3) audio format and similar formats disrupted both the distribution of music and the consumers' listening habits on various levels at the turn of the millennium [44.30]. MP3 allows storage or transmission of digital audio data at a fraction (typically around a tenth) of the original data rate while attempting to minimize perceptual differences between the original and encoded audio signal by using models of human sound perception [44.31]. The audio data is thus modified, but in most cases these changes are only noticeable at low bit rates. In combination with the growing popularity of the internet in the 1990s, the importance of the MP3 file format cannot be underestimated. Not only could listeners now have instant access to their whole music library on mobile audio players such as *Apple's* iPod, they could also access music on web sites, online music stores, peer-to-peer file sharing networks such as *Napster*, and online radio stations through streaming.

Both the rise of the internet and the introduction of perceptual audio coding techniques such as MP3 enabled the user to take habits shaped by the Compact Cassette to a different level. The coded digital files can be copied without quality loss and stored on various devices, and the network to exchange and find music extended from a circle of friends to potentially millions of listeners through forums and peer-to-peer networks. Not only is the personal music collection with you all the time, but every type of music is instantly available everywhere either by streaming audio from web radios, from a cloud storage service, or from other online services.

These technologies also changed how music is purchased by listeners. Instead of buying a physical album in a store, the consumer downloads or streams music files online. Instead of having the notion of *owning* and collecting the music, it is apparently of greater importance to the listeners to have access to any music anytime and anywhere. This behavior has several implications:

- *Access to a vast database of musical styles*. The discovery of new artists and unfamiliar musical styles is no longer restricted to suggestions by lo-

cal friends, the local record store, radio stations, and magazines. Listeners can be exposed to music they might have never come across outside the internet.
- *Quasi-instant access to all music*. If the user is interested in listening to a song, she can do so immediately.
- *Decline of the album-based distribution behavior*. Despite the vinyl single format, traditional marketing and distribution was mostly focused on the album (LP and CD); online distribution, however, is nowadays song-based. The listeners tend to listen to individual songs in uncontrolled order and do not necessarily recognize the album as a holistic work of art anymore.
- *Change in music buying behavior*. The big music labels did not anticipate the transformation of the music distribution channels; the exchange of music files over peer-to-peer networks such as Napster resulted in a decline of record sales, and the industry's reaction with lawsuits and the introduction of digital rights management (DRM) alienated the consumers. *Leyshon* argues, however, that it was not only the music downloaded for free that resulted in the decline of sales but also the growing popularity of other media such as DVDs and computer games that caught the attention and the money of young consumers [44.32].

As of the year 2000, eighty percent of the music industry (licensing and distribution) was controlled by five (now four) companies: *BMG Entertainment* and *Sony Music*, *Warner Music Group*, *EMI Recorded Music*, and the *Universal Music Group* [44.33]. As *Fox* pointed out, this agglomeration resulted in difficulties for small record companies to survive, artists having only limited control over the distribution of their music, fewer releases of new recordings due to high levels of industry concentration, and music consumers being restricted to obtaining consumers to obtain music only through industry-approved and controlled channels [44.34]. Each of these points was subject to change eventually. Smaller labels took advantage of new distribution models, artists experimented with a direct connection with their fans through social media and direct distribution models, and new companies such as *Apple*, *Google*, *Pandora*, and *Spotify* entered the music distribution market. The amount of legally obtained music, however, is probably still lower than the amount of music downloads that are illegal. There are also indications of a growing expectation amongst users that music should be freely available [44.16]. The overall volume of music sales has declined since the year 2000 (compare *RIAA Year-end Shipment Statistics*, [44.35]), but recent numbers concerning increasing digital sales and licensing income fuel hopes in the industry about a trend reversal [44.36].

As the introduction of digital recording equipment with affordable prices has democratized the recording market (see Sect. 44.1.3), *Hracs* argues that the technological change eroding the power of the major record labels will ultimately lead to a democratization of the production and distribution of music [44.37].

44.4 Conclusion

Technological innovation transformed the way music is performed, recorded, produced, distributed, and consumed. It expanded creative artistic possibilities, supported the creation of new musical styles, and changed the consumers' listening habits. Music technology can be categorized as serving one or more of the following purposes:

1. Enhance the technical and perceptual quality of recording and reproduction.
2. Provide new tools to create music, synthesize sound, process audio, and access music.
3. Improve music distribution by optimizing existing distribution channels or explore new channels.
4. Provide established functionality at lower cost or higher robustness.

The music industry is a diverse industry, consisting of record labels and rights holders, distributors, recording studios, instrument manufacturers, and manufacturers of production and reproduction equipment. Each of the industry branches has been changed radically by music technology during the past 130 years and will continue to change, regardless of the technological innovation being initiated by the industry or reacted upon. What these changes will be, however, is difficult to predict. It is clear that the manufacturing industry will continue to work on new ways to create and modify sound and on new or improved interfaces for creating music. Apart from that, the following trends with the potential to shape parts of the music industry can be observed today:

- The distribution of music shifts from selling albums and songs to licensing rights to subscription services. It appears to be increasingly important to the consumers to access music rather than own it.

- The traditional distribution channels will be threatened by community-based services for sharing music and audio files such as *Soundcloud* or direct-to-fan solutions which allow musicians and independent music labels to sell music and merchandise by bypassing major record labels.
- The borders between professional music production and hobbyists blur. On the one hand, the cost of music production equipment declines; on the other modern tools can perform tasks automatically that previously were understood as craftsmanship. DJ tools, for instance, are able to create mash-ups automatically by beat-matching two or more songs, a process that required practiced skill when performing on vinyl.
- New and improved tools for editing and morphing sounds will be able to improve the *perfection* of recordings. It will be increasingly difficult to differentiate between an unedited and a heavily modified or even synthesized song.
- Online collaboration platforms for the distributed creation of music already exist but have not been overwhelmingly successful in the past. Their approaches and business models, however, are constantly being evaluated and re-thought so that successful services might be launched anytime.
- The potential of using music technology for music education has not been explored in-depth yet. Computer-assisted instruction systems to teach score reading and intonation, to assist vocal and ear training, and to actively support learning a musical instrument effectively and possibly in a game-like environment are only just starting to become more popular.
- The way we find and explore new music will continue to change with the improvement of music recommendation systems, automatic playlist generation algorithms, and general music similarity measures.
- Object-based audio has the potential to completely transform music listening: in contrast to a channel-based system such as stereo or 5.1, object-based audio transmits individual sound elements accompanied by metadata describing how the element should be reproduced. Object-based audio systems are thus independent of the reproduction system configuration and potentially allow an unprecedented interaction between the listener and the mix.

The last century has shown a remarkable evolution in all areas of music technology, and we can expect the future to bring as many exciting changes. However, instead of being overwhelmed by these changes, it is good to keep in mind that, ultimately, music technology serves only one purpose: to explore new or improved ways to create and to access music.

References

44.1 T.A Edison: Improvement in phonograph and speaking machines, US Patent 200521 (1878)
44.2 C.A. Bell, S. Tainter: Recording and reproducing speech and other sounds, US Patent 341214 (1886)
44.3 E. Berliner: Gramophone, US Patent 372786 (1887)
44.4 W. Furtwängler: *Vermächtnis: Nachgelassene Schriften* (Brockhaus, Wiesbaden 1975)
44.5 R. Kostelanetz, J. Darby, M. Santa (Eds.): *Classic Essays on Twentieth-Century Music: A Continuing Symposium* (Schirmer Books, New York 1996)
44.6 G. Gould: The prospects of recording, High Fidel. Mag. **16**(4), 46–63 (1966)
44.7 M. Katz: *Capturing Sound: How Technology Has Changed Music* (Univ. of California Press, Berkeley 2004)
44.8 E.C. Wente: Telephone-transmitter, US Patent 1333744 (1920)
44.9 R. Philip: *Performing Music in the Age of Recording* (Yale Univ. Press, New Haven 2004)
44.10 S. Weinzierl, C. Franke: 'Lotte, ein Schwindel!' – Geschichte und Praxis des Musikschnitts am Beispiel von Beethovens 9. Symphonie. In: *Proc. VDT Int. Audio Convention (22. Tonmeistertagung), Hannover* (2002)
44.11 C. Symes: *Setting the Record Straight: A Material History of Classical Recording* (Wesleyan Univ. Press, Middletown 2004)
44.12 A. Lindau, S. Weinzierl: Assessing the plausibility of virtual acoustic environments. In: *Proc. Forum Acusticum, Aalborg* (2001)
44.13 H.-J. Maempel: Musikaufnahmen als Datenquellen der Interpretationsanalyse. In: *Gemessene Interpretation – Computergestützte Aufführungsanalyse im Kreuzverhör der Disziplinen, Klang und Begriff*, ed. by H. von Lösch, S. Weinzierl (Schott, Mainz 2011) pp. 157–171
44.14 S. Jones: Music and the Internet, Pop. Music **19**(2), 217–230 (2000)
44.15 A. Ross: *The Record Effect* (The New Yorker, New York 2005)
44.16 A. Leyshon: The Software Slump?: Digital music, the democratisation of technology, and the decline of the recording studio sector within the musical economy, Env. Plan. A **41**(6), 1309–1331 (2009)
44.17 J. McCartney: SuperCollider: A new real-time sound synthesis language. In: *Proc. Int. Comput. Music Conf. (ICMC), San Francisco* (1996)
44.18 M.S. Puckette: The Patcher. In: *Proc. Int. Comput. Music Conf. (ICMC), San Francisco* (1988)

44.19 M.S. Puckette: Pure Data: Another integrated computer music environment. In: *Proc. Second Intercollege Computer Music Concerts, Tachikawa* (1997)
44.20 International Conference on New Interfaces for Musical Expression, http://www.nime.org/
44.21 Georgia Tech Center for Music Technology: http://guthman.gatech.edu
44.22 S. Jordà, G. Geiger, M. Alonso, M. Kaltenbrunner: The reacTable: Exploring the synergy between live music performance and tabletop tangible interfaces. In: *Proc. 1st Int. Conf. Tangl. Embed. Interact. (TEI), Munich* (2007)
44.23 A. Clarkson: Lecture on Dada by Stefan Wolpe, The Music. Q. **72**(2), 202–215 (1986)
44.24 MIDI Manufacturers Association: *Complete MIDI 1.0 Detailed Specification V96.1*, 2nd edn. (MMA, La Habra 2001)
44.25 T. Rodgers: On the process and aesthetics of sampling in electronic music production, Organised Sound **8**(03), 313–320 (2004)
44.26 G. Zuckerman: American Mavericks: An Interview with Steve Reich, http://musicmavericks.publicradio.org (2002)
44.27 I. Stravinsky: *An Autobiography*, ebook edn. (Project Gutenberg 2011), transcribed from M.J. Steuer, New York 1958
44.28 A.D. Blumlein: Improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems, US Patent GB394325 (1933)
44.29 J. Nathan: *Sony* (Houghton Mifflin Harcourt, Boston 2001)
44.30 C.K.M. Lam, B.C.Y. Tan: The Internet is changing the music industry, Commun. ACM **44**(8), 62–68 (2001)
44.31 ISO/IEC JTC1/SC29 11172-3: *Information technology – Coding of moving pictures and associated audio for digital storage media up to about 1.5 Mbit/s – Part 3: Audio* (ISO/IEC, Geneva 1993)
44.32 A. Leyshon: On the reproduction of the musical economy after the internet, Media, Cult. Soc. **27**(2), 177–209 (2005)
44.33 W. Coats, V. Feeman, J. Given, H. Rafter: Streaming into the future: Music and video online, L.A. Entertain. Law Rev. **20**(2), 285–308 (2000)
44.34 M. Fox: Technological and social drivers of change in the online music industry, First Monday, special edn. 1 (2005), originally published in February 2002
44.35 Recording Industry Association of America: http://www.riaa.com
44.36 International Federation of the Phonographic Industry: IFPI Digital Music Report 2013 (IFPI, Zurich 2013)
44.37 B.J. Hracs: A creative industry in transition: The rise of digitally driven independent music production, Growth Change **43**(3), 442–461 (2012)

45. Enabling Interactive and Interoperable Semantic Music Applications

Jesús Corral García, Panos Kudumakis, Isabel Barbancho, Lorenzo J. Tardón, Mark Sandler

New interactive music services have emerged, but many of them use proprietary file formats. In order to enable interoperability among these services, the International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) Moving Picture Experts Group (MPEG) issued a new standard, the so-called MPEG-A: Interactive Music Application Format (IM AF).

The purpose of this chapter is to review the IM AF standard and its features, and also to provide a detailed description of the design and implementation of an IM AF codec and its integration into a popular open source analysis, annotation and visualization audio tool known as Sonic Visualiser. This is followed by a discussion highlighting the benefits of their combined features, such as automatic chords or melody extraction time-aligned with the song's lyrics. Furthermore, this integration provides the semantic music research community with a testbed enabling further development and comparison of new Sonic Visualiser plug-ins, e.g., from singing voice-to-text conversion with automatic lyrics highlighting for karaoke applications, to source separation-based music instrument extraction from a mixed song.

The music industry has been experiencing remarkable changes over the past two decades, especially since the popularization of the MP3 standard and the use of the Internet by a large portion of the population. However, such changes are mostly related to more efficient distribution of music content rather than the way users are interacting with the music itself. Despite the widespread use of touchscreen smart phones and tablets that enable interactivity, users of formats like MP3, FLAC or AAC can still only interact with the music they listen to to a limited extent. To address this, interactive music services are emerging to enhance the listener's experience [45.1, 2]. In recent years there have been various interactive services marketed as iKlax [45.3], MOGG [45.4] or MT9 [45.5].

However, such interactive services use their own proprietary formats. In order to ensure interoperability between different interactive music players and interactive songs, a standardized file format is required.

This shortcoming has been overcome by the ISO/IEC Moving Picture Experts Group (MPEG), which has recently released a standard, known as MPEG-A: Interactive Music Application Format (IM AF) [45.6, 7]. This is a multitrack format, so it contains individual audio tracks for different musical instruments. It allows the user to change the mix of the song by changing the volume of each instrument separately.

The novelty of this format over other multitrack formats lies in the ability to add multimedia content to enrich the user experience. For example, it is possible to include text synchronized with audio, which represents the lyrics or the chords of the song.

It may also contain images related to the album or the song. The producer has the ability to define rules to prevent the song becoming unrecognizable. In addition different presets of the song (e.g., a karaoke or rhythmic version) can also be defined.

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_45

Table 45.1 Comparison between different multitrack formats

	iKlax	MT9	MOGG	IM AF
Multitrack	✓	✓	✓	✓
Pictures	✗	✓	✗	✓
Text	✗	✓	✗	✓
Presets	✓	✗	✗	✓
Rules	✓	✗	✗	✓
Groups	✓	✗	✗	✓

Table 45.1 shows a comparison of the characteristics of iKlax, MT9, MOGG and IM AF formats.

This chapter introduces the benefits of adding IM AF support to Sonic Visualiser [45.8]. Sonic Visualiser [45.9] is an open-source application designed to assist semantic audio and signal processing researchers looking at what lies inside an audio file. It has been selected among other music analysis and annotation tools because of its widespread use (10 000 active users in December 2013). An essential strength of Sonic Visualiser is its ability to support third-party plug-ins. The native plug-in format is called VAMP, which enables developers to implement algorithms for extracting descriptive information from audio data. In this context, with respect to IM AF, we will make use of VAMP plug-ins specifically designed for chords [45.10] and melody pitch [45.11].

This chapter is structured as follows: the IM AF standard is described in Sect. 45.1; in Sect. 45.2 the implementation aspects of the IM AF encoder are discussed; Sect. 45.3 introduces a discussion on how Sonic Visualiser with IM AF support can be used with respect to the aforementioned plug-ins; and future developments and conclusions are given in Sect. 45.4.

45.1 IM AF Standard

An IM AF file consists of:

- Multiple audio tracks that represent the different parts of a song (instruments and/or voices)
- Groups of audio tracks that define a hierarchical structure of audio tracks (e.g., all the guitars of a song can be gathered in the same group)
- Presets that contain predefined information about the mixing of multiple audio tracks (for example, a karaoke version of the song or a version with only the rhythmic base)
- Rules that introduce specific data related to the interaction with the user (definition of allowable actions concerning the audio tracks, group selection and volume control)
- Additional multimedia information to enrich the user interaction (text synchronized with audio to represent the lyrics of a song, images related to the song, album or artist)
- Metadata to describe the song, album or artist.

The IM AF standard supports compression of the audio tracks in various formats including PCM, MP3, AAC, and SAOC [45.12]. It also supports the JPEG file format for still pictures [45.13], the 3GPP timed text for lyrics [45.14] and the MPEG-7 Multimedia Description Scheme for metadata [45.15]. The framework of the IM AF file is based on the MPEG-4 ISO Based Media File Format (ISO-BMFF) standard [45.16].

Table 45.2 shows all the supported components in IM AF format.

IM AF files consist of a set of objects, called boxes (Fig. 45.1), which contain all data in the file. We can classify them into two types: general boxes (which may contain other boxes inside them), and FullBoxes, which just contain data (no other boxes allowed inside them).

The File Type Box *ftyp* box should appear at the beginning of the file, as it defines the type of file, declaring the brand compatibility (Table 45.3). The brand is related to the maximum number of audio tracks that can be decoded simultaneously depending on the processing capabilities of the device that will play the IM AF file.

Brands starting with *im0* are intended for mobile phones and portable audio players, while brands start-

Table 45.2 Supported components in IM AF

Type	Component Name	Abbreviation	Specification
File Format	ISO Base Media File Format	ISO-BMFF	ISO/IEC 14496-12:2008
Audio	MPEG-1 Audio Layer III	MP3	ISO/IEC 11172-3:1993
	MPEG-4 Audio AAC profile	AAC	ISO/IEC 14496-3:2005
	MPEG-D SAOC Baseline profile	SAOC	ISO/IEC 2003-2:200x
	PCM	PCM	–
Image	JPEG Image	JPEG	ISO/IEC 10918-1
Text	3GPP Timed Text	3GPP TT	3GPP TS 26.245
Metadata	MPEG-7 Multimedia Description Scheme	MDS	ISO/IEC 15938-5:2003

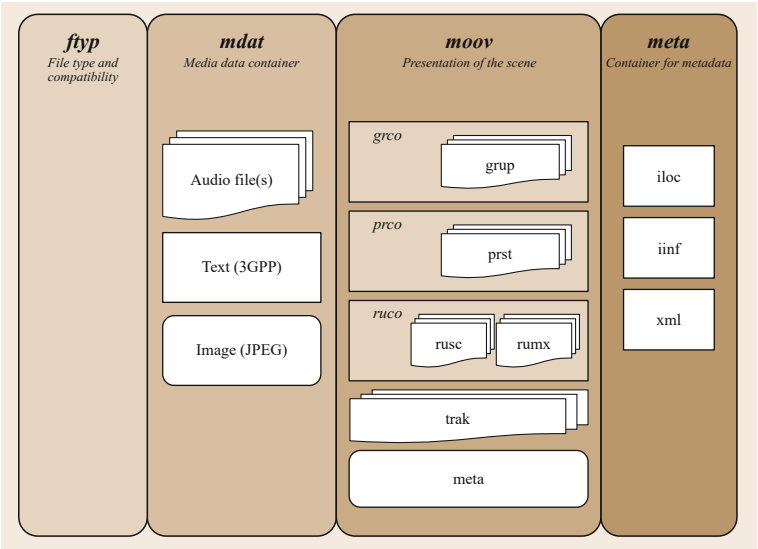


Fig. 45.1 IM AF file format structure

Table 45.3 IM AF brands

Brands	Audio AAC	MP3	SAOC	PCM	Max. number of simultaneously decoded audio tracks	Max. sampling frequency/bits	Profile/level	Application
im01	•	•			4	48 kHz/16 bits	AAC/level 2	Mobile
im02	•	•			6	48 kHz/16 bits	AAC/level 2	Mobile
im03	•	•			8	48 kHz/16 bits	AAC/level 2	Mobile
im04	•	•	•		2	48 kHz/16 bits	AAC/level 2 SAOC baseline/2	Mobile
im11	•	•		•	16	48 kHz/16 bits	AAC/level 2	Normal
im12	•	•	•		2	48 kHz/16 bits	AAC/level 2 SAOC baseline/3	Normal
im21	•			•	32	96 kHz/24 bits	AAC/level 5	High-end

ing with *im1* are intended for general music services. *im21* is the brand for high-end processing applications for professional users [45.2].

The multimedia content (audio, text, images) is stored within the Media Data Box *mdat*.

The information needed for decoding the multimedia content is stored in the Movie Box *moov*. This box can contain one or more Track Boxes *trak*. Each Track Box contains the description of one type of media (audio or text) and it may include a URL indicating where the media content is stored. Thus, IM AF files can be very light in terms of storage requirements. The information on the groups, presets and rules are stored in the Group Container Box *grco*, Preset Container Box *prco* and Rule Container Box *ruco* respectively.

The Metadata Box *meta* includes metadata, such as simple background information for a song (title, singer, album, etc.).

45.2 Implementation of the IM AF Encoder

The encoder is responsible for creating the IM AF files. It has been implemented in C programming language and does not use external libraries. In this way, we can ensure that it can be compiled on any platform. The following sections explain the implementation of the encoder. For further details see [45.17]. Each section addresses a different functionality of the encoder.

45.2.1 Audio Tracks

The File Type Box *ftyp* appears at the beginning of the file and defines the characteristics of the audio tracks contained in the IM AF file. We have implemented the brand *im02*, which allows up to six simultaneous tracks of audio. Thus the IM AF file created can be easily re-

produced in any kind of device without requiring high processing capabilities. The total size of this box is 24 bits.

The next step is to extract the audio samples of each of the MP3 files and store them in the Media Data Box *mdat*. The encoder searches the header of the first frame of the MP3 file, skipping any ID3 tags, and stores the samples byte by byte. Note that the audio tracks are stored one after another in the IM AF file. The player will need information in order to extract these audio tracks contained in the IM AF file. This information is stored in the Sample Table Box *stbl*.

An IM AF file stores a Track Box *trak* for each audio track, containing several tables with the information necessary for the player to play the media data. Those tables are saved in boxes inside the Sample Table Box *stbl*.

The Sample Description Box *stsd* stores the technical specifications of an audio track. It contains information on the type of encoding used, bit rate, sample rate and specific information for decoding. The encoder we developed sets a sample rate of 44 100 Hz and a bit rate of 128 kbps.

The Sample Size Box *stsz* stores the number of frames of the audio track and a table with the number of bytes in each frame. In the case of MP3 files with a constant bit rate of 128 kbps, the frames may be 417 or 418 bytes. The IM AF encoder reads byte by byte the MP3 file, searching all possible headers. When there is a known header, the algorithm counts the number of bytes to find the next header known. Thus, the number of frames and their size are calculated.

The Sample Table Box *stts* contains a table that stores for each entry two values: the number of contiguous frames that have the same length, and the duration of these frames (in milliseconds).

Frames are grouped into chunks. The Sample to Chunk Box *stsc* defines a table that includes the number of frames of each chunk and the number of chunks that have the same number of frames. For constant bit rate MP3s, this table includes only one entry, because all the frames are equally sized.

The Sample Size Box *stsz* stores the number of MP3 frames once more and a table giving the size (in bytes) of each of these audio frames.

Finally, the Chunk Offset Box *stco* indicates the position of the first frame of the audio track in the Media Data Container Box *mdat* of the IM AF file.

Figure 45.2 illustrates the process explained above.

45.2.2 3GPP Timed Text

Another feature of IM AF is that it allows timed text, which may represent the lyrics of the song for karaoke applications. The format of the text track is defined in 3GPP TS 26.245 standard [45.14]. IM AF files store character strings to be displayed, text modifiers that describe how the text should be shown, and temporal information in order to display the text in perfect sync with the music.

Strings and text modifiers are stored in the Media Data Box *mdat*, and information to display the text in sync with the audio is stored in the boxes as *stts*, *stsc*,

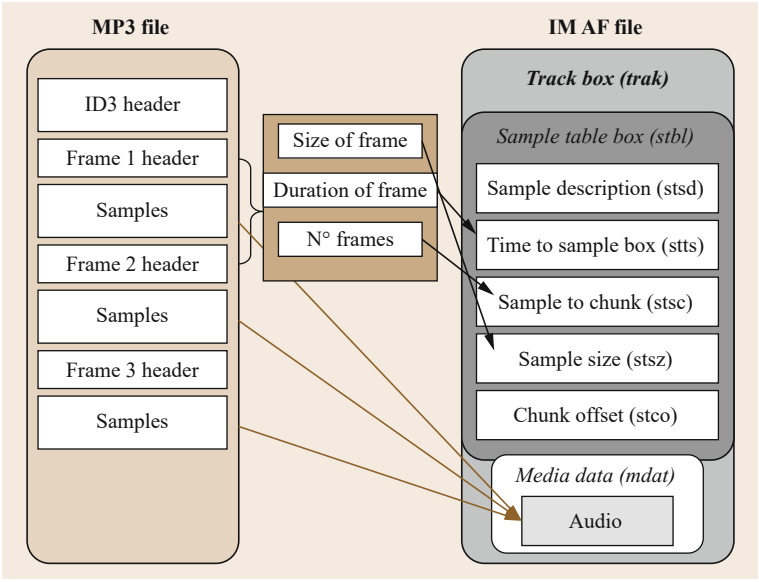


Fig. 45.2 Audio samples and associated data transfer from MP3 to IM AF

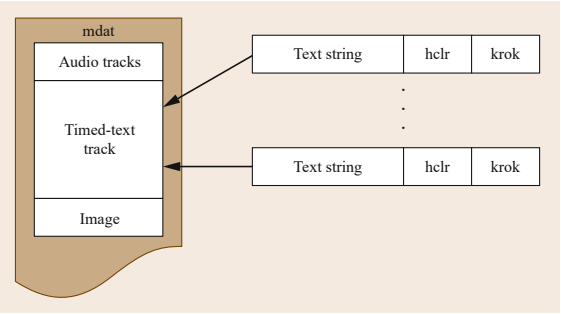


Fig. 45.3 Lyrics encoding

and *stco*. The box *stsd* stores the way in which the text is to be displayed.

Text strings are stored in an external text file with the following syntax,

[start time] Text String [end time] .

The instant of time (in seconds) in which the text string is displayed appears before the text string. The instant of time (in seconds) at which the string is no longer displayed on the screen is stored after the text string.

The encoder seeks every text string inside the text file and it copies them in the *mdat* box of an IM AF file. After a text string is stored, the encoder writes the text modifier boxes. The first modifier box is the Text Highlight Color Box *hclr*, which specifies the highlight color (expressed as hexadecimal RGB color) of the text. The second modifier box is the Text Karaoke Box *krok*. This box specifies the highlighting color start/end time of words in a phrase (Fig. 45.3).

Storage of timing information is similar to audio tracks, as the same boxes are used (*stts*, *stsc*, *stsz* and *stco*, as shown in Fig. 45.2). Duration for every string is saved in the *stts* box (sample delta = duration of interval ∗ timescale). In *stsz* the size of every string is stored including the modifiers. Another important box is *tx3g*, contained inside Sample Description Box (*stsd*), which defines font type and size, horizontal and vertical justification or background color.

45.2.3 Metadata

There are two types of metadata included in the IM AF file:

- Metadata for the song and track
- Metadata for the album.

Table 45.4 shows the boxes containing the various types of metadata.

The *meta* box has two distinct functions: it stores the descriptive metadata (see Table 45.5) and provides information about the JPEG picture included in the IM AF file by locating its length (in bytes) and its offset within the file.

Table 45.4 Levels of metadata in IM AF

Metadata	Location
Track level	Trak/meta box
Song level	Moov/meta box
Album level	Meta box for file

Table 45.5 Descriptive metadata in IM AF

Description	Level		
	Album	Song	Track
Title	●	●	●
Singer	●	●	–
Composer	–	●	–
Lyricist	–	●	–
Performing musician	–	–	●
Genre	●	●	–
File date	●	●	●
CD track number of the song	–	●	–
Production	●	●	–
Publisher	●	●	–
Copyright information	●	●	–
ISRC (International Standard Recording Code)	–	●	–
Image	●	●	–
URL Site address related to the music and the artist (e.g., album homepage, music video)	●	●	–

The container for metadata information is the Meta Box *meta*. This box includes a Handler Box *hdlr* indicating the structure of the contents. The handler type is set to *mp7t*.

The descriptive metadata is located in the XML Box *xml* in XML format. The IM AF encoder simply includes such information in the resulting file by writing the string value. For now, this value can only be changed in the source code.

45.2.4 JPEG Still Pictures

Another feature of the IM AF interactive player is that it can display images while a song is played, by supporting the JPEG file format for still pictures [45.13]. Images can be associated, e.g., to music album's cover and/or artist's photos.

In the box *mdat* the entire content of a JPEG file is stored, while the box *meta* contains the size of the picture and the picture offset within *mdat*.

The first step performed by the encoder is to calculate the size of the image. This task is performed by the getImageSize() function. The image size is used to cal-

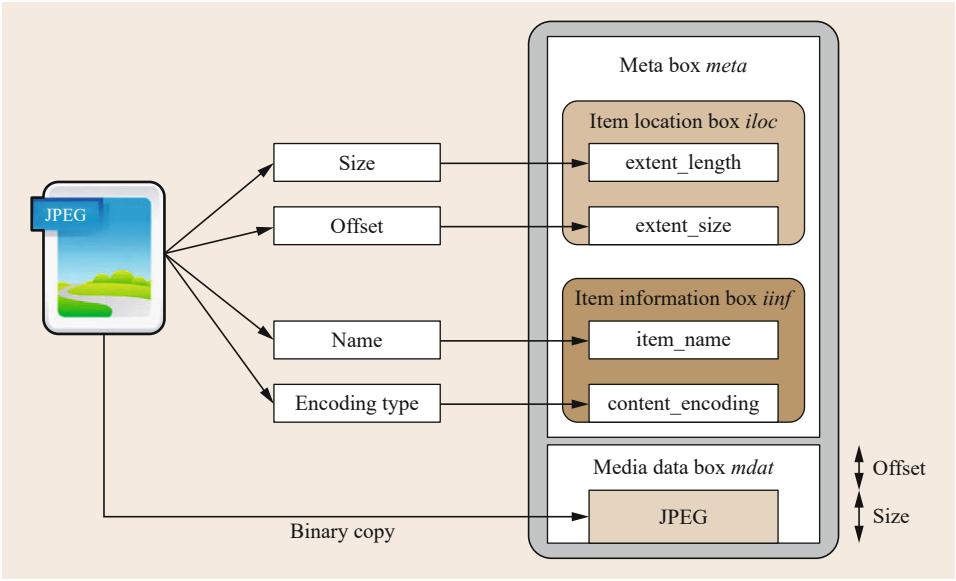


Fig. 45.4 JPEG inclusion process in IM AF

culate the size of the box *mdat*, which is the sum of the size of the audio tracks, the track size of text and the size of the image.

The Item Location Box *iloc* stores the image's size and its offset in *mdat*, in particular into the *extent-length* and *extent-offset* values. The Item Information Box *iinf* contains the picture's name (*item-name*) and its encoding type (*content-encoding*).

The insertImage() function inserts the image in the IM AF file within the box *mdat*.

Figure 45.4 shows the JPEG picture insertion process in an IM AF file during encoding.

45.2.5 Groups

For the creation of groups, two boxes are used: Group Container Box *grco* and Group Box *grup*. The Group Box is located inside the Group Container Box.

The groupcontainer() function is responsible for creating these boxes and returns the size of the Group Container Box.

The Group Container Box stores the number of groups that have been defined inside the IM AF file. The encoder currently allows you to create one group, although the code can be modified easily to allow the creation of multiple groups.

The Group Box contains all other information necessary to define a group. The name assigned to the group is stored in the variable *groupname*. An array stores the identifiers of the audio tracks that make up the group. The variable *group-activation-mode* contains a flag that defines the way the elements of a group shall be activated when a group is switched on. The encoder assigns the value 1 to the variable. This means that the decoder plays all audio tracks contained in the group. The variable *group-reference-volume* specifies the volume gain applied on this group when it is switched on. The variable *flags* stores the value 0x02, so the group information contained in the Group Box is able to display on the player, but it is not able to be edited by the user.

45.2.6 Presets

The IM AF standard defines some default presets. Despite the availability of 11 presets, only six of them are implemented in the reference software. Table 45.6 shows the available presets.

There are two categories of presets: static (fixed volume in each audio track) and dynamic (volume of the tracks can vary over time). There are two boxes involved in presets: Preset Container Box *prco* and Preset Box *prst*. The Preset Box is contained inside the Preset Container Box. The function presetcontainer() creates the presets and returns the size of *prco* box. A switch-case structure selects the implementation, with *preset-type* value as the switching variable. The static presets require the definition of the playback volume gain for each audio track or object, by setting the *preset-volume-element*. The presets that involve objects require the definition of the *output-channel-type*, that is, the number of channels of each included audio track. The encoder sets this value to 1 for stereo tracks.

There is only one working dynamic preset, *Dynamic track approximated volume preset*. This preset is used

Table 45.6 Presets defined in IM AF standard and their availability in reference software

Preset type	Availability	Description
0	✓	Static track volume preset
1	✓	Static object volume preset
2	✓	Dynamic track volume preset
3	✓	Dynamic object volume preset
4	✓	Dynamic track approximated volume preset
5	✗	Dynamic object approximated volume preset
6	–	Value reserved
7	–	Value reserved
8	✓	Static track volume preset with EQ
9	✗	Static object volume preset with EQ
10	✗	Dynamic track volume preset with EQ
11	✗	Dynamic object volume preset with EQ
12	✗	Dynamic track approximated volume preset with EQ
13	✗	Dynamic object approximated volume preset with EQ

to create a fade-in or fade-out effect. The variables *start-sample-number* and *duration-update* indicate the sample where the volume change takes place and the number of samples that it occurs for respectively.

45.2.7 Rules

The function rulescontainer() is responsible for creating rules. It defines the structure of the Rule Container Box *ruco*, which hosts a Selection Rule Box *rusc* and a Mixing Rule Box *rumx*.

Two switch-case structures are responsible for creating the desired Selection and Mixing rules.

Tables 45.7 and 45.8 show the parameters needed by the encoder to create each of the rules.

The *element-ID* is an integer that represents the ID of the element involved in a selection rule. In the case of the Min/Max rule, it must be the same ID as the group. In order to represent the ID of the element on which the rule is applied the encoder uses the *key-element-ID*. The *rule description* is a character string that allows a rule to be named.

Table 45.7 Parameters in Selection Rules

Selection rules						
	Mixing rule type	Element ID	Key element ID	Min num elements	Max num elements	Rule description
Min/Max	0	Group ID	–	✓	✓	✓
Exclusion	1	✓	✓	–	–	✓
Not mute	2	✓	–	–	–	✓
Implication	3	✓	✓	–	–	✓

Table 45.8 Parameters in Mixing Rules

Mixing rules						
	Selection rule type	Element ID	Key element ID	Min volume	Max volume	Rule description
Equivalence	0	✓	✓	–	–	✓
Upper	1	✓	✓	–	–	✓
Lower	2	✓	✓	–	–	✓
Limit	3	✓	–	✓	✓	✓

45.3 IM AF in Sonic Visualiser

Sonic Visualiser (SV) is an application for viewing and analyzing the content of audio files. This open-source software was developed at the Centre for Digital Music at Queen Mary, University of London. The IM AF codec has been fully integrated into Sonic Visualiser. Figure 45.5 shows the interface of the encoder in Sonic Visualiser. The source code as well as executables for both Windows and Mac OS/X of Sonic Visualiser with IM AF support are available through the http://soundsoftware.ac.uk repository [45.18].

The integration of the IM AF codec in SV allows the user to create IM AF files in a quick, convenient and reliable way.

Furthermore, importing an IM AF file into SV opens up and display at once both individual music tracks and mixed versions (presets) of a song. This is a huge relief for SV users that otherwise would have to search their folders and import all of the music audio tracks one by one. That is, SV with IM AF support can be used by music producers as an archival

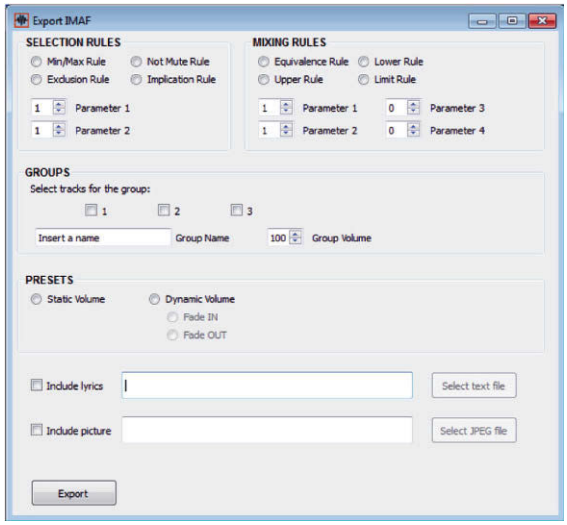


Fig. 45.5 IM AF encoder interface in Sonic Visualiser

tool for saving and importing their music projects at once, thus increasing their time efficiency and productivity.

Furthermore, SV with IM AF support has been converted from an audio analysis and visualization tool to a multimedia editing suite (e.g., multitrack audio, lyrics, pictures, etc.).

Including text synchronized with audio in IM AF offers great potential for the format to be used in conjunction with some SV plug-ins. The possibility of extracting the chords of a song by using the Chordino [45.10] plug-in and presenting them in perfect sync with the song's lyrics can be of great help to those musicians who wish to practice the songs of their favorite artists and improve their musical or vocals skills.

It is equally possible to extract the melody of a song by using the Melodia [45.11] plug-in and display it perfectly aligned with the lyrics.

In order to facilitate interaction with the audio tracks, a panel with gain knobs to control the volume of each of the audio tracks is also included. Subsequent versions of our IM AF codec will allow users to record their own mixes of songs in real time using these gain knobs and share them on social networks.

In the following section some use case scenarios are presented together with instructions on how to use the IM AF codec in Sonic Visualiser. The first subsection is devoted to the steps necessary to create an IM AF file in Sonic Visualiser with certain characteristics. The following two subsections describe the steps required to extract the chords and the melody of a certain musical instrument of an IM AF song.

45.3.1 Creation of an IM AF File

We will show the steps to create an IM AF song with multiple audio tracks, two rules, a fade-in effect, a timed-text track and a picture. A similar process can be used for any combination of features that may be needed:

1. Download and run Sonic Visualiser.
2. File → Open...
3. Select an audio track.
4. To include more audio tracks: File → Import More Audio...
5. Select an audio track. Repeat steps 4) and 5) until you have opened all the desired audio tracks.
6. File → Export IMAF File... At this point the interface of the IM AF encoder pops up.
7. Select *Implication Rule* and choose Track A = 1 and Track B = 2. Thus if Track B is active then track A will also be active. If track A is inactive the track B will also be inactive.
8. Select *Limit Rule*, choose Track A = 1, Min Volume = 10, Max Volume = 20. The minimum volume level of Track A will be 10, and the maximum volume level will be 20.
9. Select *Dynamic Volume* and *Fade IN*. A fade-in effect is a gradual increase in the volume level of an audio track from silence at the beginning.
10. Select *Include lyrics* and press *Select text file*. Choose the text file. The syntax used to create the text file must be the same as shown in Sect. 45.2.2.
11. Select *Include picture* and press *Select JPEG file*. Choose the image file.
12. Press *Export*.
13. Select the name of the IM AF file and the folder. Press *Save*.
14. If the file has been created properly you will see the next message: *IMAF file successfully created!*

45.3.2 Chord Extraction Time-Aligned with Lyrics

One of the possibilities offered by the integration of the IM AF codec in Sonic Visualiser is the potential to extract the chords of a song and present them as text synchronized with the audio (Fig. 45.6). We will use a Sonic Visualiser plug-in for chord extraction, known as Chordino [45.10]:

1. Run Sonic Visualiser.
2. File → Import IMAF File...
3. Select an IM AF file.
4. You will see the waveforms of the audio tracks, a pane with the lyrics and a panel with the gain knobs of the different tracks.

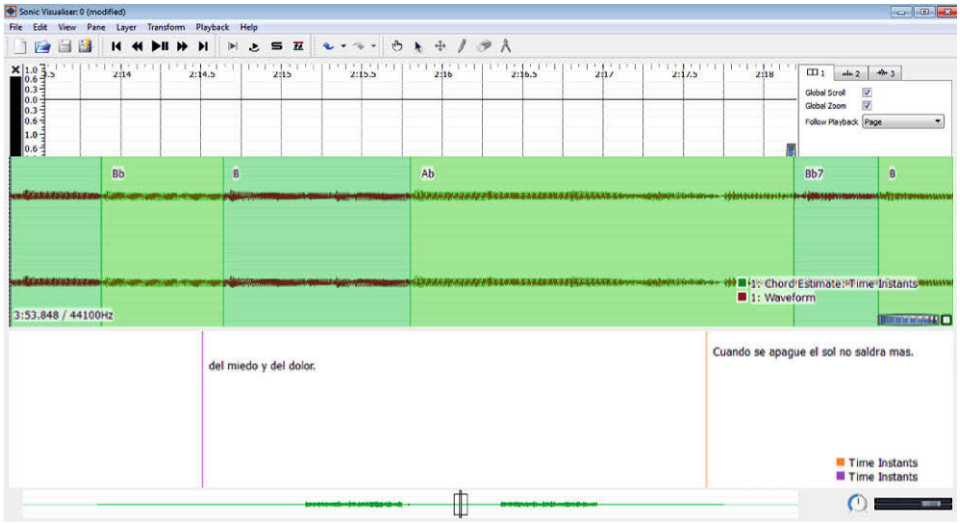


Fig. 45.6 Lyrics aligned in time with chords in Sonic Visualiser with IM AF support

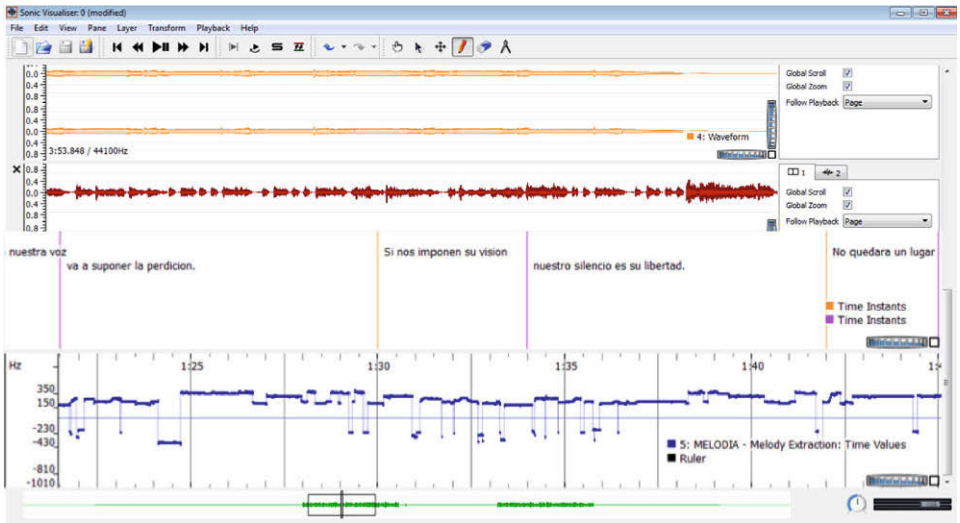


Fig. 45.7 Lyrics aligned in time with melody pitch in Sonic Visualiser with IM AF support

5. Transform → Analysis by Plugin Name → Chordino → Chord Estimate... Select the audio track to apply the Plugin in *Input Material*. Press *Accept*.

45.3.3 Melody Extraction Time-Aligned with Lyrics

Another interesting application of SV is the extraction of the melody pitch of a song. Using the VAMP plug-in called Melodia [45.11], an algorithm for melody pitch estimation, it is possible to gain a more efficient pitch extraction from only the vocal track, rather than the mixed song. Furthermore, the melody pitch is now presented aligned in time with the song's lyrics (Fig. 45.7):

1. Run Sonic Visualiser.
2. File → Import IMAF File...
3. Select an IM AF file.
4. You will see the waveforms of the audio tracks, a pane with the lyrics and a panel with the gain knobs of the different tracks.
5. Pane → Add New Pane.
6. Transform → Analysis by Plugin Name → MELODIA → Melody Extraction... Select the audio track to apply the Plugin in *Input Material*. Press *Accept*.

45.4 Future Developments and Conclusions

The integration of IM AF in Sonic Visualiser opens up new possibilities for researchers in the area of signal processing, offering a testbed for the development of new plug-ins such as:

- On extracting the lyrics from the vocal track (singing to text) [45.19] and automatic highlighting of lyrics, thus relieving IM AF creators from having to insert the lyrics and their time-stamps for highlighting. In particular, the latter task is very tedious, so a plug-in to make this process happen automatically would be extremely useful for karaoke fans.
- On source separation for extracting music instruments from a mix-down song version [45.20].

As has been already mentioned, it would be interesting to enable users to record each tracks' volume while disc-jockeying in Sonic Visualiser and share their own presets with their friends using social networks (e.g., Facebook). Due to the flexibility of the IM AF format, the file does not have to contain the audio tracks themselves but only their URL. In that way the tracks can be recalled from a server. As a result, it enables efficient exchange and sharing of IM AF files, which may include only the mixing parameters.

Furthermore, an alternative cross-platform and cross-browser HTML5 IM AF player [45.21] has also been developed for users that would only like to listen to their friends' mixes with no need to download Sonic Visualiser.

This chapter described the design and implementation of an IM AF codec and its integration in SV as well as highlighting the benefits of their combined features with some use case scenarios.

Acknowledgments. Panos Kudumakis acknowledges that this work was partially done during his visit at the University of Malaga in the context of the program Andalucía TECH: Campus of International Excellence and in conjunction to UK EPSRC project EP/H043101/1 SoundSoftware.ac.uk. This work has been partially funded by the Ministerio de Economía y Competitividad of the Spanish Government under Project No. TIN2016-75866-C3-2-R.

References

45.1 P. Kudumakis: MP3: Something's gotta change!, Audio! **1**(3), 6 (2011)
45.2 I. Jang, P. Kudumakis, M. Sandler, K. Kang: The MPEG interactive music application format standard, IEEE Sig. Process. Mag. **28**(1), 150–154 (2011)
45.3 iKlax Media: http://www.iklaxmusic.com (last accessed 12.01.14)
45.4 MOGG files: Multitrack Digital Audio Format, http://moggfiles.wordpress.com (last accessed 12.01.14)
45.5 MT9: http://en.wikipedia.org/wiki/MT9 (last accessed 12.01.14)
45.6 ISO/IEC 23000-12:2010 – Information technology – Multimedia application format (MPEG-A) – Part 12: Interactive music application format
45.7 ISO/IEC 23000-12:2010/Amd.2:2012 – Information technology – Multimedia application format (MPEG-A) – Part 12: Interactive music application format, AMENDMENT 2: Compact representation of dynamic volume change and audio equalization
45.8 J.C. Garcia, C. Taglialatela, P. Kudumakis, L.J. Tardon, I. Barbancho, M. Sandler: Interactive music applications by MPEG-A support in Sonic Visualiser. In: *AES 53rd Int. Conf. Semant. Audio, London* (2014)
45.9 C. Cannam, C. Landone, M. Sandler: Sonic Visualiser: An open source application for viewing, analysing, and annotating music audio files. In: *Proc. ACM Multimedia Int. Conf.* (2010)
45.10 M. Mauch, S. Dixon: Approximate note transcription for the improved identification of difficult chords. In: *Proc. Int. Symp. Music Inf. Retriev.* (2010) pp. 135–140
45.11 J. Salamon, E. Gómez: Melody extraction from polyphonic music signals using pitch contour characteristics, IEEE Trans. Audio Speech Lang. Proc. **20**(6), 1759–1770 (2012)
45.12 ISO/IEC 23003-2:2010 – Information technology – MPEG audio technologies – Part 2: Spatial Audio Object Coding (SAOC)
45.13 ISO/IEC 10918-1:1994 – Information technology – Digital compression and coding of continuous-tone still images (JPEG)
45.14 ETS 3GPP TS 26.245-2004 – Transparent end-to-end Packet switched Streaming Service (PSS); Timed text format
45.15 ISO/IEC 15938-5:2003 – Information technology – Multimedia content description interface – Part 5: Multimedia description schemes
45.16 ISO/IEC 14496-12:2008 – Information technology – Coding of audio-visual objects – Part 12: ISO base media file format
45.17 C. Taglialatela: *MPEG IM AF encoder: Features development*, BSc Thesis (Seconda Università degli Studi di Napoli, Napoli 2013)

45.18 P. Kudumakis: MPEG developments https://code.soundsoftware.ac.uk/projects/mpegdevelopments (last accessed 12.01.14)
45.19 T. Hosoya, M. Suzuki, A. Ito, S. Makino: Lyrics recognition from a singing voice based on finite state automation for music information retrieval. In: *Proc. Int. Symp. Music Inf. Retriev.* (2005) pp. 532–535
45.20 J. Han, Z. Rafii, B. Pardo: Audio source separation and REPEAT, Research projects of Northwestern University, Dep. of Elec. Eng. and Comp. Sc., http://music.cs.northwestern.edu (last accessed 12.01.14)
45.21 G. Herrero, P. Kudumakis, L.J. Tardon, I. Barbancho, M. Sandler: An HTML5 interactive (MPEG-A IM AF) music player. In: *10th Int. Symp. Comput. Music Multidiscip. Res. (CMMR), Marseille* (2013)

46. Digital Sensing of Musical Instruments

Peter Driessen, George Tzanetakis

Acoustic musical instruments enable very rich and subtle control when used by experienced musicians. Musicology has traditionally focused on analysis of scores and more recently audio recordings. However, most music from around the world is not notated, and many nuances of music performance are hard to recover from audio recordings. In this chapter, we describe hyperinstruments, i. e., acoustic instruments that are augmented with digital sensors for capturing performance information and in some cases offering additional playing possibilities. Direct sensors are integrated onto the physical instrument, possibly requiring modifications. Indirect sensors such as cameras and microphones can be used to analyze performer gestures without requiring modifications to the instrument. We describe some representative case studies of hyperinstruments from our own research as well as some representative case studies of the types of musicological analysis one can perform using this approach, such as performer identification, microtiming analysis, and transcription. Until recently, hyperinstruments were mostly used for electroacoustic music creation, but we believe they have a lot of potential in systematic musicological applications involving music performance analysis.

46.1 Digital Music Instruments

Traditional musical instruments are some of the most fascinating artifacts created by human beings throughout history and across cultures. The complexity and richness of control afforded by acoustic musical instruments (such as a cello or saxophone) to professional musicians is impressive and takes a lifetime of practice to develop. Musicology has traditionally focused on musical scores as the primary representation used to analyze music, despite the fact that the majority of music throughout history has not been notated. In addition to musical scores, more recently audio recordings in conjunction with signal processing techniques such as spectrography have also been used in musicology and especially computational ethnomusicology [46.1]. An audio recording captures information about the end result of making music but makes it hard to analyze the actual process or cultural aspects of making music.

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_46

Over the past 20 years, there has been increasing interest in developing new interfaces for musical expression using actuators, computers, and sensors. The main venue for this type of work is the New Interfaces for Musical Expression (NIME) conference [46.2], http://www.nime.org/archive/. Some of these new instruments are completely novel contraptions, while others either draw inspiration from or digitally augment existing acoustic instruments. The term *hyperinstrument* [46.3] has been used to refer to an acoustic musical instrument that can be played conventionally but has also been augmented with various sensors to transmit information about what is being played. The commonest use of hyperinstruments has been in the context of live electroacoustic music performance, where they combine the wide variety of control possibilities of digital instruments such as Musical Instrument Digital Interface (MIDI) keyboards with the expressive richness of acoustic instruments. A less explored, but more interesting from a musicological perspective, application of hyperinstruments is performance analysis. The best-known example is the use of acoustic pianos fitted with a robotic sensing and actuation system on the keys to capture the exact details of the player's actions and replicate them. Such systems allow the exact nuances of a particular piano performer to be captured so that, when played back on the same acoustic piano with mechanical actuation, they will sound identical to the original performance. The captured information can be used to analyze specific characteristics of the music performance, such as how the timing of different sections varies among different performers [46.4, 5].

46.2 Elements of a Hyperinstrument

In this section, we describe the elements of a hyperinstrument in a general way that is common to all such instruments. A hyperinstrument may be defined as an acoustic musical instrument that can be played in the conventional way for that instrument but has also been augmented with sensors to transmit information about what is being played and, in some cases, actuators that create sound.

At this point, some readers may prefer to skip this section and read ahead to the example instrument and application case studies of specific hyperinstruments. These readers can return later to this section, which provides a systematic but somewhat abstract description of the elements of a hyperinstrument and its performance modes.

In this section, we first describe the elements of a conventional acoustic instrument, followed by the additional elements that are needed for a hyperinstrument. The hyperinstrument can include direct and indirect (surrogate) sensors and elements for sonic and visual display, and a computer means of interpreting the sensor data and generating appropriate signals to directly or indirectly control the display elements. The hyperinstrument also has performance modes using both event-based and continuous control and playing scenarios for very slow, slow, and fast playing. In the following subsections, each of these hyperinstrument elements and the performance modes are described.

46.3 Acoustic Instrument

A hyperinstrument includes first the elements of a conventional acoustic instrument (string [46.6], wind [46.7, 8], brass [46.9], percussion [46.10], keyboard [46.11]):

- The physical acoustic instrument per se, including physical elements made from raw materials such as wood, brass, gut (for strings), and membranes (for percussion). This physical object (instrument) includes two further essential design elements
- A sensor (or interface) that accepts a gestural input from a human performer to actuate and control some of the physical elements
- An acoustic structure using these physical elements that can produce sound when actuated (moved).

These two design elements may rely on the same physical element; For example, a guitar string or drum membrane is an interface that accepts a pluck or tap gesture from a finger and also comprises an acoustic structure that produces sound.

These two design elements may also use different physical elements. A woodwind or brass instrument accepts the airflow and embouchure tension gestures from the player on the mouthpiece that actuate the reed or the vibration of the lips that produces the sound.

For some instruments, the sensor accepts gestural input directly from the performer's body (hands or mouth). For other instruments, the sensor (or interface) is a separate component (bow for some string instruments, stick or mallet for some percussion instruments, piano action mechanism) that accepts gestural inputs from the performer. For these instruments, the gestural input from the performer only indirectly actuates the physical element that produces the sound.

For most acoustic instruments, the acoustic structure also includes an acoustic amplification mechanism (partially enclosed volume or resonant cavity) by which the sound arising from the human gesture is amplified, for example, guitar or violin body, drum shell, main body of woodwind or brass instrument, or piano soundboard.

For some acoustic instruments, the amplification mechanism may include a separate interface that accepts gestural inputs from the hands to depress keys or valves that modify the resonance properties of the enclosed volume, thus changing the pitch. This interface may be designed to accept discrete inputs (key or valve up or down) but may create idiosyncratic sounds when not fully up or down.

46.4 Hyperinstrument

In addition to the above elements of a conventional acoustic instrument, a hyperinstrument includes one or more additional elements:

- Electronic and/or mechanical sensors that detect the gestural input of the human performer and change their electrical properties or emit an electronic signal in accordance with that input, including integrated sensors such as a Wii remote or tablet (referred to as *direct* sensors)
- A means of converting the sensor output to a form readable by computer
- A computer program or other means to interpret the sensor data and then map this data to control signals needed to control the following elements:
- Elements that create a sonic and visual display to supplement the performer's own display:
  - Mechanical machinery such as motors and actuators that can actuate the physical elements on the instrument or on a second separate (robotic) instrument
  - Electrical transducers such as loudspeakers, video screens, and projectors that can reproduce sound, music, images, and video
  - A computer program that uses the interpreted sensor data to cause the mechanical and/or electrical elements to change in some way to create both sound and a visual display.

A hyperinstrument may be inspired by but built without any of the elements of a conventional acoustic instrument. In this case, we may label it as a new interface for musical expression. In this chapter, we focus on the idea of a conventional instrument that is augmented as described above.

In the following subsections, we describe each of these elements in turn.

46.5 Direct Sensors

A direct sensor is added to a conventional acoustic instrument to make a hyperinstrument. Direct sensors accept gestural input from the performer and change an electronic property in a way that is related to that input. Direct sensors may be classified along several dimensions:

- Gestural input source: fingers, hands, feet, head, arms, legs, torso, or other
- Nature of gestural input that creates the sensor response: impact, pressure, position, velocity, or acceleration
- Sensor electronic properties in response to input:
  - Passive analog (change in resistance, capacitance, and/or inductance)
  - Active analog (change in output voltage or current)
  - Active digital (change in output data stream), where the data stream may be in MIDI or open sound control (OSC) format [46.12] but also in video format if the sensor is a digital camera
  - Integrated (smartphone with touchscreen and multiple embedded sensors, motion capture

system with cameras, markers, and software).
- Discrete amplitude or continuous amplitude sensor output in one, two, or three dimensions
- Discrete time event-based (sequence of asynchronous time-stamped events) output or continuous time sensor output (or sequence of numbers at constant sample rate) in one or more dimensions
- With or without tactile feedback.

These sensors require a means of converting the sensor output signal to a form readable by computer. This may be an analog electronic circuit that detects changes in resistance, capacitance, inductance, voltage or current, followed by an analog to digital converter (ADC), or a digital circuit that reads the output data stream. A commonly employed means for both cases is an Arduino or similar electronic prototyping platform. The integrated sensors include an internal computer that reads the data.

The resulting computer-readable sensor data may be represented or interpreted as a waveform (signal) showing amplitude versus time for each dimension. This sensor data has five important characteristics:

- Amplitude dynamic range (maximum signal to noise ratio)
- Amplitude precision
- Time resolution
- Time precision
- Latency (delay).

The amplitude dynamic range of a sensor with continuous output is the (logarithm of the) ratio between the strongest and weakest output signal. The dynamic range is equivalent to the maximum signal to noise ratio, where the weakest signal is barely perceptible above the noise. Ideally, the dynamic range is sufficient to allow the performer to fully utilize the dynamic range available in the aforementioned mechanical machinery and electrical transducers that create the sound and visual display. In the best case, the dynamic range will match that of the relevant human perceptual system: 140 dB or 14 orders of magnitude for the auditory system, or 5 orders of magnitude static and 9 orders with adaption for the human visual system. The amplitude precision will determine the number of distinguishable amplitude levels within the dynamic range limits. If the gestural input is position, then the amplitude resolution and precision correspond to the spatial resolution and precision.

The time resolution of an analog sensor is determined by the ADC sampling rate. For digital sensors, the time resolution may be inherent in the design (e.g., the frame rate of a camera). The time resolution should ideally be sufficiently high to resolve rapid gestures (up to 160 km/h or 44 m/s for the hand in baseball and cricket, higher for drumstick tip), as well as subtle variations in gesture. In other words, the sampling rate must be at least double the highest spatial frequency in the gesture. The time precision determines the number of distinguishable time intervals. Time precision may also be affected by jitter in the sampling clock. Latency or delay may be caused by hysteresis in the sensor or delayed response to tactile feedback. Additional latency may be caused by the means of converting sensor data to computer-readable form, e.g., memory buffers.

In this section, we consider some examples of direct sensors:

- Switch
- Force-sensitive resistor
- Accelerometer
- Integrated position sensor
- Camera
- Radio drum
- Sound plane.

46.5.1 Switch

A switch is perhaps the simplest sensor, accepting input from any part of the performer's body that can push or touch it. It responds to pressure (light or heavy) or touch, and changes its resistance from infinity (open circuit) to zero (closed circuit). It is a discrete sensor in one dimension with only two possible states (on or off), and may have tactile feedback (mechanical switch) or not (touch or heat sensitive). The switch may be a simple pushbutton, or one key on an organ or harpsichord or qwerty keyboard. Multiple switches may be considered as having outputs in multiple dimensions.

46.5.2 Force-Sensitive Resistor

A force-sensitive resistor is well described by its name. The input source is typically a finger. It responds to pressure (force) by changing its resistance continuously over a range of values. It may or may not have tactile feedback, depending on the surface on which it is mounted.

46.5.3 Accelerometer

An accelerometer responds to the rate of change of velocity in one, two or three dimensions from any gestural input source. Typically, the output voltage changes in

step with the acceleration, has a continuous output in one dimension, and no tactile feedback.

46.5.4 Contact Microphone

A contact microphone or piezo senses audio vibrations through solid objects. When mounted on the tip of a drumstick, it detects the instant of a strike when the stick vibrates. The piezo responds to an impact and has no output otherwise, producing an active analog output. The output at the moment of impact is continuous in amplitude but discrete in time and has tactile feedback.

46.5.5 Integrated Inertial Sensor

An integrated inertial sensor comprising an accelerometer, gyroscope, and magnetometer accepts any gestural input source, responds to changes in position, has multiple electronic signal outputs, and has continuous output in three dimensions with no tactile feedback. These sensors are also used in inertial motion capture systems.

46.5.6 Camera

A camera accepts any gestural input source (the entire body), responds to position, is an active digital sensor (makes changes to the output video data stream), is a continuous sensor in two dimensions (although a third dimension can be inferred), and has no tactile feedback. The Microsoft Kinect is a specialized camera that uses structured light to infer a third (depth) dimension. An optical motion capture system is another specialized camera system that detects markers on the body.

46.5.7 Key or Valve Position Sensor

A trumpet valve or piano key position sensor accepts gestural input from the fingers and responds to the position of the key or valve, ranging from fully up to fully down. In the case of a trumpet valve position sensor, the optical sensor is active analog (measuring the time delay of an infrared laser pulse reflected inside the valve chamber). The sensor is continuous in one dimension for each valve. In the case of a piano key position sensor, an optical sensor may be used also, in contrast to the traditional velocity-sensitive piano key that measures the time delay between two positions, almost fully up and almost fully down. The valve or key provides tactile feedback that can be adjusted via the amount of spring tension (for trumpet valves) or the mechanical design of the action (for piano keys).

46.6 Indirect or Surrogate Sensors

Some hyperinstruments may also include sensors that do not accept gestural input from the performer (referred to as *indirect* sensors):

- Auditory sensors (microphones, vibration sensors) that detect the sound emanated by the instrument
- Sensors that observe the environment and its parameters and generate data, including temperature, air pressure, light sources and their intensities, air motion speed and direction
- Sensors that detect audience interaction, i. e., movement, gestures, expressions, and sounds; these may be simple infrared or ultrasound motion sensors, or cameras
- A computer program to interpret the sensor data, causing the mechanical and/or electrical elements to change in some way.

An indirect sensor does not accept gestural inputs. Instead, it senses some attribute of the environment, including sound and vibration emanated by the performer's acoustic instrument, as well as environmental (weather) parameters. As for direct sensors, the sensor data has five important characteristics:

- Amplitude dynamic range (maximum signal to noise ratio)
- Amplitude precision
- Time resolution
- Time precision
- Latency (delay).

A surrogate sensor uses a direct sensor to train a machine learning model that can then be used with an indirect sensor [46.13].

46.6.1 Microphone

A microphone is a very useful indirect sensor in that its output can be interpreted by a computer program to identify specific characteristics of the acoustic instrument sound, such as pitch, spectrum (related to timbre), loudness (related to sound pressure level and timbre), dynamics (attack and release times), and rhythmic patterns. The microphone as a system includes

a preamplifier, and an ADC so that the data can be read by computer. The computer software uses digital signal processing techniques taken from the literature of music information retrieval. The microphone sensor is a surrogate sensor in that it can detect pitch without using the gestural data from the performer's fingers depressing keys, valves or strings on a fingerboard. As a surrogate sensor, the microphone can also detect loudness without directly sensing the gestural data from the performer's airflow or piano key velocity.

46.6.2 Eddy-Current Sensor

An eddy-current sensor detects the vibrations in a metal bar of a vibraphone via a pickup coil placed close to but not touching it. The bar moves closer and further away from the coil, thus changing its inductance. The coil is tuned with a capacitor to make an *RLC* resonant circuit (bandpass filter) at a frequency near 1 MHz. A signal is injected offset from this resonant frequency on the slope (skirt) of the bandpass filter. Changes in the resonant frequency then cause changes in the amplitude of the injected signal. A separate eddy-current sensor is placed underneath each bar of the vibraphone and can pick up the vibrations of each bar independently without crosstalk from other nearby vibrating bars.

The eddy-current sensor is listed in this section as an indirect sensor, since it does not respond directly to gestural input. It may be viewed as a surrogate sensor for the bar itself, which is a direct sensor of an acoustic instrument as defined above, similar to the guitar string or drum membrane. However, it does respond when a bar is struck, just as a guitar string responds to a pluck from a finger, and thus may also be viewed as a direct sensor that responds to the hand gesture via the mallet, responds to the velocity of the mallet, is active analog with continuous amplitude output in one dimension, and offers tactile feedback to the hand via the mallet only at the moment of time when struck but not other times.

46.7 Instrument Case Studies

In this section, we describe some representative case studies of digital sensing of acoustic musical instruments drawn from our own research, as we have more detailed information about these compared with systems developed by other groups and they illustrate different design considerations.

46.7.1 E-Sitar

The sitar is a 19-stringed, pumpkin-shelled, traditional north Indian instrument. Its bulbous gourd (Fig. 46.1), cut flat on the top, is joined to a long-necked hollowed concave stem that extends three feet long and three inches wide. The sitar contains seven strings on the upper bridge, and 12 sympathetic strings below. All strings can be tuned using tuning pegs. The upper strings include rhythm and drone strings, known as chikari. Melodies, which are primarily performed on the uppermost string and occasionally the second copper string, induce sympathetic resonances in the 12 strings below. The sitar can have up to 22 movable frets, tuned to the notes of a raga (the melodic mode, scale, order, and rules of a particular piece of Indian classical music) [46.14].

It is important to understand the traditional playing style of the sitar to comprehend how the controller captures its hand gestures. The controller design has been informed by the needs and constraints of the long tradition and practice of sitar playing. The sitar player uses his left index finger and middle finger to press the string to the fret to play the desired swara (note). The frets are elliptically curved so the string can be pulled downward, to bend to a higher note. This is how a performer incorporates the use of shruti (microtones), which is an essential characteristic of traditional classical Indian music. On the right index finger, a sitar player wears a ring-like plectrum, known as a *mizrab*. The thumb of the right hand remains securely on the edge of the dand (neck) as the entire right hand gets pulled up and down over the seven main strings, letting the *mizrab* strum the desired melody. An upward stroke is known as Dha, and a downward stroke is known as Ra [46.14]. These two

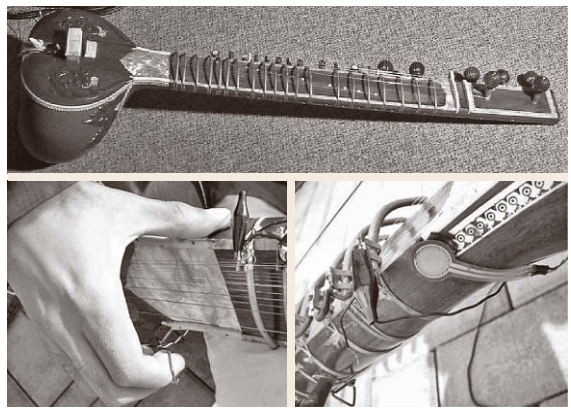


Fig. 46.1 E-Sitar and thumb sensor

main gestures are captured using sensors, and audio-based analysis is then used to model the following:

1. Pitch/fret position
2. *Mizrab* stroke direction.

The E-Sitar [46.15] was built with the goal of capturing a variety of gestural input data. A variety of different sensors, such as fret detection using a network of resistors, are used, combined with an Atmel AVR ATMega16 microcontroller for data acquisition. Fret detection is achieved using a network of resistors attached in series to each fret on the E-Sitar. A voltage is applied to the string, and a connection is established when the string is pressed down onto a fret. This results in a unique voltage based on the amount of resistance in series up to that fret. The voltage is then calculated and transmitted using the MIDI protocol.

The direct sensor used to deduce the direction of a *mizrab* stroke is a force-sensing resistor (FSR), which is placed directly under the thumb of the right hand (Fig. 46.1). The thumb never moves from this position while playing, but the force applied varies based on the *mizrab* stroke direction. A *Dha* stroke (upward stroke) produces more pressure on the thumb than a *Ra* stroke (downward stroke). A continuous stream of data is sent from the FSR via MIDI, because this data is rhythmical in time and can be used compositionally for more than just deducing the pluck direction.

46.7.2 EROSS

One of the problems with many hyperinstruments is that they require extensive modifications to the actual acoustic instrument in order to install the sensing apparatus. The Easily Removable, wireless Optical Sensor System (EROSS) [46.16] can be used with with any conventional piston valve acoustic trumpet (Fig. 46.2). Optical sensors are used to track the displacement of the three trumpet valves continuously. These values are transmitted wirelessly to a host computer system. The hardware has been designed to be reconfigurable by using three-dimensional (3-D) printing to form the housing, with dimensions that can be adjusted for any particular model of trumpet.

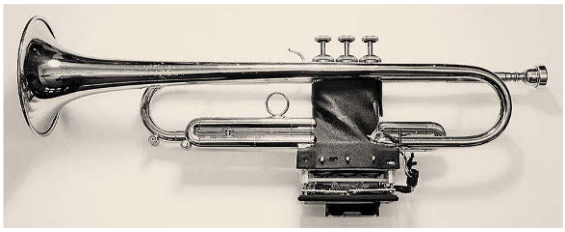


Fig. 46.2 EROSS mounted on a trumpet

46.7.3 Radio Drum

The radio drum (Fig. 46.3) [46.10, 17] is essentially a three-dimensional sensor that detects the *xyz* position of a drumstick tip above a planar surface with high temporal accuracy. The radio drum responds to gestural input from the hands indirectly via the drumsticks. The nature of the gestural input is the position of the drumstick tip. The radio drum response to this input is an active analog sensor with six continuous analog signal waveform outputs, one for each dimension for each of two sticks. The radio drum provides tactile feedback only when striking or dragging along the surface but not when moving the sticks above the surface. The amplitude dynamic range is about 30 dB, while the time resolution is determined by the ADC sampling rate.

Fig. 46.3 Radio drum

46.8 Application Case Studies

In this section, we describe some examples of how data acquired through digital sensing can be used for different types of music performance analysis. These case studies illustrate the potential of digital sensing of musical instruments for musicological research that would be difficult or even impossible to carry out based on scores or audio recordings.

46.8.1 Sitar Transcription

Automatic music transcription is a well-researched area [46.18] and is typically based on analyzing an audio recording. The novelty of this work is that it looks beyond the audio data by using sensors to avoid octave errors and problems caused by polyphonic transcription. In addition, it does not share the bias of most research that focuses only on Western music. The sitar is a fretted stringed instrument from north India. Unlike many Western fretted stringed instruments (classical guitar, viola de gamba, etc.) sitar performers pull (or bend) their strings to produce higher pitches. In normal performance, the bending of a string will produce notes as much as a fifth higher than the same fret position played without bending. In addition to simply showing which notes are audible, the system also provides information about how to produce such notes. A musician working from an audio recording (or transcription of an audio recording) alone will need to determine which fret they should begin pulling from. This can be challenging for a skilled performer, let alone a beginner. By representing the fret information on sheet music, sitar musicians may overcome these problems. The E-Sitar was the hyperinstrument used for these experiments. Automatic pitch detection using an autocorrelation-based approach was utilized with adaptive constraints (minimum and maximum pitch) based on the sensor fret data from the E-Sitar. To compensate for noisy fret data, median filtering in time was applied. To obtain an accurate final result, pitch information from the audio signal chain is fused with onset and pitch boundaries calculated from the fret signal chain. The fret provides convenient lower and upper bounds on the pitch: a note cannot be lower than the fret, nor higher than a fifth (i. e., seven MIDI notes) above the fret. Using the note boundaries derived from the fret data, the median value of the pitches (inside the boundaries supplied by the fret data) are found. These are represented by the vertical lines in Fig. 46.4, and are the note pitches in the final output.

46.8.2 Microtiming in Afro-Cuban Music

Using a percussion sensing apparatus such as the radio drum, it is possible to extract onsets of rhythmic events in rhythmically complicated music such as Afro-Cuban music based on the *clave* pattern. A special domain-specific beat-tracking technique called rotation-aware dynamic programming has been developed for this challenging scenario [46.19]. The output of either direct sensing or beat tracking based on the indirect audio signal provides the time locations of clave *notes*. A performance typically consists of about 625–1000 such clave notes. Simply plotting each point along a linear time axis would either require excessive width, or would make the figure too small to see anything; this motivates bar wrapping. Conceptually, one starts by

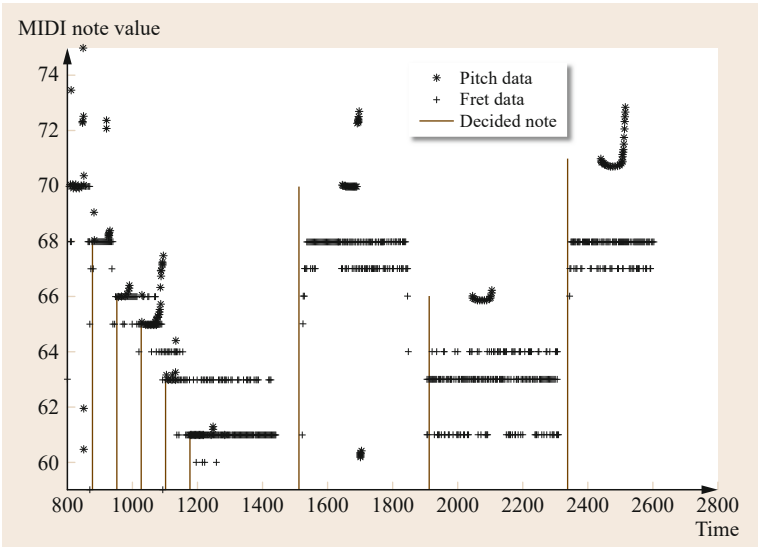


Fig. 46.4 Fret data, audio pitches, and the resulting detected notes. The final three notes were pulled

marking each event time on a linear time axis. If one imagines this time axis as a strip of magnetic tape holding our recording, then metaphorically we cut the tape just before each downbeat, so that we have 200 short pieces of tape, which we then stack vertically, so that time reads from left to right along each row, then down to the next row, like text in languages such as English. Each of these strips is then stretched horizontally to fill the figure width, adding a tempo curve along the right side to show the original duration of each bar. Figure 46.5 depicts the times of our detected onsets for for a music recording of afro-cuban music with this technique. The straight lines show the theoretical clave locations. Close inspection of this figure reveals that the fifth clave note is consistently slightly later than its theoretical location. This would be hard to notice without precise estimation of the tempo curve.

46.8.3 Performance Analysis of Kazakh Dombra

Wright [46.20] used a computer to compare the spectral content and note timings for two recorded performances of a composition for the Kazakh dombra, a long-necked, fretted, two-stringed plucked lute. He implemented a simple custom beat-tracking algorithm optimized for the extremely short-term tempo changes of this style: given hand-marked note onset times for both performances, it produces continuous curves showing local tempo as a function of time, as shown in Fig. 46.6. Comparison of these curves indicates that both performers varied the tempo in much the same way at specific points in the composition (supporting the idea that specific tempo modulations are an intrinsic part of this musical style and this particular piece), while also showing stylistic differences between the two musicians; For example, though both performances tended to have bumps in the tempo curve (i. e., brief changes in local tempo followed by a return to the original speed) at the same places in the piece, the younger musician tended to spend more time at the second tempo before returning to the first, while the older musician tended to return more quickly to the original tempo. This difference is very apparent for the sudden slowing down near the 45-second point in Fig. 46.6.

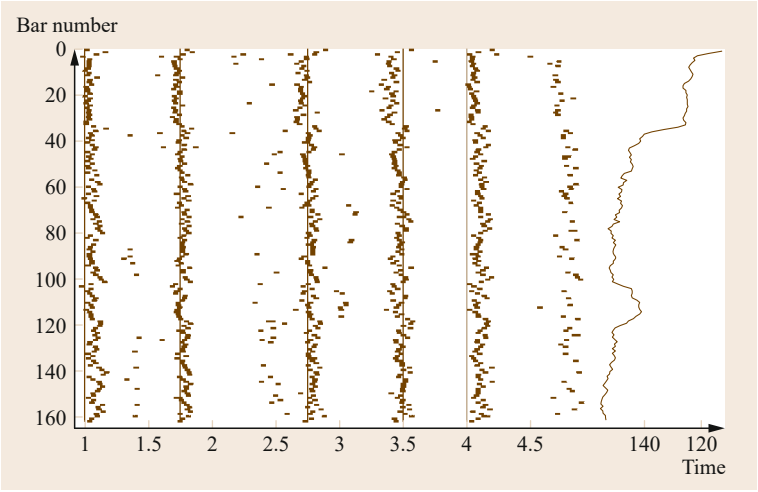


Fig. 46.5 Display of micro-timing in afro-cuban clave showing automatically detected onsets and theoretical clave locations by *straight lines*, normalizing tempo variations

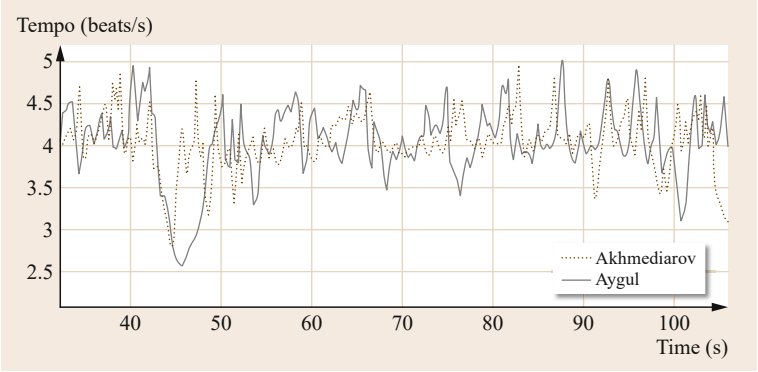


Fig. 46.6 Instantaneous tempo curves of two performances of a composition for the Kazakh dombra (time-stretched for alignment, as they have different durations)

46.8.4 Multimodal Musician Recognition

An interesting analysis task that trained humans are able to do is to recognize a particular musician from the sound of their playing, even if they are performing the same piece. Using a combination of sensor data from the E-Sitar as well as audio features, it has been shown that accurate automatic identification of the musician playing is possible [46.9, 21]. The sensor data and audio features are used to train machine learning classifiers for the performer identification. Three datasets corresponding to exercises, a composition, and an improvisation were used for the identification. Table 46.1 presents the classification accuracy results for different classifiers and datasets (MLP is a multilayer perceptron, sequential minimal optimization (SMO) is a linear support vector machine, and NB is a naive Bayes classifier).

Table 46.1 Classification accuracy for musician identification

	Exercise	Practice	Improv.	Combined
MLP	100	100	100	100
SMO	97	100	93	86
NB	85	100	94	67

46.9 Conclusions

Digital sensing technology can be used in several ways to extract performance information from musical instruments. They can be classified into direct sensing methods, in which sensors are attached to the instrument, and indirect sensing, in which sensors external to the instrument such as cameras and microphones are utilized. Some representative examples of particular hyperinstruments and a high-level overview of their corresponding sensing apparatus were also presented. The last section of the chapter describes some application case studies that illustrate the potential of digital sensing for various types of analysis such as microtiming, performer identification, and transcription. Digital sensing of musical instruments is a relatively new research area and has mostly focused on music creation rather than analysis, so its use in musicology is still in its infancy. We believe that such technologies have great potential for extraction and analysis of interesting music performance information for the enormous diversity of musical instruments from around the world.

References

46.1 G. Tzanetakis, A. Kapur, W.A. Schloss, M. Wright: Computational ethnomusicology, J. Interdiscipl. Music Stud. **1**(2), 1–24 (2007)

46.2 E.R. Miranda, M.M. Wanderley: *New Digital Musical Instruments: Control and Interaction Beyond the Keyboard*, Vol. 21 (AR, Middleton 2006)

46.3 T. Machover: *Hyperinstruments: A Progress Report, 1987–1991* (MIT Media Laboratory, Cambridge 1992)

46.4 W. Goebl, R. Bresin, A. Galembo: Touch and temporal behavior of grand piano actions, J. Acoust. Soc. Am. **118**(2), 1154–1165 (2005)

46.5 M. Bernays, C. Traube: Expressive production of piano timbre: Touch and playing techniques for timbre control in piano performance. In: *Proc. Sound Music Comput. Conf.* (2013)

46.6 T. Machover: Hyperinstruments: A composer's approach to the evolution of intelligent musical instruments. In: *Cyberarts: Exploring Arts and Technology*, ed. by L. Jacobson (Miller Freeman, San Francisco 1992) pp. 67–76

46.7 M. Burtner: The metasaxophone: Concept, implementation, and mapping strategies for a new computer music instrument, Organ. Sound **7**(2), 201–213 (2002)

46.8 C. Palacio-Quintin: The hyper-flute. In: *Proc. New Interfaces Music. Expr. (NIME) Conf.* (2003)

46.9 P. Cook, D. Morrill, J.O. Smith: A MIDI control and performance system for brass instruments. In: *Proc. Int. Comput. Music Conf. (ICMC)* (1993)

46.10 A.R. Tindale, A. Kapur, G. Tzanetakis, P. Driessen, A. Schloss: A comparison of sensor strategies for capturing percussive gestures. In: *Proc. Conf. New Interfaces Music. Expr.* (2005)

46.11 A. McPherson: TouchKeys: Capacitive multi-touch sensing on a physical keyboard. In: *Proc. New Interfaces Music. Expr. (NIME) Conf.* (2012)

46.12 M. Wright: Open sound control-a new protocol for communicating with sound synthesizers. In: *Proc. Int. Comput. Music Conf. (ICMC)* (1997)

46.13 A. Tindale, A. Kapur, G. Tzanetakis: Training surrogate sensors in musical gesture acquisition systems, IEEE Trans. Multimed. **13**(1), 50–59 (2011)

46.14 S. Bagchee: *Understanding Raga Music* (BPI, New Delhi 1998)

46.15 A. Kapur, A. Lazier, P. Davidson, R.S. Wilson, P. Cook: The electronic sitar controller. In: *Proc. New Interfaces Music. Expr. (NIME) Conf.* (2004)

46.16 L. Jenkins, W. Page, S. Trail, G. Tzanetakis, P. Driessen: An easily removable, wireless optical

sensing system (EROSS) for the trumpet. In: *Proc. Int. Conf. New Interfaces Music. Expr. (NIME)* (2013)
46.17 W.A. Schloss: Recent advances in the coupling of the language Max with the Mathews/Boie radio drum. In: *Proc. Int. Comput. Music Conf* (1990)
46.18 A. Klapuri, M. Davy: *Signal Processing Methods for Music Transcription*, Vol. 1 (Springer, New York 2006)
46.19 M. Wright, W.A. Schloss, G. Tzanetakis: Analyzing afro-cuban rhythms using rotation-aware clave template matching with dynamic programming. In: *ISMIR* (2008) pp. 647–652
46.20 M. Wright: Empirical Comparison of Two Recordings of the Kazakh Dombra piece 'Akbai'. https://ccrma.stanford.edu/~matt/dombra/ (2005)
46.21 J. Hochenbaum, A. Kapur, M. Wright: Multimodal musician recognition. In: *Proc. Int. Conf. New Interfaces Music. Expr.* (2010)

Part G

Music Ethnology

Part G Music Ethnology

Ed. by Rolf Bader

During its foundation in modern times, around the year 1900, systematic musicology was termed comparative musicology: comparing music from all over the world to find universals in music. Comparative studies arose in many disciplines during the 19th century, e.g., in archeology, where Julius Braun founded a comparative archaeology study around 1850 comparing the ancient Egyptian, Assyrian and Greek cultures. To compare music from around the world, a recording device was needed, which Edison invented in 1877. The Edison phonograph recorded music through a large horn on wax cylinders. Carl Stumpf and his assistant Erich von Hornbostel recorded a Thai *phi pha* orchestra playing in the Berlin Tiergarten in 1900 and therefore started the first phonogram archive in Berlin. From there on musicologists, archeologists or other world travelers were encouraged by the archive to record music on these cylinders and send them to Berlin.

Right from the start comparative musicology combined musical acoustics and music psychology with music ethnology. The instrument classification of Horbostel/Sachs published in 1914, and still valid today, sorts instruments according to their driving mechanism: a physical parameter rather than a place of origin. The classification lists an enormous number of instruments from all over the world, as does the Sachs Real-Lexikon of musical instruments from 1913.

Still, not all researchers in the field were interested in comparing the music traditions of the world and finding universals; some were interested in documenting single music cultures. Jaap Kunst (1891–1960) started collecting music during his free weekends as a colonial official in Bandung, Java by taking a boat to nearby islands in the Indonesian archipelago, walking into the villages and playing his violin, therefore encouraging the natives to show and play their instruments. After sending many wax cylinders to Berlin and publishing his work he was made the first fully paid ethnomusicologist in modern times. Therefore, after World War II next to comparative musicology the field of ethnomusicology, which recorded and documented single musical cultures without comparing them or looking for universals, was strong, especially in the US.

Comparative musicology was also misused by musicologists who argued that collecting music from around the world should be done to prove the superiority of the Western music tradition. These attempts were strong in Nazi Germany and were counteracted by close interaction between the German *Gesellschaft für Vergleichende Musikwissenschaft* and the *American Society of Comparative Musicology*, with a double membership to bring comparative musicologists out of Germany in times of danger. Charles Seeger, the father of Pete Seeger, the folk musician and political activist, was president of both societies in the 1930s and 1940s.

Still, music ethnology has a much longer history. During the 3rd and 2nd century BC in China there was an institute collecting and preserving the musical traditions of China, which had more than 800 researchers. This tradition in China continues to this day and large encyclopedias of the ethnic music traditions of China and neighboring countries exist. India is another example with a long and huge tradition of ethnomusicological research. However, these traditions are merely case studies and do not compare traditions.

During the 1980s and 1990s many case studies were conducted. The field also started discussing music in different contexts, including gender studies, cultural studies or the like. Still it was found that many publications might discuss the social or political context of the music but no longer the music itself. Indeed many publications were not authored by musicologists but conducted in related fields like political studies, cultural studies or anthropology.

During the last ten to fifteen years there has been a renewed interest in comparing melodies, tonal systems, textures or rhythms and discussing music itself with a broad background of technical and musicological knowledge. These are now subsumed under terms like computational ethnomusicology, analytical ethnomusicology or again comparative musicology. Music of the world is studied using musical acoustics, neuromusicology, gene expressions, and complex computational techniques, along with the invention of new phonogram archive standards. Related fields like popular music studies or music theory also try to study musical traditions in a systematic way.

The present section tries to give some insight into ongoing research directions in modern ethnomusicology. The papers provide insight into new trends and research paradigms, and with through many examples they show links between ethnomusicology and systematic musicology, or showcase exemplary research structures. It cannot be comprehensive in terms of giving an overview of musical traditions in the world or a history of world music, which has already been excellently done, e.g., in the *Garland Encyclopedia of World Music*. Also the computational part is left to Part F of this volume on Music and Media.

Jukka Louhivuori gives an overview of the relation between systematic musicology and Ethnomusicology in Chap. 47. He addresses the historical background of ethnology based on colonial interests as a collection of

case studies and the development of comparative musicology as an interest in general rules and relations in music concerning melodies, tonal systems, musical instruments, rhythms and other musical features. He gives many examples of modern methods in the field, like the use of multidimensional scaling techniques, implementation-realization models or methods of music information retrieval.

Comparing the American with the European tradition, *Leslie Tilley* in Chap. 48 discusses concepts, methods and views of ethnomusicology. Starting from Frances Densmore and Cunningham Fletcher investigating native American Indian music, and the Hornbostel/Stumpf/Sachs/Abraham European comparative musicology tradition, she goes on to discuss the post-World War II cultural relativistic approaches of Merriam, Hood and Blacking. From there, she gives many examples of the continuous analytical approaches of Kolinski starting from the 1960s, Alan Lomax and Victor Grauers' cantomerics project, and modern fields like analytical ethnomusicology co-initiated by Michael Tenzer, computational ethnomusicology, or studies in music and genetics.

As an example of music analysis at work, in Chap. 49 *Simha Arom* gives an overview of the structure of sub-Saharan music. After listing several general features of music-making in context, musical features like tonal systems, timbre, rhythms and polyphonic structures are listed and described. The analysis includes many musical styles, as referenced in the paper, supporting the findings and exemplifying the comparison of musical traditions within a geographic region.

Another analysis example of the music in a geographic region is given by *Håkan Lundström* in Chap. 50 about music of the *zomi*, the hill people of Southeast Asia. Despite the huge diversity of tribes like the *Lisu, Hmong, Kammu, Karen* etc. in Laos, Thailand, Vietnam or Yunnan, their music is subsumed under musical features like timbre related to the musical instrument material, singing styles associated with tonal languages or ensemble playing related to social and ritual habits.

An overview of music archeology is given by *Ricardo Eichmann* in Chap. 51. After discussing the historical roots of the discipline from the 19th century to the present he gives an overview about important excavations of musical instruments starting from flutes that are 40 000 years old, or of harps and flutes in ancient Sumeria, Mesopotamia or Egypt. As the music itself is not preserved the main focus of analysis is tonal systems derived, e.g., from flute tone hole spacing or from instrument geometry developments.

In Chap. 52 *David Borgo* gives insight into modern theories on music improvisation appearing in nearly all music cultures around the world. He discusses improvisation as a cognitive task, from creativity to pattern concatenation; time perception and time hierarchy; and the relation between composition and improvisation or the reasons for improvising at all. The modern term *free improvisation* as found in a Western music based on free jazz and contemporary classical music is considered.

As an example of the relation of music, musicology and politics, *Anthony Seeger* in Chap. 53 gives an overview of the history of American protest songs. Important musicians like Woody Guthrie, Pete Seeger and Bob Dylan are discussed in the context of political and industrial development and movement in the US. The influence of musicologists like Charles Seeger and Alan Lomax in the foundation and distribution of that movement is described.

47. Interaction Between Systematic Musicology and Research on Traditional Music

Jukka Louhivuori

The origin of systematic musicology is strongly linked to the studies of music cultures of non-Western origin. From the methodological point of view, folk music research applied systematic methods to collect and analyze data. Anthropology of music and later ethnomusicology had a different focus: musical phenomena should be interpreted in their cultural context. The cognitive approach was the third paradigm change in the field of systematic musicology, which again changed both methodology as well the point of view of research topics. In cross-cultural music cognition, as well as in cognitive ethnomusicology, previous approaches in systematic musicology, ethnomusicology, and cognitive science of music are combined: systematic analysis of data related to human cognition interpreted in a cultural context.

47.1 Background

From the very early stages of systematic musicology, researchers have been influenced by *folk music*, later more often called traditional music. (In the early stages of systematic musicology the term *folk music* was often used, but later the term was replaced with the term *traditional*, which does not take a stance on the social position of the people whose musical culture is under study.) Western musicology has been dominated by historical musicology, which typically concentrates on studies of European art music tradition. The music of countrymen (hunters, fishermen, farmers) was not in the core of interest until around the second half of 19th century.

The growing interest of Western musicologists in music cultures of non-Western origins (Africa, South America, Asia, Near and Far East) was closely related to colonialism. (*Jean de Léry* [47.1] was probably the first Western person who wrote about non-Western music cultures, in this case Brazil.) Connections with previously unknown (music) cultures increased interest in habits and cultures of the people living in colonial states. The first step was to collect instruments, costumes, music, dance and artifacts of other cultural activities.

New findings, such as musical instruments, playing/singing styles, scales, rhythms, harmonies and other musical behavior not known before, and new data collecting methods and equipment (phonograms, photographs and films) increased interest in and knowledge about folk/traditional music and music of non-Western origins. New equipment made it possible to analyze data in detail, which unveiled new musical phenomena.

New findings had a strong influence on the development of systematic musicology from the point of view of research topics, methodologies and theories about music. Early studies of folk/traditional/ethnic music did not take very seriously connections of musical behavior with the social and cultural context of musical behavior. Thus, the core of an ethnomusicological approach was missing from the first steps of folk/traditional music research done by systematic musicologists.

In the following text the interaction between folk/traditional, ethnomusicology and systematic musicology will be discussed. An overview of the historical background will be given, and the development of major research topics and methodologies will be described, as well as the most recent developments.

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_47

47.2 Folk/Traditional Music Research

As an example of early influences of folk/traditional music in systematic musicology are the findings of previously unknown scales and intonation systems. These forced music researchers to develop new theories about music and new conceptual tools to analyze music. In order to precisely describe interval sizes and intonation in scales *A.J. Ellis* (1814–1890) invented a system of cents, which is today still largely used in systematic musicology (Fig. 47.1) [47.2]. *Carl Stumpf* was interested in non-Western music cultures and his findings had a strong influence on theories about music [47.3, 4]. Later, *Erich von Hornbostel* (1877–1935) studied African instruments [47.5], and based on the findings he developed together with *Curt Sachs* a hierarchical classification system of instruments [47.6].

In addition to colonialism, nationalism had a strong impact on the field. Romanticism and – in some cases – political changes had an influence on researchers representing new nations or nations dreaming of independency. In Finland Ilmari Krohn (1867–1960) began to collect folk music systematically at the end of 19th century. His aim was to develop a system by which melody variants could be categorized into logical (lexical) order [47.8, 9].

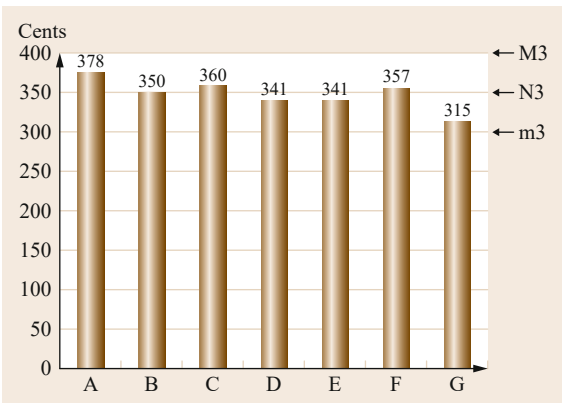


Fig. 47.1 Comparison of intonation differences in cents in Lutheran hymns in an African context. Neutral thirds and other intonation systems not used in classical Western art music tradition were difficult for early systematic musicologist to recognize. A common interpretation was (and still is for those not familiar with different intonations) to interpret the use of neutral thirds as *out of tune singing*. Invention of cents by *Ellis* [47.2] gave a precise method to measure previously unknown scales and intonations. The figure above describes intonation of thirds by seven Pedi singers in Limpopo province, South Africa. Six of the seven singers used neutral third instead of major or minor third (M3 = equal-tempered major third, 400 cents; N3 = equal-tempered neutral third, 350 cents; m3 = equal-tempered minor third, 300 cents) (after [47.7])

The systematic work by Krohn inspired other music researchers in Europe and a lively discussion started about which would be the best way to categorize folk melodies. Krohn soon understood the difficulties of the task, which resulted from the fact that folk melodies existed in thousands of variations. People sung and played music differently from one village to another, and from one performance to another. This finding caused theoretical discussions about similarity, variations, and melody families. *Oswald Koller*, Bela Bartók and many other researchers and composers participated in this theoretical discussion and searched for the *best method* to categorize folk music [47.10–12].

Krohn based his system on the cadence structure of melodies. He looked for stable positions of melodies and found out that the most invariant structural parts of melodies were cadences at the end of phrases [47.9]. Theoretical discussion on classification of folk/traditional music continued for decades and is still active [47.13–20].

The quick development of music technology has made it possible to record and store music easily, which has caused new problems related to music archives and databases. Transcription of music turned out to be a very slow process. A need to develop automatic transcription methods and to apply computer technology to statistical analysis of music was obvious [47.21–23].

The development of computer technology inspired researchers to invent new coding systems for the needs of music archives and researchers. Digitalization impacted on origination of new formats (MIDI, WAV, etc.) to store folk/traditional music. (See for example ESSEN collection, which contains about 10 000 Chinese and German folksongs [47.24, 25]. The collection of Finnish Folk Tunes, containing about 12 000 folk melodies, was published on the Internet in MIDI format in 2004 [47.26].) Recently the number of databases containing almost any kind of music is growing rapidly [47.27–30]. Today musical databases store millions of melodies and – in some cases – also contextual data. As an example, the Garland Encyclopedia of World Music Online is an online resource for music researchers interested in studying ethnic music cultures including both audio, video and text material [47.31].

Automatic transcription has appeared to be a very complex task and even today automatic music transcription contains many theoretical and practical problems, which need to be solved [47.32–34].

The size of databases has in turn caused the need to develop methods that might be helpful in musical

data mining (music information retrieval, MIR) [47.35–38].

From the point of view of systematic musicology, folk/traditional music and music of non-Western origins offers challenges for the researchers of our times. In addition to automatic transcription, the *traditional* theoretical topic about similarity is still under lively discussion [47.39–42].

The question of melodic similarity is closely related to the topics typical to music psychology (perception, cognition etc.). Because of the interest of systematic musicology in folk/traditional music, research topics moved naturally towards approaches and methodologies typical to cognitive sciences. Perhaps it is not incorrect to argue that interest in folk/traditional music and music of non-Western origins was the key factor in the paradigm change in systematic musicology in 1990s towards new approaches, such as comparative, cross-cultural and cognitive musicology.

47.3 Comparative Musicology

Systematic musicology as a term appeared in *Adler*'s writings in which he divided musicology into two fields: historical and systematic musicology [47.43]. Systematic musicology included:

1. Investigation and founding of laws in harmony, rhythm, and melody
2. Aesthetics of tonal art
3. Musical pedagogics and didactics
4. *Musicology* (examination and comparison for ethnographic purposes).

Thus, in this first definition of systematic musicology *examination and comparison for ethnographic purposes* was mentioned. Later the term comparative musicology appeared in the writings of researchers belonging to so call Berlin School [47.44–46].

If the study of folk/traditional music had a strong impact on the development of archiving music, analysis (categorization), analytical tools and methods, increasing data from non-Western cultures increased interest on comparative approaches.

A well-known systematic attempt to look at musical cultures from a global perspective was the research done by Alan Lomax. He analyzed thousands of songs and singing styles collected from different parts of the world. The aim was to categorize musical styles in a global perspective. Different from typical approaches in systematic musicology in which scales, rhythmic patters, melodic structure etc. are in the core of analysis, Lomax took into account social and cultural factors like group cohesion, orchestral cohesion, and features of voice production (vocal quality, breathiness etc.). He came to the conclusion that music cultures of the world can be categorized into eight major groups: Eurasian, Siberian-Amerindian, Pigmy, sub-Saharan African, Australian, Melanesian, and Polynesian [47.47–49].

Research by Allan Lomax differed from previous studies in systematic musicology especially because he took into account social and cultural factors. By doing this he came closer to approaches typical to ethnomusicology. Work by Lomax can be seen as a link between systematic folk/traditional music research and ethnomusicology.

47.4 Cognitive Approaches – Cross-Cultural Music Cognition and Cognitive Ethnomusicology

Typical to the early stages of systematic musicology was the interest in music psychology, especially Gestalt Theory [47.3, 50]. Later research oriented more towards *mechanical* classification of music, and automatic transcription methods (Fig. 47.2a–d) especially for the needs of music archives. The problem of categorization appeared to be much more complex than researchers had thought. It became clear that in order to solve the problem of similarity, categorization, classification etc., the cognitive mechanism of music perception, learning and production should be understood.

Linguist theories [47.51–53] inspired music researchers to apply cognitive and generative theories to the study of music. One of the first attempts was the study by *Sundberg* and *Lindblom* in which they developed a generative model that generated Swedish folk song variations [47.54, 55]. Later *Lehrdahl* and *Jackendoff* introduced the generative theory of tonal music [47.56]. Generative approaches developed by linguists and adapted by musicologists had a strong influence on the development of systematic musicology and in the growing inter-

est in the cognitive background of musical behavior.

Cognitive psychology – as with many other disciplines – has been a strongly Western-oriented field of science. What guarantees that theories developed and tested only in Western contexts are valid among other cultural contexts? Cross-cultural comparisons are needed to test the universal character of theories. Cross-cultural approaches aim to look at cognitive processes from a global perspective. How do people perceive music and what is the role of culture in perception? Methodologies of cognitive science were applied to studies of music cultures around the world. Cross-cultural music perception and cognition aimed to test existing theories in new cultural contexts [47.57–74].

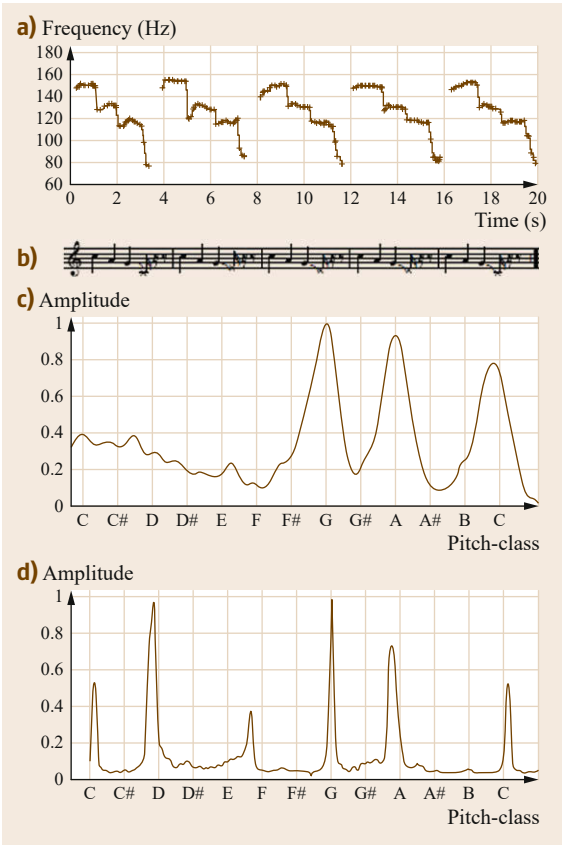


Fig. 47.2a–d Different representations used in the analysis of the singing of a Pedi traditional healer in South Africa [47.57, p. 240]. A quite common practice in traditional music contexts is the gradual rising of the pitch. This is one of the problems for automatic transcription. In the above example, the basic motive is repeated and the pitch rises gradually towards the end of the song. Recording solo performances during rituals is a difficult task. In this example the traditional healer followed the ritual from the video in a studio setting. He joined in with the singing heard from the video. By recording the singing of the traditional healer in the studio, it was possible to capture one improvised version of the song typical to Pedi traditional healers' rituals. **(a)** Pitch shift in Hz, **(b)** traditional notation, **(c)** frequency of pitches transposed to same octave, **(d)** absolute frequency of pitches

Cognitive approaches applied to folk/traditional and non-Western music research changed the direction of systematic musicology towards a more psychological (cognitive psychology) orientation. By this change the interest turned the field of systematic musicology closer to the approaches in the earlier stages of the field.

This cognitive turn had a strong impact on analytical methods. Modern statistical methods (computational methods, multidimensional scaling, see Fig. 47.3), computational modeling (self organizing maps, SOM), brain scanning (electroencephalogram, functional magnetic resonance imaging), and many others were applied to the methodological repertoire applied by researchers in systematic musicology [47.17, 20, 75–77].

Physical modeling of musical instruments has been a popular research topic in systematic musicology. In addition to computational modeling of Western instruments, computational modeling was applied also to the study of instruments used in folk/traditional music and instruments of non-Western origin [47.78–81]. An interesting and important development has been to create transcription methods, which take into account the way the sound is produced by the musicians [47.82].

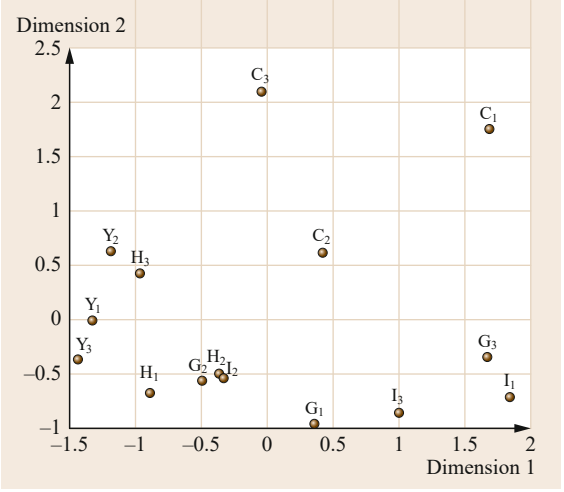


Fig. 47.3 Multidimensional scaling applied to categorization of folk music styles of Greek (C), Hungarian (H), German (G), Irish (I) and Sami people (Y) (after [47.75, p. 3])

47.5 Anthropology of Music – Ethnomusicology – Cultural Musicology

The terms *ethnic*, *folk*, and *traditional* used in this article have appeared to be problematic from many points of view. The term *ethnic* refers to something that is *far* (mostly geographically), *folk* refers to social position of the group of people under study (countrymen, fishermen, farmers, etc.), and *traditional* can refer as well to traditional classical music as other musical genres. The Society for Ethnomusicology describes ethnomusicology as the study of music in its cultural context:

> *Ethnomusicologists approach music as a social process in order to understand not only what music is but why it is: what music means to its practitioners and audiences, and how those meanings are conveyed.* [47.83]

The use of the term ethnomusicology is relative young; it was first introduced by *Jaap Kunst* in 1950 [47.84]. In the early years of systematic musicology terms such as folk music research, research on traditional music and comparative musicology were more often used. The key difference between approaches typical to systematic musicology studying folk/traditional music and ethnomusicology is that systematic musicologists have been more interested in collecting, categorizing and analyzing music of *folk* or *ethnic* origins, while ethnomusicologists emphasize the necessity to interpret findings in social and cultural context.

Definitions of systematic musicology have taken a different stance to the position of ethnomusicology in systematic musicology. In the definition by *Adler*, systematic musicology is described as a subdiscipline of musicology the aim of which was *examination and comparison for ethnographic purposes* [47.43]. Later, ethnomusicology was described as an independent, third major discipline in musicology (historical musicology, systematic musicology, and ethnomusicology) [47.68, 85, 86].

If the beginning of 20th century focused research on classification, categorization and on the question of similarity, in the second half of the century awareness of the function of culture grew. Especially writings by *Allan Merriam* and *Bruno Nettl* caused a paradigm change and increasing interest in interpreting musical behavior in social and cultural context [47.87–89]. Anthropological orientation gave new directions to the studies of folk/traditional music. The new questions were, for example, *what happens if two musical cultures meet?*, *what are the features that are assimilated by another culture?* etc.

Typical for contemporary systematic musicology influenced by anthropology of music and ethnology is an enduring interest in analyzing musical structures, acoustics and physics of ethnic musical instruments [47.90, 91], in developing new information retrieval methods or to analyze *big data* etc., but at the same time awareness of the function of society and culture is growing (Fig. 47.4). On the other hand ethnomusicology has broadened its field by applying traditional ethnomusicological methodology to new musical genres like Western classical music, pop/rock, jazz, contemporary art music and music and media. The geographical distance or ethnicity of music cultures is not anymore what makes research *ethnomusicological*. Melodic, rhythmic, harmonic structures, instruments, dances, costumes and other aspects present in musical performance are not of interest on their own, separated from their social and cultural context. Thus, an ethnomusicologist can study Beethoven symphonies, rap artists' performances, contemporary art music, or Indian classical music in addition to more traditional ethnomusicological topics like African *folk music*. Contemporary ethnomusicologists can study any music of the world, but not by applying any methodology or not taking social and cultural context into account. The new approach is called *cultural musicology* or *new musicology* [47.92].

The demand of understanding the social and cultural background of musical behavior has several methodological consequences. Understanding musical behavior in social and cultural context requires deep knowing of people's social and cultural habits. It is relatively easy to record a song, transcribe it and analyze the intervallic and rhythmic structure of it. It is much more challenging to describe the function of a specific song, for example the function of a song in ritual use, or to explain why certain instruments are allowed to be played only by men or women [47.93, 94].

It is common in ethnomusicology to increase cultural knowledge by living a relatively long time among the people under study (fieldwork). By sharing common experiences with people, observing their everyday life, and participating in musical and other cultural events and activities, understanding and trust between the researcher and the people increases. Trust is in many cases a necessary condition to get reliable and valid information from participants. (One of the first researchers conducting fieldwork by living long periods within music cultures studied was *Richard Waterman*, who studied Australian aboriginal music culture [47.95].)

a)

b)

c)

d)

	Sami	Finnish	Western	African
Original (*df* = 105.4)	$R^2 = 0.22$	$R^2 = 0.15$	$R^2 = 0.28$	$R^2 = 0.29$
Registral direction	0.04*	0.01	0.01	0.01
Intervallic difference	0.09*	0.06*	0.07*	0.06*
Proximity	0.06*	0.03	0.01	0.00
Unison	0.07*	0.05*	0.03*	0.05*
Registral return	0.00	0.00	0.01	0.01
Revised (*df* = 108.1)	$R^2 = 0.12$	$R^2 = 0.09$	$R^2 = 0.26$	$R^2 = 0.20$
Proximity (revised)	0.03*	0.05*	0.16*	0.12*
Pitch reversal	0.03	0.00	0.02	0.00

* Significant predictor at $p < 0.05$ level

Fig. 47.4a–d In searching for musical universals, the perception of Sami yoiks by Sami and South African traditional healers was studied. The photos and table above describe different stages of cross-cultural cognitive musicology (cultural musicology), which have connections to traditions both in ethnomusicology and systematic musicology. In addition to (quasi)experimental settings typical to systematic musicology (laboratory experiments), the author of this article aimed at being familiar with the culture of traditional healers (ethnomusicology – social and cultural context) by observation, interviews, and participation (participatory methodology) in ritual dance. Running music-psychological experiments in challenging environments creates several methodological problems (language interpretation, reading and writing skills of participants, reliability of technical infrastructure etc.). **(a)** Observation (ethnomusicology), **(b)** participation (ethnomusicology), **(c)** experiment (systematic musicology), **(d)** data analysis in cultural context (cross-cultural cognitive musicology; cultural musicology) (after [47.57]) ►

Fieldwork is traditionally understood as the basic condition for ethnomusicological research, but today – when the musical genres and objects of ethnomusicological studies are broadened – a more important precondition is that results are interpreted in social and cultural context. If a person has studied classical Western music for several decades, for example studied Chopin's Mazurkas, the need for *fieldwork* is totally different from the situation where he/she would like to study Pedi reed pipe music without any previous understanding of the Pedi society and culture.

Kenneth Pike [47.96] introduced the concepts of ethic and emic approaches, which have appeared to be useful for understanding the role of the researcher in relation to the culture under study. The ethicist approach refers to a situation in which the researcher studies the culture as an *outsider*, and emic to a situation in which he/she is an *insider* looking at the culture from an inside perspective. The ethicist approach gives the researcher an *objective* position in relation to participants. The researcher has no need to interpret certain aspects of the culture more positively in order to please the people. At the same time, representatives of the culture are not perhaps willing to openly speak about their thoughts and about their culture or to get the researcher to participate in certain culturally important events.

The emic approach is often understood to be more typical for ethnomusicology, because it makes possible a deeper understanding of the social and cultural background of people's behavior. In systematic musicology the emic approach has been very rare, perhaps partly because of the special needs of the researchers (living long periods abroad, sometimes in quite primitive environments etc.).

An ethnomusicological approach requires from the researcher skills in methodologies not typical to systematic musicology. Meaningful interpretation of data needs additional information acquired by doing fieldwork, interviewing and observing people, and even

participating in musical and other activities (participatory methodology). It is quite common that ethnomusicologists learn to play local instruments, sing the songs and dance the dances of the cultures under study.

Writings by John Blacking (1928–1990) have had an important influence both on ethnomusicology and systematic musicology. His studies among the Venda people in South Africa were strongly rooted in ethnomusicological tradition. He lived several years among Venda people, systematically collected their music, observed musical behavior and tried to learn to understand Venda music in its social and cultural context. At the same time he understood the significance of music psychology in getting deeper explanations of musical behavior [47.97–102].

47.6 New Trends

In the beginning of the new millennium interest towards the role of the body in musical activities grew [47.103–109]. Musical behavior, such as playing and dancing, is closely connected to body movements. It is actually strange that it took so long for researchers to look at the close connection of music and body (embodiment). Interest in embodiment is rooted in several research fields in musicology, but findings of the role of the body in traditional music was one of the reasons that inspired researchers to look more carefully at the role of the body in musical behavior.

Studies on embodiment focused first on the role of a single body, but recently the interest in interaction (synchronization, entrainment [47.111, 112]) of bodies is growing (social embodiment) (Fig. 47.5). This tendency directs research on embodiment closer to the core of ethnomusicology by emphasizing function of social and cultural factors in musical behavior [47.110, 113, 114].

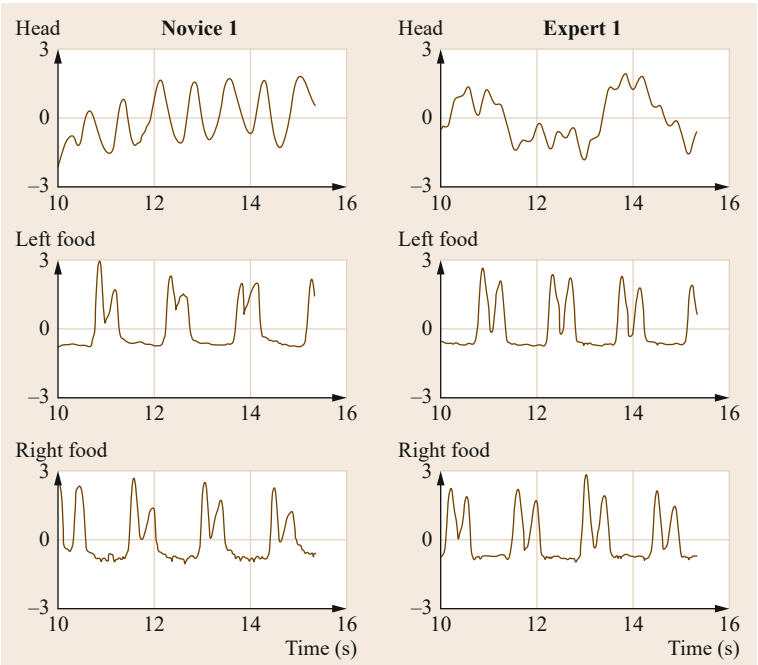


Fig. 47.5 Comparison of South African (experts) and Finnish (novice) choir singers' movements (vertical head and foot acceleration) in singing South African songs has been studied by applying contemporary motion capture technology (after [47.110])

47.7 Function of Ethnomusicology in Systematic Musicology

The function of folk/traditional/ethnic music research in systematic musicology has been strong from the very beginning of the field. Pioneers of systematic musicology, such as Adler, Stumph, Sachs and von Hornbostel, were influenced by *ethnic* music cultures and musics of *other* nations. Findings from research among music cultures not known before by Europeans challenged existing theories about music and increased understanding of the limitations of a purely *European* view of what music may be.

The systematization of data collected by folk music researchers and stored in large archives is one of the areas in which folk/traditional music research, ethnomusicology and systematic musicology has interacted. Folk/traditional music research and ethnomusicology have offered previously unknown findings about music and musical behavior. Folk/traditional music research and ethnomusicology has had an impact on the development of theoretical thinking in systematic musicology. Scales and intonation systems, new instruments, the role of dance and movement and many other findings forced researchers to rethink what music is and especially what is the role of society and culture in musical behavior.

Music as understood by the *Western world* and old theories were challenged by the new findings from *ethnic* cultures. Folk/traditional music research and ethnomusicology sped up theoretical reform and invention of new methodologies in systematic musicology. In addition, especially ethnomusicology increased understanding and need to look at ethical questions and background of people participating in experiments. Questions like *what is the impact of systematic musicology in our society?* or *how relevant is systematic musicology for the society?* came under discussion. Ethical questions are crucial for ethnomusicologists, but were not in the core of discussions in the field of systematic musicology in its early stages.

The most recent new research topics, like embodiment in music, are reflections of increased understanding of the holistic nature of musical behavior. Music is not only *art for the brain*, but *art for the human beings* living in a historical, cultural, and social context and interacting with the environment with the biophysical body.

Figure 47.6 describes the different fields of research and approaches that have interacted with systematic musicology. Over time, the role of these approaches has changed, being sometimes bigger and sometimes smaller. The process of interaction between folk/traditional music research, ethnomusicology and systematic musicology has been, and is still, dynamical in nature.

A few features typical to ethnomusicological research, which probably will be in the future more important also for the studies in systematic musicology, should be discussed. Ethnomusicologists must know well the social and cultural background of the people whose musical behavior they are studying. Due to globalization, cultural homogeneity of Western countries is quickly changing. In conducting empirical research, typical to systematic musicology, participants are needed, but do we know well enough their social and cultural background? Quite typical is to use music excerpts taken from the classical music genre, thinking that every *European* is familiar with this musical style.

How do we select participants for experiments? In what degree do the distribution of participants' cultural background represent the distribution of the population in the country? In some cases also language skills should be taken into account. If the lingual background is different from Europeans, we are faced with new methodological and practical problems in running experiments. These challenges are quite similar to those faced by ethnomusicologists: how do we communicate with participants (language skills)? Do we need language interpretation? Are participants familiar with participating in experiments? If not, how does the unfamiliar testing situation reflect on the experiment and results? How much should we understand the cultural background of participants in order to be able to make relevant conclusions? Quite rarely are these kinds of questions, typical to ethnomusicology, taken into account by systematic musicologists.

The question of sharing results with the participants, and, more generally, the responsibility of researchers for the impact of research on the society, is another topic not taken often into account by systematic musicologists. Conducting research among *ethnic* cultures, among people who have perhaps not much contact with Western cultures, necessarily causes changes to the lives of people participating in the study. If certain members of the society are interviewed and their music, dance or other kind of musical behavior is stored (audio/video recordings, photographs, films etc.), the position and status of the person in the society can change permanently. After researchers visit, he/she is the person towards whom *highly educated people from abroad* were interested. It is not sure if this special new role has a positive or negative consequence to his/her future life. This kind of change of status of participants may be a reality also when we conduct *normal* research in typical Western context.

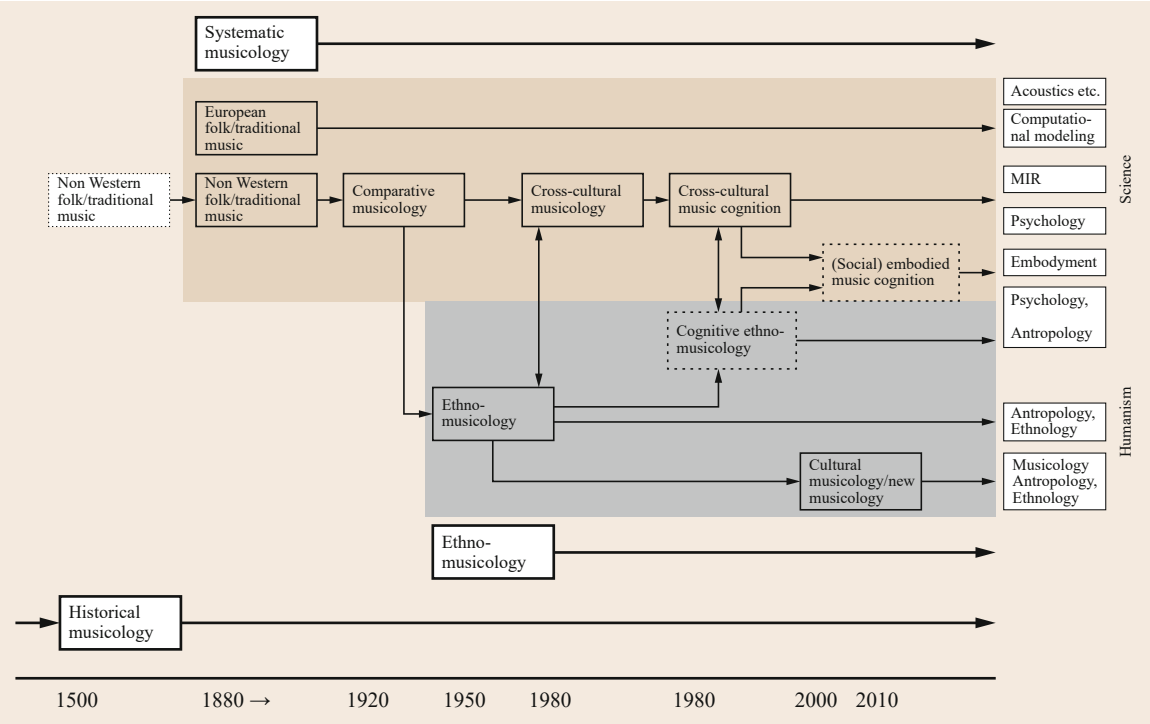


Fig. 47.6 Interaction between different approaches in folk music research, ethnomusicology and systematic musicology from historical perspective

Sharing the results and taking into account the question of immaterial rights is a serious ethical question. Who owns the rights to the data we have collected? How do we ensure that the data benefits the people under study? Or perhaps we don't care about that. We just run experiments, publish the results, and get the academic credits. In the field of ethnomusicology this kind of ethical questions must be taken seriously: What is the impact of the research for the society in general? How do we ensure that the data collected from people would be given back? In systematic musicology it is normally much easier to give feedback about the results by sharing the articles or meeting the people face to face.

It is typical in experimental research to ask permission to conduct the experiment. In Western contexts it is reasonably easy to get permissions and to find the officials from whom these should be asked. In ethnomusicological contexts this process can be much more complex. What is the official authority from where the permissions should be obtained? Among ethnic cultures in some cases local chiefs must be contacted and their acceptance to run experiments must be acquired. This process is not always a straightforward one. The researcher should somehow convince the local authority of the benefits of the research for the people. Often an intermediate person is needed who the local author (chief) knows and trusts.

During these times of increasing diversity of cultural backgrounds of people living in Western cultures, the need to ensure the relevance of research topics becomes more important. In Western context the meaningfulness of measuring for example perception of melodies is not normally doubted. What if the music culture of people participating in the study contains a very different understanding of the key concepts like *melody*, *music*, etc?

The education of most researchers in the field of systematic musicology consists of typical Western classical music education, including courses like tonality, scales, history of Western music (today also often so-called World music), harmony and counterpoint etc. Educational background has a strong impact on how we perceive and understand music as well as on what research topics we choose and believe to be of interest and importance. If one has never heard music sung or played in intonation systems not normally present in the context of Western music, such as neutral thirds, hearing music in *strange* intonation may cause serious misinterpretations, such as interpreting intonation as *out of tune* singing.

An example of *misunderstanding* certain basic elements in musical behavior is related to role of movement (dance). Although in pop/rock/jazz and folk music genres moving along with music is a normal element in performance, also in Western contexts researchers have not thought seriously about the key role of bodily movements in musical behavior. The separation of music from bodily movements has caused difficulties in understanding correctly musical behavior in cultures in which this division is not relevant at all. A better understanding of the function of the body as an integral part of the song helps us look at music from a more holistic perspective. When including bodily movements into analysis of a *simple* and *repetitive* song, the song/dance appears to be a most complex cognitive and physical activity.

African songs are normally transcribed without simultaneous transcription of the dance. This is a serious shortage, because then we leave out one of the essential elements of the musical behavior. Bodily movements create a complex counterpoint with acoustical (musical) elements of the performance. The melodic and harmonic structure can sound simple to the inexperienced listener, but can appear to be most complex when the bodily aspects are taken into account.

Embodiment in instrumental playing is another topic in which ethnomusicology and systematic musicology have a common interest. Interest in performance practices started from the studies of medieval, renaissance and baroque music practices. Later interest in music performance studies has enlarged to the study of performance in general (not limited to certain historical periods).

In systematic musicology (especially music acoustics and physics of music) study of ethnic instruments has had a key role from the early stages [47.2]. Recently, understanding of the role of embodiment in instrument playing and singing is growing. In addition to knowing physical and acoustical characters of instruments, more awareness exists today about the need to understand how the instrument is played, how the sound is produced, what the position of the musician is, how the sound is produced from the bodily aspect, etc.

In order to get a better understanding of the role of the body in instrumental playing, new methodologies have been developed. By combining knowledge of physics of the sound (sound analysis and synthesis) with the analysis of the bodily movements, it has been possible to reach a new level of description of musical behavior [47.82, 115–117]. This approach gives an interesting possibility to diminish the limitations of staff notation.

47.8 Summary

Figure 47.6 gives an alternative to look at the interaction between folk/traditional music, ethnomusicology and systematic musicology discussed in the previous chapters. In the figure the division of systematic musicology into scientific and humanistic tradition introduced by *Richard Parncutt* is taken into account [47.86].

In the early stages of systematic musicology (late 19th century), the field was inspired strongly by folk/traditional music. In addition to studies of folk music of researchers' own cultures, interest towards non-Western cultures grew rapidly. New findings and awareness of musical behavior beyond Western context had a strong influence on the theoretical development of the field. Research on non-Western music cultures created the need to compare musical cultures, which led to the development of comparative musicology. In the middle of the 20th century, anthropology and ethnology had a strong impact on musicology and directed interest of researchers towards looking at music cultures in their social and cultural context (ethnomusicology). This new approach was close to humanistic methodology and quite different from methodologies typical of systematic musicology at the time (empirical research).

In the 1980s, interest in systematic musicology grew towards cognitive sciences, cognitive psychology and computer science. New methodologies were applied to studies of music of non-Western cultures. In systematic musicology it had not been typical to take social and cultural aspects into account. Closest to ethnomusicology was the cross-cultural approach, in which an attempt was often taken to take social and cultural aspects into account.

A cognitive turn in the 1990s was strongly affected by cognitive sciences and especially computer science. Modeling of cognitive processes in music forced researchers to take more seriously cultural aspects. Cross-cultural music cognition and cognitive ethnomusicology were approaches in the 1990s. These approaches were methodologically quite close to each other. The cross-cultural approach typically applies more scientific and cognitive ethnomusicology humanistic methodologies [47.86, 118].

The problems with terms like *folk*, *traditional* or *ethnic* and the interest of ethnomusicologists to enlarge their studies into new areas, such as Western classical music, pop/jazz/rock genres and music and media,

made the use of the term ethnomusicology problematic. The core of ethnomusicology is not today so much in the object of the study, but more on the methodology: to study musical behavior in social and cultural context. This change of research object was the origin of the appearance of a new term, cultural musicology – sometimes called *new musicology* [47.92].

Cultural musicology and systematic musicology do not differ so much in research topics: both fields can study music of non-Western cultures, classical music, pop/jazz/rock genres, music and media etc. The more typical difference between these fields is that the systematic approach applies scientific and cultural musicology humanistic approaches. Social embodied music perception and cognition is one of the fields of systematic musicology in which social and cultural (humanistic) aspects have an important role.

It is apparent that most of the research topics and questions typical to the early stages of systematic musicology still exist, such as interest in folk/traditional music (variation, similarity etc.), acoustics and physics of music. In addition to the traditional research topics and methodologies, new topics and approaches have emerged into which ethnomusicology has had a clear impact, such as cross-cultural music cognition.

It remains an open question when systematic musicology and cultural musicology (ethnomusicology) and the third major field in musicology, historical musicology, will find more common interest. In order to get a holistic understanding of a complex musical phenomenon in its historical, social and cultural context, a closer cooperation between these three major subjects is needed.

References

47.1 J. de Léry: *History of a Voyage to the Land of Brazil* (Univ. California Press, Berkeley 1993)

47.2 A.J. Ellis: On the musical scales of various nations, J. Soc. Arts **33**, 485–527 (1885)

47.3 C. Stumpf: *Tonpsychologie*, Vol. 1 (Hirzel, Leipzig 1883)

47.4 C. Stumpf: *Die Anfänge der Musik* (Barth, Leipzig 1911)

47.5 E.M. von Hornbostel: The ethnology of African sound-instruments. Comments on *Geist und Werden der Musikinstrumente* by C. Sachs, Africa **6**(2), 129–157 (1933)

47.6 E.M. von Hornbostel, C. Sachs: Systematik der Musikinstrumente. Ein Versuch, Z. Ethnol. **46**(4/5), 553–590 (1914)

47.7 J. Louhivuori: Lutheran hymn singing in African context: Lobe den Herren, den mächtigen König der Ehren vs. Rena Morena wa matla! In: *Systematic and Comparative Musicology: Concepts, Methods, Findings*, Hamburger Jahrbuch für Musikwissenschaft, Vol. 24, ed. by A. Schneider (Peter Lang, Frankfurt am Main 2008) pp. 339–358

47.8 I. Krohn: *Über die Art und Entstehung der geistlichen Volksmelodien in Finnland* (Otava, Helsinki 1899)

47.9 I. Krohn: Welche ist die beste Methode, um Volks- und volksmäßige Lieder nach ihrer melodischen (nicht textlichen) Beschaffenheit lexikalisch zu ordnen?, Sammelbd. Int. Musikges. **4**, 643–660 (1903)

47.10 O. Koller: Die beste Methode, Volks-, und volksmäßige Lieder nach ihrer melodischen Beschaffenheit lexikalisch zu ordnen?, Sammelbd. Int. Musikges. **IV**, 1–15 (1902)

47.11 B. Suchoff: *Preface to the Hungarian Folk Song, by B. Bartók* (State Univ. New York Press, Albany 1981), pp. ix–lv

47.12 M. Domokos: Bartóks Systeme zum Ordnen der Volksmusik. Studia Musicologica Academiae Scientiarum Hungaricae, T. 24, Fasc. 3/4. In: *Rep. Int. Bartók Symp* (Akadémiai Kiadó Stable, Budapest 1982) pp. 315–325

47.13 W. Wiora: Systematik der musikalischen Erscheinungen des Umsingens. In: *Jahrbuch für Volksliedforschung*, Vol. 7 (1941) pp. 128–195

47.14 A. Elscheková: Methods of classification of folktunes, J. Int. Folk Music Counc. **18**, 56–76 (1966)

47.15 A. Elscheková: Systematizierung, Klassifikation und Katalogisierung von Volskliedweisen. In: *Handbuch des Volksliedes*, Vol. II, ed. by R. Brednich, L. Röhrich, W. Suppan (Fink, München 1975)

47.16 M.S. Keller: The problem of classification in folksong research: A short history, Folklore **95**(1), 100–104 (1984)

47.17 Z. Juhász: Contour analysis of hungarian folk music in a multidimensional metric-space, J. New Music Res. **29**(1), 71–83 (2000)

47.18 Z. Juhász: The structure of an oral tradition: Mapping of Hungarian folk music to a metric space, J. New Music Res. **31**(4), 295–310 (2002)

47.19 Z. Juhász: Segmentation of Hungarian folk songs using an entropy-based learning system, J. New Music Res. **33**(1), 5–15 (2004)

47.20 Z. Juhász: A systematic comparison of different European folk music traditions using self-organizing maps, J. New Music Res. **35**(2), 95–112 (2006)

47.21 B.H. Bronson: Mechanical help in the study of folk song, J. Am. Folk. **62**(244), 81–86 (1949)

47.22 B.H. Bronson: Some observations about melodic variation in British-American folk tunes, J. Am. Musicol. Soc. **3**, 120–134 (1950)

47.23 B.H. Bronson: Melodic stability in oral transmission, J. Int. Folk Music Counc. **3**, 50–55 (1951)

47.24 H. Schaffrath: The retrieval of monophonic melodies and their variants – Concepts and strategies for computer-aided analysis. In: *Computer Representations and Models in Music*, ed. by A. Marsden, A. Pople (Academic, London 1992) pp. 95–109

47.25 H. Schaffrath: The Essen associative code: A code for folksong analysis. In: *Beyond Midi: The Handbook of Musical Codes*, ed. by E. Selfridge-Field (MIT Press, Cambridge 1997) pp. 343–361

47.26 University of Jyväskylä: http://esavelmat.jyu.fi/index_en.html, Cited 15 August 2014

47.27 Essen University: Themefinder, http://essen.themefinder.org, supported by Stanford University

47.28 Wikipedia: List of online music databases, http://en.wikipedia.org/wiki/List_of_online_music_databases

47.29 Virtual Library for Musicology, http://www.vifamusik.de/home.html

47.30 Wikipedia: Berliner Phonogramm-Archiv, http://en.wikipedia.org/wiki/Berliner_Phonogramm-Archiv

47.31 Garland Encyclopedia of World Music Online: http://glnd.alexanderstreet.com (2014), cited on 28 August 2014

47.32 G. Aloupis, T. Fevens, S. Langerman, T. Matsui, A. Mesa, Y. Nunez, D. Rappaport, G.T. Toussaint: Algorithms for computing geometric measures of melodic similarity, Comput. Music J. **30**(3), 67–76 (2006)

47.33 A. Klapuri: Multipitch analysis of polyphonic music and speech signals using an auditory model, IEEE Trans. Audio Speech Lang. Process. **16**, 255–266 (2008)

47.34 J.P. Bello: Measuring structural similarity in music, IEEE Trans. Audio Speech Lang. Process. **19**(7), 2013–2025 (2011)

47.35 R. Typke: *Music Retrieval Based on Melodic Similarity*, Ph.D. Thesis (Univ. Utrecht, Utrecht 2007)

47.36 P. van Kranenburg, J. Garbers, A. Volk, F. Wiering, L. Grijp, R. Veltkamp: Towards integration of MIR and folk song research. In: *Proc. ISMIR* (2007) pp. 505–508

47.37 P. van Kranenburg, J. Garbers, A. Volk, F. Wiering, L.P. Grijp, R.C. Veltkamp: Collaboration perspectives for folk song research and music information retrieval: The indispensable role of computational musicology, J. Interdiscip. Music Stud. **4**(1), 17–43 (2009)

47.38 D. Cohen, R. Granot, C.O. Pratt, M. Lesaffre, D. Moelants, M. Leman: Access to ethnic music: Advances and perspectives in content-based music information retrieval, Signal Process. **90**(4), 1008–1031 (2010)

47.39 W. Steinbeck: *Struktur und Ähnlichkeit: Methoden Automatisierter Melodieanalyse* (Bärenreiter, Kassel 1982)

47.40 D. Müllensiefen, K. Frieler: Cognitive adequacy in the measurement of melodic similarity: Algorithmic vs. human judgements, Comput. Musicol. **13**, 147–147 (2004)

47.41 E. Selfriedge-Field: Towards a measure of cognitive distance in melodic similarity, Comput. Musicol. **13**, 93–112 (2004)

47.42 R. Typke, F. Wiering, R.C. Veltkamp: Transportation distances and human perception of melodic similarity, Music. Sci. **11**, 153–181 (2007)

47.43 G. Adler: Umfang, Methode und Ziel der Musikwissenschaft, Vierteljahrsschr. Musikwiss. **1**, 5–20 (1885)

47.44 E.M. von Hornbostel: Die Probleme der vergleichenden Musikwissenschaft, Z. Int. Musikges. **7**, 247–270 (1905), English translation by R. Campbell: The Problems of Comparative Musicology (1975)

47.45 G. Herzog: Comparative musicology, Music J. **4**, 11 (1946)

47.46 W. Rhodes: Toward a definition of ethnomusicology, Am. Anthropol. **58**, 457–463 (1956)

47.47 A. Lomax: *Folk Song Style and Culture* (American Association for the Advancement of Science, Washington 1968)

47.48 A. Lomax: *Cantometrics: An Approach to the Anthropology of Music* (Univ. California Extension Media Center, Berkeley 1976)

47.49 A. Lomax: *Cantometrics: A Method in Musical Anthropology* (Univ. California Extension Media Center, Berkeley 1977)

47.50 A. Truslit: *Gestaltung und Bewegung in der Musik* (Vieweg, Berlin 1938)

47.51 A.B. Lord: *The Singer of Tales* (Harvard Univ. Press, Cambridge 1960)

47.52 N. Chomsky: *Reflections on Language* (Pantheon, New York 1975)

47.53 N. Chomsky: *New Horizons in the Study of Language and Mind* (Cambridge Univ. Press, New York 2000)

47.54 B. Lindblom, J. Sundberg: Towards a generative theory of melody, Sven. Tidskr. Musikforsk. **52**, 71–88 (1970)

47.55 J. Sundberg: On melodical similarity in versions of a Swedish folk-tune, STL-QPSR **16**(2/3), 61–66 (1975)

47.56 F. Lehrdahl, R. Jackendoff: *A Generative Theory of Tonal Music* (MIT Press, Cambridge 1983)

47.57 T. Eerola, J. Louhivuori, E. Lebaka: Expectancy in Sami Yoiks revisited: The role of data-driven and schema-driven knowledge in the formation of melodic expectations, Music. Sci. **13**(2), 231–272 (2009)

47.58 D.L. Harwood: Universals in music: A perspective from cognitive psychology, Ethnomusicology **20**, 521–533 (1976)

47.59 M.A. Castellano, J.J. Bharucha, C.L. Krumhansl: Tonal hierarchies in the music of North India, J. Exp. Psychol. General **113**, 394–412 (1984)

47.60 J.R. Cowdery: A fresh look at the concept of tune family, Ethnomusicology **28**(3), 495–504 (1984)

47.61 E.J. Kessler, C. Hansen, R.N. Shepard: Tonal schemata in the perception of music in Bali and the West, Music Percept. **2**, 131–165 (1984)

47.62 L. Davidson, B. Torff: Situated cognition in music, World Music **34**(3), 120–139 (1992)

47.63 E.C. Carterette, R.A. Kendall, S.C. DeVale: Comparative acoustical and psychoacoustical analyses of gamelan instrument tones, J. Acoust. Soc. Jpn. E **14**(6), 383–396 (1993)

47.64 D. Huron: The melodic arch in western folksongs, Comput. Musicol. **10**, 3–23 (1995)

47.65 A.H. Gregory, N. Varney: Cross-cultural comparisons in the affective response to music, Psychol. Music **24**, 47–52 (1996)

47.66 C.L. Krumhansl, J. Louhivuori, P. Toiviainen, T. Järvinen, T. Eerola: Melodic expectation in Finnish Spiritual Folk Hymns: Convergence of statistical, behavioral, and computational approaches, Music Percept. **17**, 151–196 (1999)

47.67 C.L. Krumhansl, P. Toivanen, T. Eerola, P. Toiviainen, T. Järvinen, J. Louhivuori: Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks, Cognition **76**, 13–58 (2000)

47.68 I. Cross: Music, cognition, culture, and evolution, Ann. N.Y. Acad. Sci. **930**, 28–42 (2001)

47.69 R. Bod: Memory-based models of melodic analysis: Challenging the Gestalt principles, J. New Music Res. **31**(1), 27–37 (2002)

47.70 S. Ahlbäck: *Melody beyond Notes: A Study of Melody Cognition*, Ph.D. Thesis (Göteborgs Universitet, Göteborg 2004)

47.71 G. Kubik: *Theory of Afrian Music* (Univ. Chigago Press, Chigago 2010)

47.72 C.L. Krumhansl: Music psychology and music theory: Problems and prospects, Music Theory Spectr. **17**, 53–90 (1995)

47.73 C.L. Krumhansl: Effects of musical context on similarity and expectancy, Syst. Musikwiss. **3**, 211–250 (1995)

47.74 C.L. Krumhansl: Tonality induction: A statistical approach applied cross-culturally, Music Percept. **17**(4), 461–479 (2000)

47.75 T. Eerola, T. Järvinen, J. Louhivuori, P. Toiviainen: Statistical features and perceived similarity of folk melodies, Music Percept. **18**, 275–296 (2001)

47.76 S.M. Demorest, S.J. Morrison, E.H. Aylward, S.C. Cramer, K.R. Maravilla: An fMRI study of cross-cultural music comprehension-http://strukidemo.le-tex.de/help?locale=de. In: *Proc. 7th int. Conf. Music Percept. Cogn*, ed. by C. Stevens, D. Burnham, G.G. McPherson, E.E. Schubert, J.J. Renwick (Causal Productions, Adelaide 2002) p. 141

47.77 G. Tzanetakis, A. Kapur, W.A. Schloss, M. Wright: Computational ethnomusicology, J. Interdiscip. Music Stud. **1**(2), 1–24 (2007)

47.78 C. Erkut, M. Karjalainen, P. Huang, V. Välimäki: Acoustical analysis and model-based sound synthesis of the kantele, J. Acoust. Soc. Am. **112**, 1681–1691 (2002)

47.79 J. Pölkki, C. Erkut, H. Penttinen, M. Karjalainen, V. Välimäki: New designs for the kantele with improved sound radiation. In: *Proc. Stockh. Music Acoust. Conf., Stockholm* (2003)

47.80 H. Penttinen, J. Pakarinen, V. Välimäki, M. Laurson, H. Li, M. Leman: Model-based sound synthesis of the guqin, J. Acoust. Soc. Am. **120**, 4052–4063 (2006)

47.81 L. Henbing, M. Leman: A gesture-based typology of sliding-tones in guqin music, J. New Music Res. **36**(2), 61–82 (2007)

47.82 M. Kuuskankare, M. Laurson: Expressive notation package, Comput. Music J. **30**(4), 67–79 (2006)

47.83 The Society for Ethnomusicology: http://www.ethnomusicology.org/

47.84 J. Kunst: *Musicology: A Study of the Nature of Ethno-musicology, its Problems, Methods, and Representative Personalities* (Indisch Instituut, Amsterdam 1950)

47.85 B. Aarden, D. Huron: Mapping european folksong: Geographical localization of musical features, Comput. Musicol. **12**, 169–183 (2001)

47.86 R. Parncutt: Systematic Musicology and the history and future of Western musical scholarship, J. Interdiscip. Music Stud. **1**, 1–32 (2007)

47.87 A.P. Merriam: *The Anthropology of Music* (Northwestern Univ. Press, Evanston 1964)

47.88 B. Nettl: *Theory and Method in Ethnomusicology* (Free Press of Glencoe, New York 1964)

47.89 B. Nettl: *The Study of Ethnomusicology*, 2nd edn. (Univ. Illinois Press, Urbana 2005)

47.90 A. Schneider: *Tonhöhe – Skala – Klang. Akustische, Tonometrische und Psychoakustische Studien auf Vergleichender Grundlage* (Orpheus, Bonn 1997)

47.91 O. Elscheck: *Ethnologische, Historische und Systematische Musikwissenschaft* (ASCO Art & Science, Bratislava 1998)

47.92 L. Kramer: *Classical Music and Postmodern Knowledge* (Univ. of California Press, Berkeley 1995)

47.93 S. Arom: Combining sounds to reinvent the world. World music, sociology and musical analysis. In: *Analytical and Cross-Cultural Studies in World Music*, ed. by M. Tenzer, J. Roeder (Oxford Univ. Press, New York 2011) pp. 388–413, in collab. with D.-C. Martin)

47.94 T. Njoora, R. Mtwali, J. Louhivuori, O. Moilanen: Sengenya dance among coastal communities: Documentation of music, instruments and social context, Muziki **12**(2), 53–76 (2015)

47.95 R.A. Waterman: Aboriginal songs from Groote Eylandt. In: *Proc. Centen. Workshop Ethnomusicol.*, ed. by P. Corssley-Holland (Univ. of British Columbia, Vancouver 1970) pp. 102–116

47.96 K. Pike: *Language in Relation to a Unified Theory of the Structure of Human Behavior* (Mouton, The Hague 1954)

47.97 J. Blacking: *How Musical is a Man?* (Univ. Washington Press, Seattle 1973)

47.98 J. Blacking: Musical expeditions of the Venda, Afr. Music **3**(1), 54–72 (1962)

47.99 J. Blacking: Tonal organization in the music of two Venda initiation schools, Ethnomusicology **14**, 1–56 (1970)

47.100 J. Blacking: Some problems of theory and method in the study of musical change, Yearb. Int. Folk Music Counc. **9**, 1–26 (1977)

47.101 J. Blacking: Movement and meaning: Dance in social anthropological perspective, Dance Res. **1**(1), 89–99 (1983)

47.102 G. Malacari, P.S. Campbell: *The Work and Legacy of John Blacking* (Univ. of Western Australia, Crawley 2003)

47.103 F. Hoerburger: On relationships between music and movement in folk dancing, J. Int. Folk Music Counc. **12**, 70 (1960)

47.104 A. Camurri, P. Morasso, V. Tagliasco, R. Zaccaria: Dance and movement notation. In: *Human Movement Understanding: From Computational Geometry to Artificial Intelligence*, ed. by P. Morasso, V. Tagliasco (Elsevier, Amsterdam 1986) pp. 85–124

47.105 F.J. Varela, E. Thompson, E. Rosch: *The Embodied Mind – Cognitive Science and Human Experience* (MIT Press, Cambridge, London 1991)

47.106 M. Clayton, R. Sager, U. Will: In time with the music: The concept of entrainment and its significance for ethnomusicology, ESEM Counterpoint **1**, 1–82 (2004)

47.107 M. Leman: *Embodied Music Cognition and Mediation Technology* (MIT Press, Cambridge 2008)

47.108 R.I. Godøy, M. Leman: *Musical Gestures: Sound, Movement, and Meaning* (Routledge, New York 2010)

47.109 P. Toiviainen, G. Luck, M. Thompson: Embodied meter: Hierarchical eigenmodes in spontaneous movement to music, Music Percept. **28**(1), 59–70 (2010)

47.110 T. Himberg, M. Thompson: Learning and synchronizing dance movements in South African songs – cross-cultural motioncapture study, Dance Res. **29**(2), 305–328 (2011)

47.111 L. Naveda, M. Leman: The spatiotemporal representation of dance and music gestures using topological gesture analysis (TGA), Music Percept. **28**(1), 93–111 (2010)

47.112 B.H. Repp: Sensorimotor synchronization: A review of the tapping literature, Psychon. Bull. Rev. **12**(6), 969–992 (2005)

47.113 M. Leman: The societal contexts for sound and music computing: Research, education, industry, and socio-culture, J. New Music Res. **36**(3), 149–167 (2007)

47.114 A. Gritten, E. King: *Music and Gesture* (Ashgate, Aldershot 2006)

47.115 M. Müller: *Information Retrieval for Music and Motion* (Springer, New York 2007)

47.116 L. Naveda, M. Leman: A cross-modal heuristic for periodic pattern analysis of Samba music and dance, J. New Music Res. **38**(3), 255–283 (2009)

47.117 M. Leman, L. Naveda: Basic gestures as spatiotemporal reference frames for repetitive dance/music patterns in Samba and Charleston, Music Percept. **28**(1), 71–91 (2010)

47.118 P. Moisala: Cognitive study of music as culture – basic premises for "cognitive ethnomusicology", J. New Music Res. **24**(1), 8–20 (1995)

48. Analytical Ethnomusicology: How We Got Out of Analysis and How to Get Back In

Leslie Tilley

Analysis has had a long and somewhat tenuous history under the umbrella of ethnomusicology. In this chapter, we examine the trajectory of analytical ethnomusicology, from its parallel beginnings in late 19th-century Europe and North America through its relative obscurity in the field in the mid-20th century to its panoply of new methods in the late 20th and early 21st centuries. The aim of the chapter is threefold. Looked at in one way, it is a simple historical overview of analysis in ethnomusicology: an examination of the major players, from Erich Moritz von Hornbostel to Alan Lomax to many of today's central scholars, as well as the major trends and intellectual frameworks influencing its execution, from cultural evolutionism to cultural relativism to interdisciplinarity. Yet it is also designed as an exploration of the myriad methods and approaches in the analytical ethnomusicologist's toolkit, from transcription and trait listing to structural analysis, computational analysis, and the new comparative analysis. And finally, woven throughout is the story of the place of analysis in ethnomusicological research: its strengths and weaknesses, successes and mistakes, practitioners and detractors. Through these discussions, we then begin to unpack the ebbs and flows of its use, reception, and usefulness in the field.

Analysis has had a long and somewhat tenuous history under the umbrella of ethnomusicology. Beginning as one of the central activities of both North American music ethnologists and European comparative musicologists in the late nineteenth century, it fell out of favor in the post-World War II era with the rise of anthropology-based ethnomusicological studies, and since the late 1960s has been rather relegated to a status of red-headed stepchild within in the discipline. *Joseph Kerman*'s provocatively titled *How We Got into Analysis, and How to Get out* might equally have been leveled at ethnomusicologists [48.1]. Yet there has remained throughout a subset of ethnomusicologists dedicated to developing newer, more applicable, and more culturally sensitive analytical methods, which have increasingly diversified through the twenty-first century to include computational and interdisciplinary approaches, among others. This chapter will explore the changing trajectory of analytical ethnomusicology over the course of the last 130 years, examining its practitioners and detractors, its insights and mistakes, and its mosaic of methods.

48.1 Ethnomusicology's Analytical Roots

In this first section, we will explore the analytical methods that arose in the early history of the field – both in Europe and in North America – and examine some of the larger goals behind these early analyses. Some of this history will be familiar to scholars of systematic musicology (SM), particularly as regards the early European scholars. Yet, while the current chapter will act as a complement to other chapters in this volume, it is designed as an overview of analytical techniques used by ethnomusicologists and their predecessors, and

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_48

will thus not focus on those techniques unique to SM. Readers interested in the historical connections between these fields should refer to *Albrecht Schneider*'s 2006 article on the topic [48.2].

Although the commonly accepted narrative of ethnomusicology's history has its inception in mid-1880s Europe, we will begin instead by examining some of the studies of early North American music ethnologists. Generally less concerned with formulating grand, far-reaching theories of musical evolution and origins than their European counterparts in comparative musicology, these North American scholars tended to present more modest lists of musical traits. Starting our examination of ethnomusicological analysis here will allow us to begin grappling with analytical techniques and concepts from a simpler vantage point before then turning to the methods of European comparative musicologists.

48.1.1 Modest Beginnings: North America

The beginnings of ethnomusicological research in North America were largely focused on preservation: early American music ethnologists like Alice Cunningham Fletcher (1838–1923) and Frances Densmore (1867–1957) set out to collect and transcribe what they considered to be dying traditions, mostly Native North American musics. This was actually a key interest of many late-19th and early-twentieth century music scholars, including important European collectors not discussed in depth in this chapter. Among these were Hungarian scholars Béla Bartòk (1881–1945) and Zoltán Kodály (1882–1967), who compiled extensive collections of folk music from across Eastern Europe, as well as the less-often-cited Finnish folk music collector Ilmari Krohn (1867–1960), whose folk music categorization method later influenced both Bartòk and Kodály. Key to the efforts of all such scholars, of course, was the invention of Edison's phonograph in 1877, which allowed music scholars, for the first time, to record music in the field and then replay it for more accurate transcription and analysis.

Both Densmore and Fletcher were prolific collectors of Native American music. Densmore, for instance, *studied the music of* 76 *tribes, recorded more than* 2500 *songs, and published at least* 22 *monographs and* 175 *articles*, all between 1901 and 1940 [48.3, p. 53]. Many of these publications involved song classification and categorization alongside basic analysis, and all relied very heavily on transcriptions in Western notation. *Densmore*'s 1910 collection of Chippewa songs [48.4], for instance, comprises transcriptions of 200 songs categorized according to their social function. In this study, Densmore engages in two main styles of analysis, both of which were common in these early years: trait listing and descriptive analysis.

Trait Listing

The most basic analysis style in Densmore's study is trait listing. In this approach, the analyst begins with a list of seemingly objective musical parameters regarding scale type, melodic characteristics (range, contour, and intervals), ornamentation, tonal organization, rhythm and rhythmic organization, and form. S/he then makes a chart with all parameters listed and fills in the details for a given piece. Trait listing may be used to analyze characteristics of a single piece of music or to present a tabulated analysis of multiple songs. Figure 48.1 shows Densmore's tabulated analyses of accidental use and rhythm in Chippewa songs.

Trait listing was designed to be scientific and objective, a central goal of many nineteenth and early twentieth century music scholars. And, though more complex and comprehensive forms of analysis have arisen in the interim, similar approaches are still used today, subsumed under Mervyn McLean's category of *standard analysis*. This componential approach to music, *McLean* maintains, *is simple, relatively easy to apply, and* [. . .] *has served its purpose well* [48.3, p. 292].

Descriptive Analysis

The second approach that Densmore and other early American music ethnologists used was descriptive: prose outlining general facts about the music as well as presenting more specific observations on individual songs. In Densmore's work, each song category is first introduced with a few paragraphs for cultural context, and each song is given a handful of sentences regarding its source. Then, following each transcription is a brief analysis of its individual musical characteristics. Figure 48.2 shows an example of *Densmore*'s transcription and descriptive analysis style, this time from her study of Native songs in British Columbia [48.5].

This descriptive analysis, like basic trait listing, makes no special attempt to either uncover or reinforce a larger theory; it is simply an outlining of the observable musical characteristics of a single song. In this example, Densmore's analysis focuses on meter, phrase length, pitch use (in relation to an assumed tonic), and the more subjective evaluation of melodic character. Other analyses in the same collection touch on small-scale thematic development, melodic motion, interval use, and so on. Yet, what will hopefully be immediately apparent is that these descriptions and trait analyses are as much an analysis of the author's *transcriptions* as they are of the songs themselves. Thus, one of the first lessons we learn from examining the work of these early scholars is that transcription is, in fact, a form of analysis.

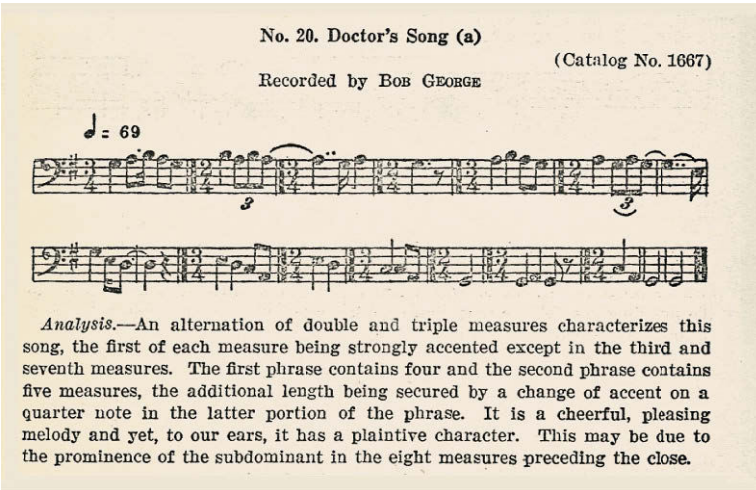
ACCIDENTALS

Songs containing no accidentals	3	112, 203, 224
Sixth lowered a semitone	1	181
Total	4	

RHYTHMIC ANALYSIS

Beginning on accented portion of measure	2	181, 203
Beginning on unaccented portion of measure	2	112, 224
Total	4	
Metric unit of voice and drum different	3	181, 203, 224
Recorded without drum	1	112
Total	4	

Fig. 48.1 *Densmore*'s trait listing, tabulated analysis (after [48.4, p. 160])



No. 20. Doctor's Song (a)

(Catalog No. 1667)

Recorded by BOB GEORGE

♩ = 69

Analysis.—An alternation of double and triple measures characterizes this song, the first of each measure being strongly accented except in the third and seventh measures. The first phrase contains four and the second phrase contains five measures, the additional length being secured by a change of accent on a quarter note in the latter portion of the phrase. It is a cheerful, pleasing melody and yet, to our ears, it has a plaintive character. This may be due to the prominence of the subdominant in the eight measures preceding the close.

Fig. 48.2 *Densmore*'s descriptive analysis (after [48.5, p. 37])

Transcription as a Form of Analysis

Undertaking transcription of oral musics involves myriad decisions. In Fig. 48.2, for instance, how does Densmore choose the key of the piece? Is a Western conception of tonal hierarchy even relevant in this tradition, and if so, how did Densmore establish that it was? Further, why the changing meter in this transcription? *Densmore* has stated of her collections that *the transcription of a song is divided into measures according to the vocal accent* [48.4, p. 5]. But why discount the possibility of syncopation within a stable metric framework, or even the idea that the music is not strictly metered at all? Further, what about allowance for variation in song performance? Densmore often records multiple versions of a single song, but only one is ever transcribed [48.5, p. 39]. Without a discussion of the differences between variants or the reasons for one particular variant being chosen over others, this reifies a single version of the song for analysis. Even seemingly small details must be decided upon when transcribing. For instance, how accurate should a transcription be in terms of pitch placement? *Densmore* admits that *ordinary musical notation does not, in all instances, represent the tones sung* in the music she transcribes [48.4, p. 3], but still staunchly clings to the idea of music based in tones and semitones, ignoring *Alexander Ellis*' strong arguments to the contrary [48.6]. She states [48.4, p. 4]:

> *At present the only standard generally available for the measurement of musical intervals is the tempered musical scale. This is artificial, yet its points of difference from the natural scale are intervals less frequently used in primitive music than those which the two scales have in common. Chippewa*

> *singers have been found who sang all the tones correctly except the fourth and seventh.*

Here I draw the reader's attention to the use of the term *correctly*, which *presumes* a musical system that uses a Western European standard of tuning, and implies that the Chippewa singers have simply not yet mastered it. Thus, every decision in transcription comes with a set of assumptions. John Comfort Fillmore (1843–1898), who often collaborated with Alice Fletcher as a music expert, had even more audacious ideas about this question of precision in pitch placement for ethnomusicological transcription. To Fillmore, the actual pitches sung were *a matter of comparatively little importance*. He claimed that [48.7, p. 288]:

> *the really important question is what tone* [the performers] meant *to sing, and on this point there can be no doubt whatever. The song as given* [in the transcription] *is exactly as they meant and sang it.*

What's more, Fillmore often transcribed Native American melodies with piano accompaniment, erroneously presuming that *the forms assumed by primitive songs are determined (unconsciously to those who make them) by a latent sense of harmony* [48.8, p. 305].

Some early scholars of non-Western music attempted to avoid the potential misrepresentation exhibited by Densmore, Fillmore, and others, either by including an enormous level of detail in their transcriptions or by creating unconventional staves. Hungarian music collector and composer Béla Bartòk was known for his meticulous detail in pitch placement (with the use of arrows to show deviance from tempered tuning) as well as in the notation of rhythm and ornamentation. American music scholar *Benjamin Ives Gilman* (1852–1933) went a different route, inventing a 45-line quarter-tone staff for the notation of Hopi songs [48.9]. Unfortunately, while putatively objective and exact, the visuals of such a transcription lead to an analysis that favors the minutiae of pitch placement and intervallic content over melodic motion or phrasing, for instance, which are visually obfuscated by the level of pitch detail present.

The choices made in the transcription process, then, are very much a part of the analytical process. As *Hornbostel* has stated, *notation, in order to be readable, must reduce facts to formulas* [48.10, p. 38]. And when a scholar's analysis is based not on the performance but on the transcription (and all of its assumptions and reductions), much of the music's nuance can be lost. That none of these early scholars performed the music they studied is equally relevant; there was no opportunity to verify their findings cognitively or experientially, nor to discuss questions of performance practice or cognition with the culture-bearers themselves. Through our twenty-first-century lenses, of course, it's easy to see the ethnocentric bias in many of these earlier studies, but the issue of transcription's subjectivity has remained to the present day. Despite the availability of better technological resources and the cognitive perspectives that modern ethnomusicological fieldwork has given us, transcription in its very nature is still an imperfect art and science, both. Thus, it is always important to ask the question posed by *Jason Stanyek* in his 2014 *Forum on Transcription* [48.11, p. 104]: *how do the practices, products, and politics of transcription fit into an ever-changing landscape of ethnomusicology?*

Concluding Remarks on Early Music Ethnology

There are undeniably numerous intractable problems inherent in both transcription and standard analysis. Yet these methods can be invaluable assets to the ethnomusicologist; we must simply be aware of their limitations, as we must be of the limitations in *any* analytical tool, be it musical or cultural. Despite their shortcomings, the analytical methods used by these early music ethnologists – descriptive analysis, trait listing, and transcription – are still central features of an ethnomusicologist's analytical toolkit. They are an excellent starting point. However, as *Mervyn McLean* points out, we should always be searching for *improved ways of looking at each of the* [musical] *components and their relationships with each other* [48.3, p. 292].

48.1.2 European Comparative Musicology

The creation story told to most budding ethnomusicologists about the origins of our discipline generally does not begin with the transcription-for-preservation, descriptive analyses, and trait listing of the early American school that we have been discussing. It begins in Europe in the late 1800s with the emergence of Vergleichende Musikwissenschaft: comparative musicology (CM). As the name suggests, comparative musicologists saw as their task not just the collection and classification but also the comparison of all the world's musics. In this pursuit, they attempted to trace historical connections between traditions through the application of empirical and analytical methods, many seeking universals in music or positing other grand theories supported by their analyses.

Most of the important nineteenth- and early twentieth-century comparative musicologists hailed from German-speaking lands, and of these, many were associated with the Berlin Phonogramm-Archiv, a collection of thousands of phonograph cylinders founded in 1900 by philosopher, acoustician, and psychologist Carl

Stumpf (1848–1936) and developed by his students: chemist, philosopher, and musicologist Erich Moritz von Hornbostel (1877–1935) and physician, psychologist, and auditory perception specialist Otto Abraham (1872–1926). Initially assembled for Stumpf's *psychological interest in the sensual experience of tones and intervals and their ordering into tone systems* [48.12, p. 204], the Phonogramm-Archiv became, through Hornbostel, a repository of recordings for academic inquiry that went far beyond the confines of psychology.

In his 1905 lecture *The Problems of Comparative Musicology*, *Hornbostel* outlined the general goals of the field, beginning with the reasons for comparison [48.13, pp. 249–250]:

> *Comparison is the principal means by which the quest for knowledge is pursued. Comparison makes possible the analysis and the exact description of an individual phenomenon by comparing it with other phenomena and emphasizing its distinctive qualities. But comparison also characterizes individual phenomena as special cases in which the similarities are defined and formulated as* laws. *Systematization and theory depend on comparison.*

Hornbostel was particularly interested in methods for the comparative study of scales and tone systems, and these would remain a central concern of CM. But his 1905 lecture also presented ideas on the analysis of melodic construction and rhythm, encouraged explorations into the nature of the musically beautiful, and incorporated some quite forward-looking musings on the inherent problems of analyzing musical systems different from our own. Many of Hornbostel's early studies, and those of his contemporaries, grappled with these tasks. In 1906 for instance, *Hornbostel* published articles on the recorded musics of both the Thompson River Indians of British Columbia [48.14] and the people of Tunisia [48.15]. Like many early American studies, these articles contain Western-style transcriptions and analyses with heavy focus on melodic characteristics and tone systems. Both attempt a statistical understanding of scalar types and seek to ascertain the inherent tone hierarchy of the music's supposed *tonalities*. *Bruno Nettl* has noted that [48.16, p. 75]

> *what has sometimes been called the* Hornbostel paradigm – *focusing on scalar structures and pitch relations and giving attention to singing style and tone colour – seems to have been developed in part for establishing an approach to a description of music that might facilitate a comparative method.*

Yet, although European comparative musicologists would use many of the same standard analysis techniques as the early Americans – examining music componentially through trait listing and descriptive analysis – they had loftier goals too.

A Basis in the Sciences: Grand Theories of Origins and Evolution

Most early comparative musicologists approached the study of music as a science in the tradition of the great syntheses of *Helmholtz* [48.17] – not surprising, given the fact that many of them, like Stumpf, Hornbostel, and Abraham, were trained as acousticians, psychologists, and physicians first. And many would take their studies further than simple description, categorization, and comparison, turning to theories and discoveries in other academic fields as a scientific basis for their research. For instance, the then-current belief that there were universals in music was based in psychology. *Hornbostel*, for example, claimed that certain musical gestures, like a descending melody moving from tension to rest, were *natural, i. e., rooted in the psychophysical constitution of man, and* [could] *therefore be found all over the world* [48.10, p. 34]. And Darwin's theory of evolution was understood to support two of the more commonly espoused theoretical orientations of CM: cultural evolutionism and diffusionism.

Cultural Evolutionism. Cultural evolutionism stemmed from the belief that all cultures evolve from primitive to civilized – and their musics from simple to complex – but that each does so at a different rate. Supporting a theory of the polygenesis of musical attributes, cultural evolutionists posited that we could understand the music of our distant ancestors by studying the music of more *primitive* cultures: our so-called *contemporary ancestors*. Aesthetician *Richard Wallaschek* of the Vienna School of CM, in his 1893 *Primitive Music* [48.18], presents a musical world tour in 300 pages, describing and categorizing each so-called primitive music in terms of its evolutionary stage. At the core of Wallaschek's study is a belief in cultural evolutionism, held without question. He says [48.18, p. 145]:

> *I can take it for granted that there are still savage tribes, whose culture has remained stationary ever since the stone age. If this is so, it seems – to say the least – extremely improbable that such tribes (as Bushmen, Australians) should at the same time have made any progress in music alone.*

The book examines everything from instrument type to singing style to a people's understanding of harmony and harmonic progression, in order to place musical characteristics and (by extension) societies along a continuum of evolution, speculating on the origins of these attributes and connecting links between societies

along the way. An ambitious early example of CM, as *Mervyn McLean* notes, *it is also compiled substantially from unreliable secondary sources and is full of mistakes and misinterpretations* [48.3, p. 241].

Carl Stumpf, the father of the Berlin School of CM, was also very concerned with evolutionary theory in his work. He theorized [48.19, 20]:

> *that a tonal system with stable steps required an intellectual development* [...] *whose consecutive stages and inner properties no one ha*[d] *yet demonstrated for us in a psychologically credible way.*

In his 1911 *The Origins of Music* [48.21], *Stumpf* attempts to do just that: to discover and lay out the *main forms of primitive melodic formation and their gradual refinement* [48.21, p. 62]. In it, he discusses the ways in which melody-making developed from the use of *noisy sounds* (*Geräusche Töne*) through the use of fixed musical intervals, to the development of a sense of tonality and melodic centralization, and so on, all as a way of categorizing music on an evolutionary scale and thus speculating upon its origins.

Diffusionism. The closely related theoretical framework of diffusionism espoused a *mono*genetic theory of evolution: all cultural traits (including musical characteristics and forms) are invented in a single location and spread outward from that point. *Curt Sachs*, one of the more ardent supporters of diffusionism, justified it over the polygenetic theories explained above by appealing to a personal logic based on the relative complexity of musical instruments and characteristics [48.22, pp. 62–63]:

> *We may believe that a tool such as a hammer can be invented everywhere at a certain stage of human evolution; the progression from the use of the bare fist, through the use of a stone in the fist, to a stone on a wooden handle, is quite logical and natural. But a bull-roarer? Is it really acceptable that every human tribe must invent an oval board, held by a cord and whirled around the head, for certain magic purposes? Is it convincing that merely because of natural evolution such a bull-roarer should have been almost universally connected with the fish, and that Paleolithic hunters in France as well as modern Eskimos should both have the idea of dentating its rims?*

Convincing though his argument may sound, of course, it is still more conjecture than proof.

One of the more commonly adopted diffusionist theories among comparative musicologists was the theory of culture circles (*Kulturkreislehre*), where cultural traits were thought to spread in ever-expanding circles from their point of origin, and where those traits most distant from the center were considered to be the oldest (Fig. 48.3).

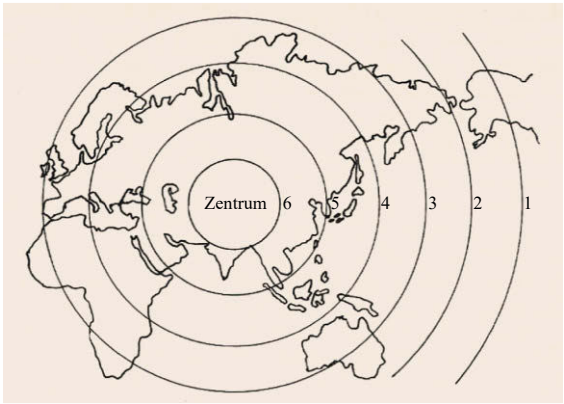


Fig. 48.3 The theory of culture circles (after [48.23, p. 29])

Diffusionist scholars theorized, among other things, that *the more widely an object is spread over the world, the more* primitive *it is* [48.22, p. 62]. *Sachs* claimed that the strung rattle, for instance, which is *used by modern primitives of a very low cultural standard as well as by Paleolithic hunters*, must be *among the earliest instruments* [48.22, p. 26]. In his 1929 *Geist und Werden der Musik Instrumente* [48.24], *Sachs* developed 23 historical strata for musical instruments based on distribution patterns, later adopted and adapted by *Hornbostel* for consideration of African musical instruments [48.25] and further refined by *Sachs* in 1940 [48.22]. Again, the most widely distributed instruments – like rattles – were considered to belong to the oldest strata.

Diffusionist theories not only allowed researchers to conceive of the age of certain musical or cultural traits; they also allowed them to hypothesize on the geographical route by which these traits traveled from one culture to another, and thus to build theories of cultural influence. In 1911, for instance, *Hornbostel* published a study on the tuning systems of xylophones in Southeast Asia and Africa [48.26]. Surprising similarities in the pitches, expressed in vibrations per second, led to a largely unsupported theory of monogenesis and cultural influence.

Concluding Remarks on Comparative Musicology

As *Savage* and *Brown* have noted [48.27, p. 158]:

> *One of the weaknesses of early comparative musicological work was a reliance on what we will call* remote comparison*, in which small numbers of songs from very distant regions were com-*

pared, often to support arguments of monogenesis *about long-distance similarity between regions. Such projects often involved the cherry-picking of particular songs that satisfied preconceptions of musical similarity.*

Moreover, the inherent racism in these evolutionary theories will probably be obvious to the twenty-first-century reader: it was not only allowed but encouraged to equate our Stone Age ancestors with our contemporaries in more so-called primitive cultures, in terms of their evolutionary stage. This supported a hierarchy of musics (and therefore cultures) ranging from primitive to advanced, thus justifying the goals of colonialism – and beliefs in Western superiority – through musical *science*.

For these reasons, the theoretical orientations of evolutionism and diffusionism, questions of origins, and concepts of universals were all eventually rejected by the academic community. And with that dismissal, the comparative approach to a large degree fell out of fashion also.

48.2 The Mid-Century Pendulum Swing: The Rise of Anthropology-Based Studies

While analysis of non-Western musics and the comparative approach did continue to have a few supporters and practitioners (discussed below) after the rejection of evolutionary theories, the second half of the twentieth century saw the development and eventual dominance of a very different form of ethnomusicological research. This was accompanied by a change in the name of the discipline from *comparative musicology* to *ethnomusicology* (a term coined by Jaap Kunst in 1950, later to lose the hyphen). As *Martin Stokes* has noted [48.28]:

Academic music theory and ethnomusicology parted company in the 1960s. Ethnomusicologists turned increasingly to Geertzian hermeneutics and ethnoaesthetics, viewing the application of western theoretical methodologies to non-western musics with concern and suspicion.

And given some of the emerging misgivings about what Western analytical techniques had wrought, this was an understandable reaction. Alongside the above-mentioned concerns of Eurocentric racism, scholars in the newly minted field of ethnomusicology feared that the kind of comparison undertaken in CM was a classic cart-before-horse blunder. *Mantle Hood* (1918–2005), who at mid-century promoted direct engagement with musics of the world through performance as a way to better understand music cultures and music systems (bimusicality), stated [48.20, p. 233]:

It seems a bit foolish in retrospection that the pioneers of our field became engrossed in the comparison of different musics before any real understanding of the musics being compared had been achieved.

He maintained that this approach had *led to some imaginative theories but provided very little accurate information* [48.29, p. 299].

Other objections to CM ran even deeper. As a new generation of American ethnomusicologists increasingly received training in anthropology, these scholars began questioning the very essence of the way we studied music. Suddenly anthropologists like Franz Boas (1858–1942), and his ideas of cultural relativism, began profoundly impacting the discipline as a whole. Scholars realized that music could not be studied in a vacuum; cultural factors *must* be taken into account so that *all people, in no matter what culture,* [would] *be able to place their music firmly in the context of the totality of their beliefs, experiences, and activities* [48.30, p. 3]. Enter *Alan P. Merriam*, an anthropologically trained ethnomusicologist, who developed a tripartite model for the study of music as culture, outlined in his influential 1964 book *The Anthropology of Music* [48.31]. According to Merriam's model, ethnomusicologists must begin by examining the culturally specific *concepts* about music as revealed to them by the practitioners themselves, use these concepts to inform observations about the *behaviors* of music-making, and only then tackle the music *sounds*, now from a deeper, more culturally sensitive place of understanding. This new approach would gain widespread appeal and support among many ethnomusicologists, particularly in the newly dominant North American branch of the field, calling into question the older practices of CM. As important English ethnomusicologist *John Blacking* (1928–1990) argued two years later [48.32, p. 218]:

A logical outcome of Merriam's approach to the study of music is surely the need for entirely new methods of analysis of music sound [. . .] If we accept the view that patterns of music sound in any culture are the product of concepts and behavior peculiar to that culture, we cannot compare them with similar patterns in another culture unless we

> *know that the latter are derived from similar concepts and behavior. Conversely, statistical analysis may show that the music of 2 cultures is very different, but an analysis of the cultural* origins *of the sound patterns may reveal that they have essentially the same meaning, which has been translated into the different* languages *of the 2 cultures.*

Scholars like *Charles Seeger* (1886–1979), alongside Blacking and Merriam, called for a discipline focused on *the advancement of knowledge of and about music* [and…] *the place and function of music in human culture* [48.33, p. 217]; as *Hood* put it, a study of music *not only in terms of itself but also in relation to its cultural context* [48.29, p. 298]. As mentioned, these new attitudes were especially strong in American ethnomusicology, where the teachings of Boas held considerable sway.

The putative goal of these mid-century scholars was to create a balanced, inclusive approach to the study of music – as the influential *Bruno Nettl* (b. 1930) [48.34, p. 26] described it:

> *a sort of borderline area between musicology (the study of all aspects of music in a scholarly fashion) and anthropology (the study of man, his culture, and especially the cultures outside the investigator's own background).*

A necessary shift in consciousness and understanding, this radical change in thinking among ethnomusicologists had a rather unfortunate, if inadvertent, side effect: a sudden and distinct disinclination among most ethnomusicologists to engage in deep musical analysis of any kind, and a subsequent paucity of analytical studies in the field. Ethnomusicology became a study of *music in culture* and *music as culture*, and its publications became overwhelmingly dominated by these anthropology-based studies. A survey by Timothy Rice of articles published in the journal *Ethnomusicology* from 1979 to 1986 showed that only 10% of these emphasized music analysis, while 34% focused on social processes and another 17% on individual processes [48.35, p. 476]. I performed a similar survey 16 years later with comparable results [48.36, pp. 108–109]:

> *In the annual list of dissertations and theses published by* Ethnomusicology *in the winter of 2001, only six of the over* 120 *papers had a distinct analytical bent. And, of the* 129 *articles published in that journal in the last ten years* [1993–2003]*, only fourteen contained major music-theory analyses. Many articles did not contain a single musical example.*

The effects of Merriam's pendulum swing, then, have been deep and long lasting. As *Nattiez* has noted [48.37, pp. 241–242]:

> *Since the 1960s ethnomusicology ha*[s] *become increasingly an* anthropology of music *under the influence of Merriam (1964) and Blacking (1973).* [… And] *because of the widespread assumption that only a knowledge of the cultural environment would permit a true understanding of music from an oral tradition, all analytical activity, which, it was suspected, substituted the tools of the Western researcher for the values and concepts of the native musician, began to disappear gradually from ethnomusicological monographs.*

There were, however, a few pockets of the ethnomusicological community doing analysis – and encouraging analysis – in the second half of the twentieth century. In the next few sections, we will discuss some of the major trends and a handful of the most important scholars and methods in these decades.

48.2.1 Analysis in a Relativist World

The few scholars left attempting close musical analysis in the second half of the twentieth century now had to do so with a new relativist understanding of the world. One of the more engaging analytical experiments undertaken in the 1960s was a symposium on transcription and analysis published in a 1964 issue of *Ethnomusicology* [48.38]. This study, which began as a colloquium presented at the Society for Ethnomusicology conference in 1963, was an unabashed demonstration of the subjectivity of transcription and therefore of analysis. Four respected scholars, Robert Garfias, Mieczyslaw Kolinski, George List, and Willard Rhodes were invited to transcribe and analyze a recording of a Hukwe song performed with musical bow. The participants were given some background information on the Hukwe people and a small description of the bow's playing technique. They were invited to transcribe and analyze the recording any way they saw fit, with very little instruction, no communication among them, and a stated desire for individual approaches. There are certainly elements common to the resulting transcriptions, such as agreements regarding pitch content and the timbral quality and rhythmic material of the bow. But as we can see in the synoptic view of the four transcriptions in Fig. 48.4, they are also utterly different. Two of the scholars choose to transcribe the whole recording while the other two focus in on smaller sections; one uses a graphic-style notation for the vocal melody; each decides on a different level of detail for the bow's pitch

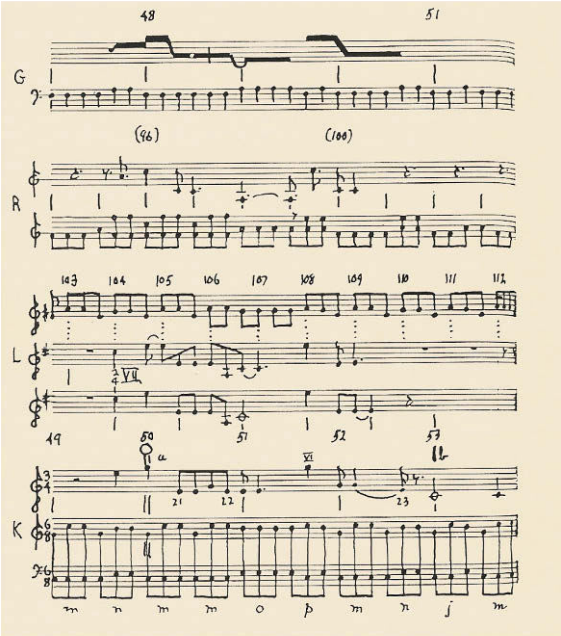


Fig. 48.4 Synoptic view of transcriptions from the 1964 symposium (after [48.38, p. 274])

content. Their accompanying analyses show even more breadth of perspective. While Garfias analyzes the piece from a culture-specific approach, both List and Kolinski apply the universalist sound-based theories still in vogue at the time, each focusing on his own particular areas of expertise: Kolinski alone, for instance, deals with tonal modality, which is of special interest to him.

The exciting thing about this symposium is that it addresses issues of subjectivity, recognizing *that in fact transcriptions bear within them the result of a transcriber's analytical understanding* of the music [48.39, p. 543], without then rejecting these methods out-of-hand. It is an experiment celebrating the relativism that turned many ethnomusicologists away from analysis in the later twentieth century, stating *one great strength of our Society lies in the varied individual approaches that are (and have been) made toward the data of our discipline* [48.38, p. 233]. It is with this attitude that the scholars discussed in the next sections moved forward with analysis in the latter half of the twentieth century, despite a virtual field-wide reaction against it.

48.2.2 Later Comparative Studies

Though comparative research largely passed out of vogue with evolutionary models, it did not entirely disappear in the second half of the twentieth century. In fact, as *Nettl* has noted, comparative study in the 1960s and 1970s actually *continu*[ed] *to occupy about the same proportion of research as it did before 1950; it* [simply] *received less attention and respect* [48.16, p. 67]. Two scholars forging new approaches to comparison in these decades were Mieczyslaw Kolinski (1901–1981) and Alan Lomax (1915–2002). Though studies of both men have been questioned and their approaches largely fallen into disuse, they were among the last to attempt to capture the full scope of the world's music in analytic terms. While Lomax was interested in cross-culturally examining singing style with enormous breadth of focus, Kolinski was concerned with deeply exploring the minutiae of very specific musical parameters across genres and cultures.

Alan Lomax's Cantometrics Project

Like many of the early comparative musicologists and music ethnologists, Alan Lomax was a prodigious collector of music, largely of the folk musics of Europe and North America. He was interested in comparing song styles and hypothesizing how differences among them might correlate with differences in the social structure of their respective cultures. *Lomax* is most well known for his multidecade project in what he called Cantometrics, or *measure of song* [48.40, 41]. An advanced form of trait listing analysis, Cantometrics measures 37 (later to be 36) distinct musical parameters of a song, from its various rhythmic and melodic features (e.g., regular versus irregular overall rhythm, melodic shape, type of polyphony, etc.) to ornamentation, level of group cohesiveness, and vocal qualities like nasality, vocal width, enunciation, and rasp. Lomax also designed a coding sheet on which each of these parameters could be judged. Figure 48.5 shows parameters 32–37 of the coding sheet – those dealing with vocal quality – including the various points on the rating scale for each parameter. The researcher would select the most appropriate point on each scale of the coding sheet and this would provide a *speedy characterization and classification* of the song's musical style [48.41, p. 8].

The same coding sheet could also be used to comparatively analyze two distinct music traditions, as we can see in Fig. 48.6. Here, the various vocal qualities of the so-called *African Gatherer* style are circled, and those of the *Urban East Asian* style are marked with rectangles. As *Lomax* asserts, these two profiles *define the extremes of the human stylistic range. There are, of course, other styles whose patterns fall along the middle of the profiling system* [48.40, p. 19].

Yet what really interested Lomax was determining how musical features might reflect social features. He sought to show, by comparing song measurements with preexisting ethnographic data, that [48.42, p. 97]

32) *Vocal Pitch* (Register)
(1) very high V-Hi (2) high Hi (3) mid Mid (4) low Low (5) very low V-Low
33) *Vocal Width*
(1) narrow NA (2) mid M (3) wide W (4) yodel Y
34) *Nasality*
(1) extreme or steady Ext (2) much Much (3) intermittent Int (4) some Some (5) little or no Ø
35) *Rasp*
(1) extreme Ext (2) great Gt (3) mid or intermittent Int (4) slight Sli (5) little or no Ø
36) *Accent*
(1) very forceful V-fo (2) forceful Fo (3) mid Mid (4) relaxed Re (5) very relaxed V-Re
37) *Enunciation*
(1) very precise V-Pre (2) precise Pre (3) moderate Mod (4) slurred Slur (5) very slurred V-Slur

Fig. 48.5
Lomax's Cantometrics coding sheet for vocal quality (after [48.40, p. 67])

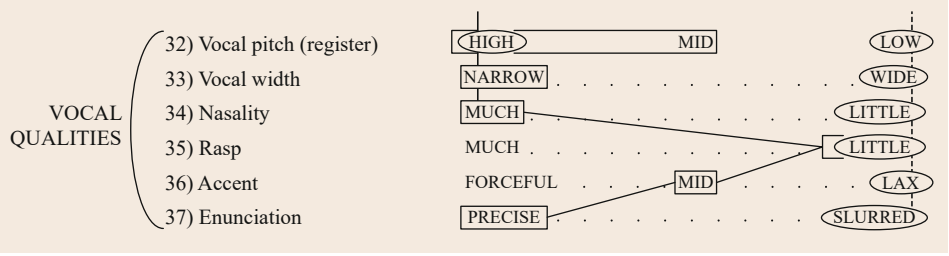


Fig. 48.6
Lomax's comparative analysis for vocal quality, *Urban East Asian* versus *African Gatherer* (after [48.40, p. 18])

> *song styles shift according to differences in productive range, political level, level of class stratification, severity of sexual mores* [...] *level of social cohesiveness,* [and so on].

Lomax's research team used a computer program to search for correlations between musical and societal features across the cultures they examined. Cultures were then organized into groups to create a geography of song style, and these larger profiles were analyzed in terms of their level of similarity in an attempt to show an evolution of style traditions. In the tree graph in Fig. 48.7, double lines represent the strongest cultural links, and the numbers on the bottom represent the proposed evolutionary stage of each culture group.

The Cantometrics system is flawed in a number of ways. It has been criticized for its evolutionist leanings and its unevenly distributed and often highly subjective parameters (how does one accurately judge a level of rasp, for instance?). Other critics have questioned its small sample sizes based on the mistaken assumption that most folk culture areas participate in a homogeneous style of music-making, and its use of *broad culture-areas as the basic units* of musical analysis [48.27, p. 155]. But, despite its not insignificant shortcomings [48.43, p. 101]

> *the Cantometrics system deserves credit for having moved vigorously in a direction previously uncharted: the description of singing style and of the nature of musical sound in general, things in the realm of what is usually called* performance practice.

Mieczyslaw Kolinski's Grand Schemes

While Alan Lomax was devising Cantometrics, Mieczyslaw Kolinski was developing a very different kind of grand scheme for comparative analysis. Yet, unlike Lomax, whose Cantometrics project is still cited in virtually all texts on ethnomusicology, Kolinski's methods are often only discussed *to recommend mostly against them* [48.3, p. 294]. Like earlier comparative musicologists, Kolinski was interested in discovering universals in music and comparing large bodies of data. He believed that all musics, no matter how diverse, could be [48.43, pp. 98–99]

> *subjected to comparison through a single classificatory system, a system reflecting and determined by the outer limits of and range of possibilities with the* [psychophysically rooted] *constraints* [of each musical style].

Kolinski published a series of articles through the 1950s, 1960s, and 1970s, each of which attempted a comprehensive examination of all possibilities for a given musical feature, from the 348 scalar and modal arrangements described in his *Classification of Tonal Structures* [48.44] to his detailed calculations of melodic movement [48.45, 46]. In these latter, Kolinski developed formulae and charts for comparatively analyzing melodic contours of diverse bodies of musical works.

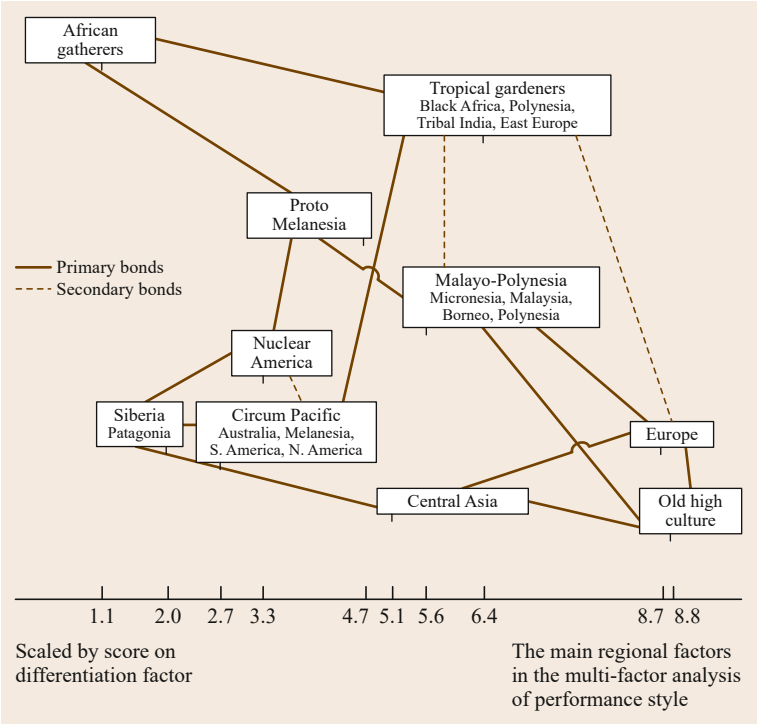


Fig. 48.7 Lomax's geography of song style (after [48.40, p. 35])

In *The General Direction of Melodic Movement* [48.45], for instance, *Kolinski* develops a method for comparatively analyzing different musics by charting the initial and final note of a melody in relation to its range. For a large body of works within a single tradition, Kolinski calculates what he calls *level formulae* – the average initial and final *melody level* expressed as a value from 0 to 100, where 0 is the lowest pitch and 100 the highest. For instance, among the songs of the Papago of Southern Arizona, Kolinski calculates the average initial pitch as being 64% of the piece's range higher than its lowest pitch, and the average final pitch as 32% higher. This gives the style a level formula of 64° : 32° and a level shift or pitch change of −32°, descending 32% of the full range over the course of the piece. By contrast, statistics of songs from the Menominee of Wisconsin exhibit a level formula of 89° : 4°, thus presenting a much sharper level shift of −85°.

Kolinski then expresses these level formulae in more detailed graphs showing the frequency distribution of songs with different initial and final levels across the repertoire, thus illustrating how the averages for level formulae were obtained. Figure 48.8 comparatively charts the distribution of initial pitch levels across the repertoire for Papago and Menominee songs. Kolinski divides the *x*-axis into 12 ranges – 0° (the lowest pitch in the range), 1–9°, 10–19°, 20–29°, and so on, up to 90–99°, and finally 100° (the highest pitch in the range). The *y*-axis shows the percentage of songs in the repertoires with an initial pitch in each of the given ranges. An average of all initial pitches in the graph gives the numbers from the level formulae above: 64° for the Papago songs and 89° for the Menominee songs. One can see that the Papago songs have a lower average initial level because a smaller percentage of songs begin on 100° than in the Menominee tradition and a larger percentage begin in the 20–29°, 30–39°, and 40–49° ranges.

The benefit of this sort of analysis is that it allows general melodic direction of different repertoires to be compared seemingly objectively regardless of the scale type or range of a given piece. Yet, for a number of reasons, Kolinski's methods have generally been rejected in the ethnomusicological community [48.47], [48.3, pp. 294–297]. This analysis of melodic direction, for example, takes no account of vastly different melodic contours presented with the same initial and final pitches and range. A melody beginning with a large leap up then proceeding with a slow descent back to the same pitch would generally be perceived as a descending melody where another differently contoured melody starting and ending on the same pitch could be perceived as wave-like or ascending. Yet these would be given the same level formula and thus erroneously analyzed as having the same melodic direction.

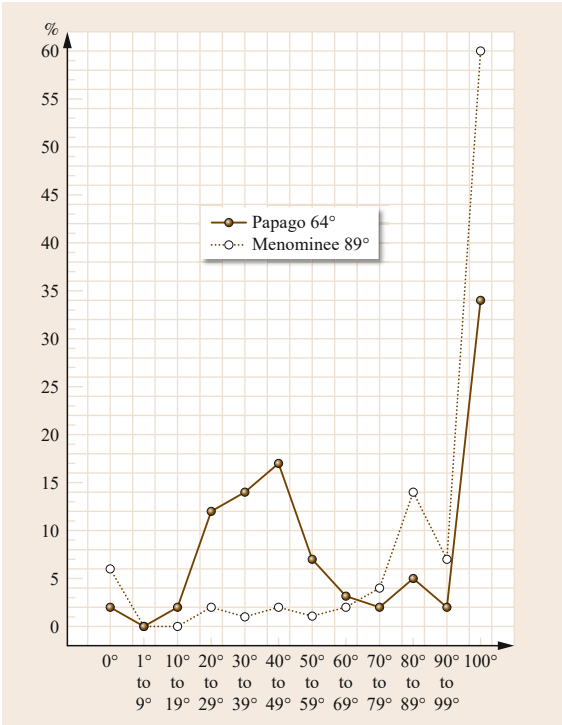


Fig. 48.8 Kolinski's comparative analysis of initial pitches; Papago versus Menominee songs (after [48.45, p. 241])

Kolinski's comparative study of tempo [48.48] has equally come under attack for its oversights and assumptions [48.47]. In this study, the author defines the tempo of a piece in terms of the number of notes per minute with no concern for the underlying beat of the music, as perceived by either cultural insider or outsider. Again, an attempt to provide a wide-spanning, seemingly objective comparative framework falls far short of its goal in terms of important perceptual details as well as *emic* (insider) considerations of the music.

Though much of the work of these mid-century comparativists has been largely rejected, we can appreciate the scope of their aspirations: using comparison, as *Kolinski* stated, as *an essential tool in* [our] *quest for a deeper insight into the infinite multifariousness of the universe of music* [48.49, p. 160].

48.2.3 Inspiration from Linguistics

Continuing the trend set by the late nineteenth-century comparative musicologists, many mid-twentieth-century analytical ethnomusicologists not predominantly interested in comparison turned to other disciplines for inspiration and guidance. The post-WWII era saw an academic atmosphere in which the study of symbols – and by extension the study of culture as a symbolic system – became increasingly popular across many disciplines, as Levi-Straussian structural anthropology [48.50] and semiotics [48.51] gained recognition. And for scholars attempting analysis in ethnomusicology, *an alternative to a behavioral approach to identifying symbols from a culture* [was] *to use language, the central symbolic code of humans, as a point of departure* [48.43, p. 305]. These scholars, then, turned to the work of structural linguists such as Ferdinand de Saussure, Roman Jakobson, and Noam Chomsky in the hopes of examining musical systems in the meticulous ways they had developed for examining linguistic systems. Scholars like George Herzog, the father of the American school of CM, were already asking questions about the overlap between musical and linguistic phenomena [48.52], but interest in applying linguistic *models* to musical analysis was something that developed through the 1960s and 1970s.

Saussurian structural linguistics is concerned, as the name suggests, with the underlying structures of language. Saussure made a distinction between the abstract linguistic system common to all speakers of a given language – what he called *langue* – and the discrete, unique utterances of individual speakers – *parole*. And for Saussure, the scientific study of language was a study of langue: a study of the rules behind the utterances. This he then divided into surface and deep structures. *Syntagmatic* analysis focused on the surface syntactical rules, such as the grammatically correct order of article, adjective, and noun in a given language: article-adjective-noun in English (the green house), article-noun-adjective in French (la maison verte), noun-adjective in Indonesian (rumah hijau). *Paradigmatic* analysis focused on the deeper paradigms, or preexisting sets of signifiers (letters, words, etc.), within the langue. This style of analysis could be applied to languages at several levels: the Roman alphabet is the paradigm from which English words are made; a full lexicon is a paradigm from which sentences are made, and this lexicon may be divided into paradigm sets according to things like word function (e.g., verbs). Paradigmatic analysis involves comparing the chosen signifiers (be they letters in a word, words in a sentence, etc.) with other signifiers that might have been chosen instead and to consider the significance of those choices. A common test in paradigmatic analysis is a commutation test, in which a signifier is selected and replaced with a different one to see whether or not the meaning – the signified – changes; this determines the distinctive signifiers within the language, defines the importance of those signifiers, and creates categories or paradigmatic classes of signifiers [48.53]. For instance, at the phonemic level, replacing *f* with *p* in an English word changes its mean-

ing; *fast* and *past* do not have the same meaning. But in Indonesian, *f* and *p* are very often interchangeable; *breath* can be spelled and pronounced either *nafas* or *napas*. This sort of commutation test becomes rather more subjective at higher levels of analysis, but its basic use is to discover the paradigms of the language: to determine at what level a change would affect the meaning; the signified.

The first musicologist to conceive of a paradigmatic analysis of music was *Nicolas Ruwet* [48.54]. He and his successors worked almost exclusively with notated music from the Western tradition. It was *Simha Arom*, in his studies of polyphonic traditions from the Central African Republic [48.55], who developed these techniques of classification and paradigmatic analysis most fully for ethnomusicological research. Arom used the natural commutation test of a cyclic music with repeated variation to discover culturally equivalent variations in diverse polyphonic musics, thus uncovering the paradigms of their musical languages. Through identifying common aspects between like patterns, Arom was able to decode a model for different rhythmic and melodic patterns and their possible variations. In Fig. 48.9, we see an example of what Arom calls a paradigmatic block: a group of rhythms that are deemed *culturally equivalent* among the Banda-Linda in the context of a specific ritual – rhythms that are freely interchangeable in their given position in the cycle.

A closely related form of linguistic analysis that also bled over into ethnomusicological research was *Chomsky*'s transformational-generative grammar [48.56], which aimed at [48.57, p. 163]

> *separat*[ing] *the grammatical sentences of a language from the ungrammatical ones and* [...] *provid*[ing] *a systematic account of the structure of grammatical sentences.*

An early example of the transformational model being applied to music is Edward Sapir's 1969 study of Diola-Fogny funeral songs [48.58]. Sapir incorporates insider (*emic*) names for different song structures while applying to those structures a transformational analysis that shows features shared by all the songs, and revealing each of the ways new variants can be derived. *Johan Sundberg* and *Björn Lindblom*'s 1976 article on generative theories [48.59] describes both Swedish nursery tunes and melodic variants of Swedish folksongs using generative rule systems. The authors then point to similarities between the two systems as potential guiding principles for composing simple melodies. They further hypothesize that connections between those musical rule systems and Chomsky and Halle's generative phonology of the English language reflect general characteristics of human cognitive capacities. *Vida Chenoweth* and *Darlene Bee*'s comparative-generative study of melodic structure in New Guinea [48.60] takes a different approach. The authors present three models of Awa song types (melodic structures) in the hopes of giving cultural outsiders simple visual tools with which to compose syntactically appropriate melodies in each structure. Figure 48.10 shows the authors' flow chart and associated formula for describing the simplest melody type. All syntactic units (notes) are described in terms of their interval relationship to the tonal center (TC), with *L* representing intervals below it (lower than) and *H* representing those above it (higher). As can be seen in the accompanying description, the flow chart and formula show all the possible choices that a composer can make, thus presenting a generative grammar:

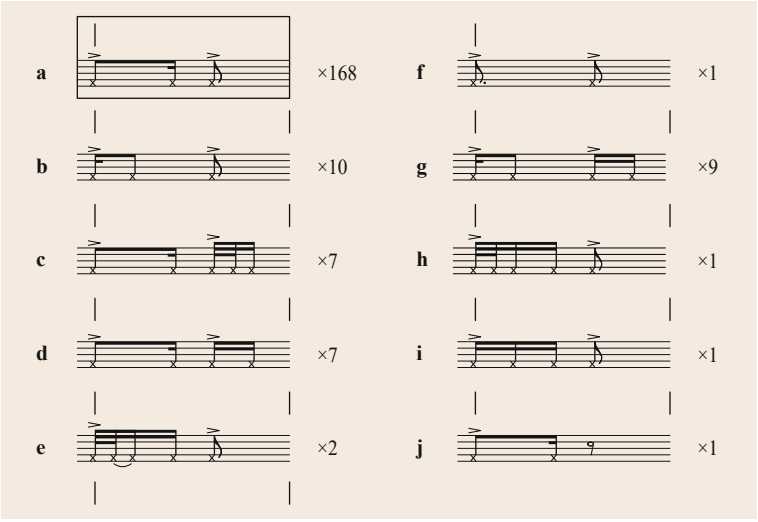
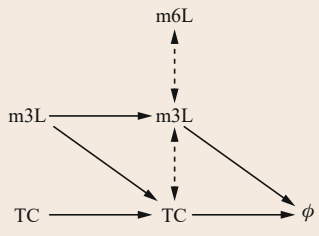


Fig. 48.9 Example of Arom's paradigmatic block (after [48.55, pp. 256–257])



{m3L; TC + TC} + {(TC) + m3L + (m6L + m3L); m3L + TC} $n + \phi$

An Awa melody composed on this pattern may begin on either a minor third lower than the tonal center or on the tonal center. If it begins on the tonal center there must be at least one repeat of the first note. The melody may end with either the tonal center or the minor third. The core of the melody consists of alternations between these two with an optional progression of a minor third plus a minor sixth and back to a minor third.

Fig. 48.10 Chenoweth and Bee's flow chart (*left*), formula (*top right*), and description (*bottom right*) for one Awa song type (after [48.60, p. 779])

a set of rules representing all the syntactically correct ways to generate new melodies in this melodic structure.

Concluding Remarks on Linguistic Approaches

Like all the methods discussed thus far, linguistic approaches to ethnomusicological analyses also have their critics. In an atmosphere of cultural relativism, one of the major points of contention is that these scholars, like many of their forebears in CM, were often unconcerned with how music related to culture. Further, ethnomusicologists like *Stephen Feld* [48.61] have questioned the practice of asserting equivalence between music and language without really demonstrating it, calling attention to studies that assume rather than explain the validity of using linguistic models for musical analysis. These detractors warn against what *Aniruddh Patel* calls the *distraction of superficial analogies between music and language* [48.62, p. 5], and promote the need for skepticism when considering studies that claim strict equivalence between them. As *Lerdahl* and *Jackendoff* argue [48.63, p. 5]:

> *Many previous applications of linguistic methodology to music have foundered because they attempt a literal translation of some aspect of linguistic theory into musical terms – for instance, by looking for musical* parts of speech*, deep structures, transformations, or semantics.* [...] *One should not approach music with any preconceptions that the substance of music theory will look at all like linguistic theory.*

Flawed though these approaches may have been, however, like the early comparative approaches, they helped to build a toolkit of methods that an ethnomusicologist could turn to – with a critical eye – to help with analysis.

48.3 Analysis in Modern Ethnomusicology

As we have seen in this chapter, analytical ethnomusicology is in its very nature subjective and imperfect. But, as *Judith Becker* has maintained (wisely if somewhat idealistically) of these many approaches attempted over the decades [48.64, p. 113]:

> *In each case, ethnomusicologists had the good sense to learn from these movements what was useful to us. In the process we became enriched theoretically and methodologically. Eventually, the realization sets in that this or that approach does not, as first assumed, solve all our problems or answer all our questions. But we move on with a richer arsenal of ways to think about music. We do not discard approaches that once seemed stunning in their ability to reveal insights into music and musical behavior, but subsequently were shown to be partial and vulnerable. Rather, we carry within ourselves, like a palimpsest, each theoretical methodological approach with which we have seriously engaged, and we are richer for it.*

This assertion, unfortunately, is the ideal, not yet the reality. Thus, in this final section, we will attempt to understand the still-uncertain status of analysis within the field – and perhaps look to the increased prominence that it *could* enjoy in the future – through a discussion of our field-defining rhetoric and an examination of a few of the new analytical methods evolving within the discipline.

48.3.1 The Still-Shaky Position of Analysis in Ethnomusicology

Despite the rocky history examined here, many ethnomusicologists do see the value of music theory and anal-

ysis, and a not insignificant number of us include them in our studies. As *Gabriel Solis* has argued [48.39, p. 533]:

> *Such music theory* [...] *should be – and, indeed, is – neither of limited value to questions about music as social practice, nor marginal to the discipline of ethnomusicology at large, but rather of central importance in practice and in principle to both.*

And *Michael Tenzer* contends [48.65, pp. 6–7]:

> *Once observed, sound patterns can be mobilized for many purposes: to demonstrate or inspire compositional depth or ingenuity, to discover an archetypical sound-structure model on which a music or repertoire is based, to symbolize or reflect a philosophy, social value or belief (of the analyst, the composer(s), performer(s), or their society), to reveal a historical process of change, to unearth suspected connections to music elsewhere, to embody a mathematical principle. Good analysis demystifies by cracking sound codes, better enabling the ear to collaborate with the mind in search of richer experience.*

That music theory and analysis are an integral part of ethnomusicological research thus seems an already accepted norm. Yet, since the so-called *great divide* of the discipline at mid-century, they have never regained equal footing with the more anthropology-based studies in the field. As *Jonathan Stock* observes [48.66, pp. 224–225]:

> *music analysis per se has been and continues to be questioned by the influential anthropological bloc within ethnomusicology* [...] *Whether the distrust by music anthropologists for what they see as the analysis of acontextual musical features is indeed academically well-founded may be contended with on a number of levels, but, in that it is analysis of the interplay between musical context, behaviour and sound that continues to dominate ethnomusicology* [...] *the anthropological view remains a powerful one in the shaping of ethnomusicological discourse.*

Mervyn McLean agrees, calling attention to the still-prevalent *notion that the earlier specifically musical approaches (especially those involving transcription and analysis) are no longer acceptable* [48.3, p. 330]. He argues fervently that a *complete* ethnomusicologist must be equally competent in both music and anthropology, and warns that [48.3, p. 333]

> *one of the most pernicious outcomes of the application (or more accurately misapplication) of Merriam's model* [of concept–behavior–sound] *is that it has led to the* sound *component of the model effectively becoming lopped off in favour of the remainder, because analysis of* mere sound *is supposed to obscure the reality of whatever it is that lies behind it.*

Indeed, many ethnomusicologists today still shy away from analysis or even warn against it. In his *Very Short Introduction to Ethnomusicology* [48.67], for instance, *Timothy Rice* expresses concern that the goals of music analysts like *Michael Tenzer* [48.65, 68], *reposition ethnomusicological study within a European, universalizing definition of what art is in contrast to the last thirty years of ethnomusicological work* [48.67, p. 62]. These analyses, in *Rice*'s estimation, are problematic because they elevate a Kantian view of music as art, which [48.67, p. 62]

> *has been used to valorize a limited, European view of art as always about beauty and to relegate non-European practices to a category of non-art or functional or applied art.*

These definitions of both music and analysis, however, seem narrow: Rice implies that analysts always consider music to be art, assumes that one would not wish to analyze something that was not high art, and by extension insinuates that analysis seeks to categorize and exclude, not to explore and discover.

Yet, though the role of music theory and analysis has been the subject of an ongoing negotiation since ethnomusicology's inception, to the credit of Tenzer, Stock, McLean, and other scholars passionately addressing the relative scarcity of analysis in our field, the tenor of this dialog has shifted in recent years. We have seen the establishment of the *Analytical Approaches to World Music* (AAWM) journal and conference, both of which rely heavily on the contributions of self-proclaimed ethnomusicologists. Further, *Tenzer*'s 2006 *Analytical Studies in World Music* [48.65] – a compendium of work from both ethnomusicologists and Western music theorists – was successful enough that Oxford University Press released a sister volume in 2011 [48.69]. And *Gabriel Solis* reports that in the *British Journal of Ethnomusicology* about 30% of recent articles have included significant music analysis [48.39, p. 535]. This is certainly an improvement over the aforementioned surveys of the North American *Ethnomusicology* journal by Rice and myself. Further, many recent published books and theses – like *Marc Perlman*'s *Unplayed Melodies* [48.70] and, in very different ways, *Thomas Turino*'s *Music as Social Life* [48.71] and *Michael Tenzer*'s *Gamelan Gong Kebyar* [48.68] – have also engaged in close musical analysis.

Yet there is still a significant imbalance of analytical versus anthropological studies; the pendulum has

yet to return to center after its mid-century swing. Solis attempts an explanation for this by shining a light on our own rhetoric surrounding analysis. He reproaches many ethnomusicologists for giving music theory and analysis short shrift in their field-defining metatheory, citing a recent article by *Timothy Rice* aimed at delineating the tasks and directions of the field [48.72]. While Rice, in his own research, does occasionally engage in what *Solis* (and I) would term a music theory and analysis of Bulgarian music, in his field-defining writing *he seldom metadiscursively acknowledges the ways that ethnomusicologists engage in theory about musical sound* [48.39, p. 531]. In fact, he even goes so far as to present an *off-hand dismissal of music theory as an exogenous and currently marginal practice to the discipline* [48.39, p. 532]. While Rice is perhaps extreme in his view of music theory's tangential role, *Solis* points to *the limited role of explicit language about music theory and analysis* [48.39, p. 546] and a general lack of deep engagement with these approaches in many of our major field-defining writings as a potentially dangerous precedent. He points to two seminal introductory texts on ethnomusicology as prime examples: *Helen Myers' Ethnomusicology* [48.73] and *Bruno Nettl*'s *The Study of Ethnomusicology* [48.43]. In these works, despite prefatory remarks embracing an integrated view of the field, *considerations of musical sounds become somewhat buried or separated from the main body of the work*. Myers, for instance [48.39, pp. 545–546],

> *includes major, very important articles* [... on transcription and analysis]*, but questions of music theory and analysis, and their role in answering anthropological questions are largely missing from other chapters in the volume.*

And ironically, though I have attempted here to present connections between anthropological and musicological concerns in ethnomusicology, the current chapter may simply be another in a long line of writings that separate analysis from the rest of the field.

Though many ethnomusicologists are now engaging in intensive music theory and analysis, then, the *metatheoretical ambivalence* [48.39, p. 541] facing those studies may perpetuate their second-class status. I am continually surprised at how many ethnomusicologists – when I talk to them at conferences – admit to being interested in music theory and analysis, because I do not hear about it in their papers. It seems almost a case of so-called pluralistic ignorance, where *virtually every member of a group or society privately rejects a belief, opinion, or practice, yet believes that virtually every other member quietly accepts it* [48.74, p. 161]. The end result of this metatheoretical silence on the topic, then, becomes a vicious cycle where each new generation learns from teachers who, themselves, were not trained to do analysis. Absent from our metatheory, music theory and analysis do not play a large role in most of our *Intro to Ethno* classes, requirements for graduate programs, or conference themes. As the graduate students of this generation become the next generation of teachers and writers, they will not necessarily think to include music theory and analysis approaches in their own writing or teaching; they have been acculturated into the thought-habits of the previous generation. Were music theory and analysis to enter our rhetoric, conferences, and grad programs in a more meaningful way, we would surely see more scholars engaging with music theory *as well as* social theory. And that deepening can only strengthen and enrich us. So how do we proceed from here?

48.3.2 A Panoply of Analytical Methods

Over the last half-century, ethnomusicologists interested in analysis have been grappling with the lessons of anthropological study and relativism, and we have learned that we cannot apply the same analytical approaches to every music culture. The paradigmatic analysis that Arom used to unravel underlying models and rules of variation for Central African polyphony, for instance, might quite aptly be applied to the Anlo-Ewe drumming examined by *Locke* [48.75]. Like those musics discussed by Arom, this tradition centers around short cyclic patterns varied by individual players, thus allowing the analyst to establish rhythmic models and equivalences. Or, as I have done for my analyses of Balinese improvised *arja* drumming [48.76], Arom's method could be emulated but then tweaked to suit the different style of cyclic variation in that tradition: one in which each set of paradigms does not trace to a single identifiable composition but rather a constellation of possible patterns then varied upon. For yet other styles, like the alap in a Hindustani rag performance, which does not have a steady rhythmic framework upon which equivalences may be found, paradigmatic analysis will probably be a much less fruitful approach. The analyst must choose from her/his toolkit of methods those that best suit the genre, so that, as *Marcia Herndon* puts it, *our conclusions about a particular piece* [can] *be checked by actual events within that piece, actual events within related pieces,* [or] *by informants* [48.77, p. 252].

Jonathan Stock has noted that detractors of analytical ethnomusicology *regularly censure music analysis as drawing on the values of the external scholar to the exclusion of those of the cultural insider*. He argues vehemently that *this criticism is intellectually unsatisfactory* and asserts that analysis lies [48.78, pp.189–190]

> *much closer to sensitive interpretation than dispassionate description when done well. Like ethnography, analysis is ultimately a means to develop and recontextualize understandings as we communicate them to our readers.*

Most of the analytical studies published in the last few decades fly in the face of the so-called anthropology-musicology divide, and many scholars are attempting to amalgamate music theory with social theory in their writings. *Thomas Turino* in his *Music as Social Life* [48.71] examines the ways that music can be socially meaningful. But the book is also [48.71, pp.1–2]

> *an introduction to some basic conceptual models that might help to illuminate why and how music and dance are so important to* [. . .] *fundamental aspects of social life,*

thus blending social theory with music theory. This more integrated approach has been at the center of many recent offerings in the field. In her article *Riffs, Repetition and Theories of Globalization*, for instance, *Ingrid Monson* presents a *detailed knowledge of musical practice* [as] *crucial in situating music within larger ideological and political contexts* [48.79, p. 33]. Marc Perlman agrees. In his work on Central Javanese *pathet* he has tried to show [48.80, p. 68]

> *how very close analyses of* the music itself *need not be divorced from issues of status, gender, or colonial history. The more detailed our technical analyses,* [he asserts,] *the more opportunities we will have to show how sounds and context are subtly intertwined.*

In my own studies of diffusion and change in Balinese *arja* drumming [48.81], a musical analysis of patterns from seemingly unconnected drummers was what led me to dig deeper, thus finding musicogenealogical links between them. Music analysis and social analysis, then, can be mutually beneficial, reinforcing and informing one another.

Anthropological approaches have also informed analysis in other ways. Methods of fieldwork have allowed many analytical studies to draw on concepts from local music theory. In his 2004 *Unplayed Melodies* [48.70], for instance, *Marc Perlman* relies heavily on the insights and expertise of master musicians from Java in order to unravel the implicit *unplayed* melodies upon which Central Javanese gamelan music is based. In his study of Balinese *gamelan gong luang* [48.82], *Wayan Sudirana*'s analyses are very much informed by ideas and terminology from interviews and casual conversations with individual musicians. And *Andrew McGraw*'s *Musik Kontemporer* [48.83] and *Radical Traditions* [48.84] equally draw upon ethnographic perspectives from contemporary Balinese composers and musicians.

In this more well-rounded world of analysis, each scholar chooses her/his own approach. What we have seen over the last 20 years is an explosion of new methods as well as attempts to resurrect and improve upon older ones. There is no style of analysis that dominates the scene. Some scholars will choose to apply well-established techniques to their studies, but often in surprising ways. *Jonathan Stock* [48.66], for instance, has suggested that Schenkerian analysis may be appropriate for some ethnomusicological studies, applying it to such diverse musics as the Kalasha praise songs of north-western Pakistan and Beijing opera. Others look farther afield, exploring interdisciplinarity in ethnomusicological analysis. Many of these scholars are revisiting Carl Stumpf's early interest in the psychology of music, borrowing theory and method from cognitive science. An important book from the 1990s incorporating musical analysis with cognitive studies is *Benjamin Brinner*'s *Knowing Music, Making Music* [48.85]. This insightful work sheds light on ways of knowing and concepts of competence through the lens of Javanese gamelan practice and performance. More recently, an article on improvisation in Indian classical singing by *Richard Widdess* [48.86] uses the concept of schemas – or cognitive maps – to examine processes of variation (*laykārī*). *Widdess* explores the idea that [48.86, p. 198]

> *both the singer's improvisation, and the listener's comprehension of it, depend on the simultaneous combination of pre-existing schemas, which enable the singer to arouse, and the listener to feel, varying degrees of uncertainty, expectancy and resolution.*

These are just two of many applications of cognitive science for analytical ethnomusicology.

In an environment where new approaches are encouraged and past approaches revisited and refined, there are far too many analytical methods in the modern toolkit to discuss them all here. In the following sections, then, we will briefly explore two quite different directions taken in recent years, both innovative, interdisciplinary approaches: the development of computational ethnomusicology and a return to comparative analysis.

48.3.3 Computational Ethnomusicology

The rise of new computer technologies has provided many opportunities for ethnomusicological analysis in recent decades. The use of computers and other machines by ethnomusicologists is by no means new – Charles Seeger developed the Melograph for graphic

transcription and the real-time analysis of pitch, dynamic, and timbre back in the 1950s – but a trend of using electronic tools for transcription and analysis did not follow for another 50 years. *Nicholas Cook* hypothesizes that much of this had to do with the strong reaction against comparative work in the mid-century [48.87, p. 103]:

> *Perversely* [...] *the possibility of computational approaches to the study of music arose just as the idea of comparing large bodies of musical data – the kind of work to which computers are ideally suited – became intellectually unfashionable.*

In the interim, some isolated studies have arisen, dispersed throughout various journals in many disciplines. Only in the last decade has there been a concerted effort to *gather relevant, high-quality research on computational methods and applications in ethnomusicology* into a unified, accessible forum for ethnomusicologists [48.88, p. 111]. Two important recent sources are *Tzanetakis* et al.'s introductory article in the *Journal of Interdisciplinary Music Studies* [48.89] and a special issue of the *Journal of New Music Research* (JNMR) from 2013 that presents several different studies on the topic (including [48.88, 90–92]).

Computational ethnomusicology, or CE, is understood to be *the design, development and usage of computer tools that have the potential to assist in ethnomusicological research* [48.89, p. 3], and beyond that, may have the capacity to develop [48.88, p. 111]

> theories *or* hypotheses *(not just* tools *as a spreadsheet or a statistical package can be) about processes and problems studied by traditional ethnomusicologists.*

Many recent CE studies are taking advantage of music information retrieval (MIR) techniques: tools that allow users to organize, search, and understand very large collections of data. Until recently, such techniques were largely used for popular applications, such as selecting music for personalized radio stations, query-by-humming [48.93], automatic genre classification, and tempo tracking. The potential benefits of such tools for the academic musical analysis of large data corpuses have only begun to be examined: analyzing pitch use or finding recurring rhythms or melodic patterns in collections far larger than a human could do by hand, transcribing for microtiming, analyzing minute physical performance gestures, searching for structural patterns in large collections, and so on. *Joren Six* et al.'s contribution to the JNMR special issue [48.92], for instance, introduces a modular software platform called Tarsos, which is designed to precisely extract pitch in recorded music of any tradition and, more importantly, to analyze pieces for their pitch distribution and organization. Figure 48.11 shows the various components of the Tarsos platform, which begins with a digital audio input, extracts pitch estimations, draws a histogram of pitch distribution, and finally creates an audio output of the result.

A tool such as this, if sensitively applied, could provide accurate cross-cultural pitch distribution analysis of a kind Ellis and Hornbostel could only dream.

Other applications of CE focus on the precise analysis of rhythm and timing, something for which computers are better suited than the human ear. In *Rhythm Analyzer* [48.94], *Kenneth Lindsay* uses computational analysis to measure *swing* in various recordings of Reggae, Afro-Brazilian music, and Western pop. *Cornelis* et al. [48.91] discuss the problems in using a computational approach to address tempo perception and levels of meter in Central African music. And even as early as 1993, *Jeffrey Bilmes* [48.95] was [48.89, p. 10]

> *work*[ing] *with multitrack audio recordings of Afro-Cuban percussion, extracting note timing, modelling note timing* [...] *and finally applying machine learning techniques to produce stylistically correct expressive timing for new phrases.*

Yet other applications of CE involve the classification of different pieces based on recurring rhythmic or melodic patterns. *Lin* et al. [48.96] have developed a method for determining the genre, or class, of a piece of music by identifying *significant repeating patterns* and then matching them to similar recurring patterns in other pieces of the same genre. Though their study shows varying levels of success depending on the genre being examined, its computational techniques provide interesting potential for more specific applications of pattern analysis. In his 2013 *Antipattern Discovery in Folktunes* [48.90], for instance, *Darrell Conklin* analyzes a large corpus of Basque folk tunes from different genres and identifies what he calls *antipatterns* – those patterns that are very common in most genres of folk tunes but rare or absent in one genre or set. In this way, he creates *negative association rules* for the identification of songs within certain sets of music, thus *predicting the absence of a genre based on the presence of a pattern* [48.90, p. 166].

Other studies using computational techniques attempt further levels of interdisciplinarity, interfacing with not only ethnomusicology but other fields of inquiry as well. In their *Computational Models of Symbolic Rhythm Similarity*, *Toussaint* et al. [48.97] test computational judgments of the relative similarities of different rhythms against human perception, thus touching not only on ethnomusicology, computer science, and mathematics, but also cognitive science.

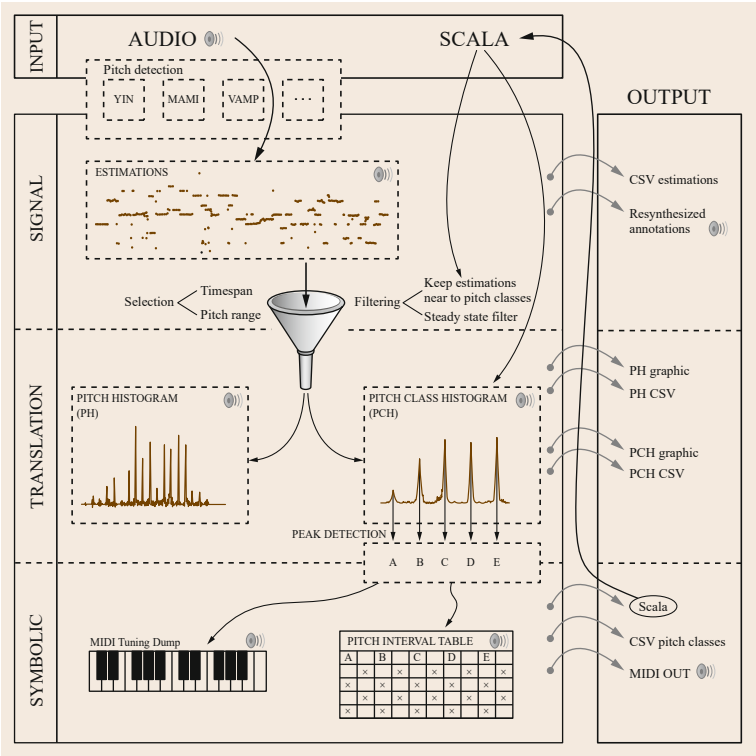


Fig. 48.11 The components of the Tarsos platform (after [48.92, p. 115])

It is hopefully clear from this limited examination that, though still a young field, computational ethnomusicology presents a number of new avenues for analysis and research. What will hopefully also be obvious, however, is that many computer-based tools are very useful for low-level analysis but, to be truly insightful, require refinement and, more importantly, interaction with music specialists. As *Tzanetakis* et al. state, we must [48.89, p. 12]:

> *actively seek interdisciplinary collaborations that include music scholars and technical researchers. Experimental results should generally be interpreted by music scholars with an understanding of the specific music(s) involved.*

Only in this way can domain-specific techniques and custom software be developed to tackle less general questions geared toward individual music traditions.

48.3.4 The New Comparative Musicology

The interest in interdisciplinary approaches currently fueling the growth of CE has also helped to reinvigorate comparative research. This renewed focus opposes the general perception in the field [48.3, p. 315]

> *that comparison is not only* old fashioned *but also in some sense unacceptable or even indefensible.* [... Of course] *most early comparative musicology was based on now long-discredited theories related to Kulturkreise and evolutionary ideas* [...] *No one wants to be tainted with such a brush, and there is irrational distaste for the whole idea of comparison as a result.*

Yet it seems premature to throw the baby out with the bath water. Comparative musicology was in many ways a groundbreaking, forward-looking, wonderfully experimental field. It brought together scholars of diverse specializations – both scientific and humanistic – and tried to grapple with some of the most basic questions of human existence: Why do we make music? What unites all musics and thus all peoples? Can we trace connections across the world through music? Perhaps these questions are impossible to answer in full, but without the latent Eurocentrism of Hornbostel's generation – or, more accurately, with a reconstructed, self-aware ethnocentrism cognizant of its own dangers – and with the benefit of a century's worth of insight and experience, old comparative methods and interests are leading to new analytical studies.

Comparative research in modern ethnomusicology takes many forms. In some cases, it involves the es-

tablishment of general theories that are cross-culturally relevant. *Michael Tenzer*, for instance, both in individual articles [48.98, 99] and edited works [48.65, 69] has established broadly applicable approaches for the classification and comparison of music according to its temporal organization. *Tenzer* asserts that *periodicity is really a universal, inseparable from a conception of music. This justifies choosing it as a framework* for analysis that may be applied across genres and across cultures [48.65, p. 23]. In his *Temporal Transformations in Cross-Cultural Perspective* [48.99], for instance, *Tenzer* examines concepts of time transformation cross-culturally by comparing instances of temporal augmentation in three distinct pieces of music – from Europe, South India, and Indonesia – and exploring how that temporal augmentation interacts both with the piece's musical structure and with the listener's orientation in musical time. He hypothesizes *that studying processes of* time transformation *cross-culturally can lead to musical and cultural insights* and, more broadly, *that cross-cultural research on musical temporality can lead to the discovery of some cognitive universals* [48.99, pp. 1–2]. In a very different way, *Judith Becker*'s *Deep Listeners* [48.100] sets up a theoretical basis for the cross-cultural analysis of high emotion and trance responses to music by drawing on new research from the fields of neuroscience and biology.

These sorts of universally applicable theories present their own unique set of problems. We know from earlier attempts at comparison that creating appropriate comparative categories can be daunting. At the first AAWM conference in 2010, Simha Arom raised this matter in a plenary session entitled *Ethnomusicology, Music Theory, and Music Analysis*. There, Arom made the contentious claim that if we wished to examine the concept of meter cross-culturally, we needed to provide the most neutral possible framework. Meter could not be defined as intrinsically multilevel and hierarchic as the growing consensus on the matter since the publication of *Lerdahl* and *Jackendoff*'s *A Generative Theory of Tonal Music* [48.63] asserted, because Arom did not see African music as having more than a bare skeleton of a hierarchy which, moreover, did not meet other accepted criteria for meter (typified by Lerdahl and Jackendoff's *preference rules*). Accounting for African meter in any encompassing definition meant abandoning what Arom suggested were biases favoring the structure of European musics, and therefore implied that asserting anything universal or cross-cultural meant only going for the minimal. At worst, this contention suggests that cross-cultural studies can only be shallow. Whether or not that is the case, the problem of defining meter nonetheless remains, and engaging in cross-cultural study demands reflection on these more difficult questions and the careful development of new solutions to old problems.

Two of the most avid proponents of reestablishing the comparative approach, *Patrick Savage* and *Steven Brown*, have proposed parameters and directions for a new comparative musicology [48.101], one that [48.27, p. 148]:

> *seeks to classify the musics of the world into stylistic families, describe the geographic distribution of these styles, elucidate universal trends in musics across cultures, and understand the causes and mechanisms shaping the biological and cultural evolution of music.*

This newer incarnation of the subdiscipline would avoid some of the pitfalls of earlier comparative studies by using larger sample sizes, focusing primarily on regional comparison, selecting appropriately sized units for analysis (e.g., individual songs, phrases, etc. as opposed to whole genres or cultures), creating culture-specific as well as more general and universal systems of classification, cross-culturally analyzing nonacoustic features of music – Merriam's *behaviors* and *concepts* – as well as music sounds, and so on [48.27].

The new comparative musicology demands mutually beneficial cross-disciplinary research, where anthropological, historical, biological, and linguistic studies, among others, help to inform discoveries in music research, but where the reverse is also true: where [48.27, p. 151]

> *knowledge of music's cultural evolution can be useful in illuminating* human history *more generally, including such phenomena as migration, colonialism, globalization, and other forms of cultural contact.*

In his provocative *Echoes of Our Forgotten Ancestors*, *Victor Grauer* [48.102] has attempted to do exactly that for the ancient populations of Africa. A collaborator on Lomax's Cantometrics project, Grauer uses song classification techniques to hypothesize that Pygmy and Bushmen groups, long isolated by geography, both maintain salient structural features of the same ancient musical practices. The striking similarities in their musics – interlocking polyphony made up of short phrases of continuous sound, repeated and varied, and sung in open-throated, blended voices with yodeling – he ascribes to a common ancestor. And he backs his theory with current research in genetic anthropology that has compared the DNA of Pygmy and Bushmen groups with that of other groups of black Africans and found

that the former two groups tend to have older mitochondrial and Y haplotypes [48.102, pp. 8–9]. Thus, the musical and the biological research support each other. And although Grauer's theories have not gone unquestioned [48.103], they provide a model for a very different kind of interdisciplinary analysis.

Much narrower in geographical scope but with the potential for large-scale insights is *Brown* et al.'s *Correlations in the Population Structure of Music, Genes and Language* [48.104]. This study examines the traditional group-level folksongs of nine indigenous populations in Taiwan using a modified Cantometrics system, called CantoCore [48.105]. The distance between each tradition in terms of its musical features is measured against existing information on mitochondrial DNA in the same populations to search for correlations. As we can see in Fig. 48.12, these measurements are most certainly connected.

In fact, this study shows stronger parallels between music and genes than between language and genes, and is thus [48.104, pp. 1,4]

> *the first quantitative evidence that music and genes may have coevolved* [and that music] *might serve as a useful marker to study human migrations and human origins more generally.*

Related to this sort of evolutionary study are the phylogenetic analyses undertaken by scholars like computer scientist Godfried Toussaint. Phylogenetics, which is a biological classification method that traces genetic links between organisms, is [48.107, p. 1115]

> *used to create a nested series of taxa based on homologous characters shared* [...] *by two or more taxa and their immediate common ancestor,* [and] *offers a means of reconstructing artifact lineages that reflect heritable continuity.*

For larger datasets in comparative study, computational methods are very useful, and *Toussaint* certainly uses these [48.97], but a smaller study will elucidate the method. *Toussaint*'s article on African ternary rhythm timelines [48.106] presents an analysis of ten 12-pulse, 7-stroke bell patterns from various African and African diaspora communities, shown in Fig. 48.13.

Toussaint measures the rhythmic similarity of these 12 patterns in terms of *swap-distance* – the minimum number of times one would have to move a note onset by one pulse in order to transform one pattern into another. These swap distances are then represented in a number of ways, including the swap distance matrix and phylogenetic SplitsTree shown in Fig. 48.14. Here the reader will notice a similarity between this style of distance tree and the one used by Savage et al. in Fig. 48.12. And, as *Toussaint* points out, with further research *these mechanisms may in turn shed light on the evolution of such rhythms* [48.106, p. 34].

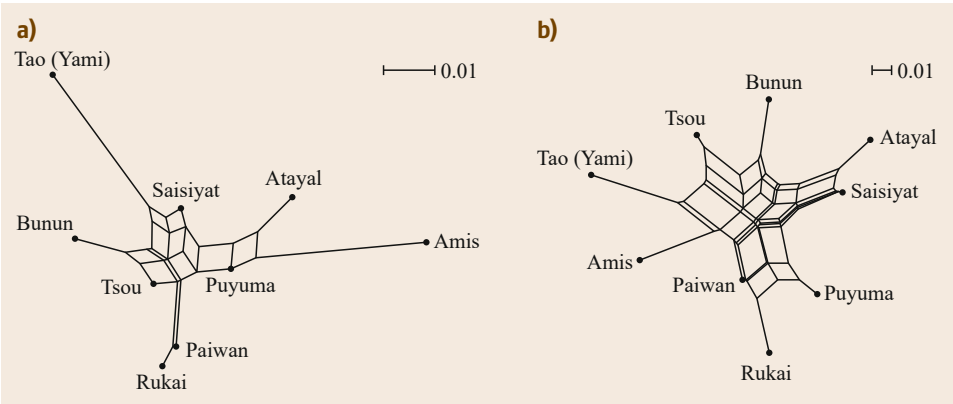


Fig. 48.12a,b Distance between musics **(a)** and genes **(b)** for nine indigenous Taiwanese populations (after [48.104, p. 3])

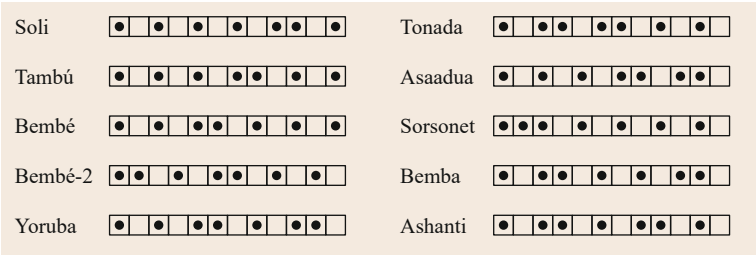
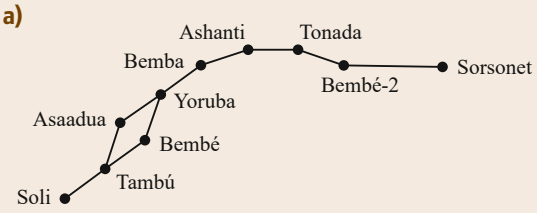


Fig. 48.13 Ten 12-pulse timeline rhythms to be analyzed by Toussaint (after [48.106])

a)



b)

	Soli	Tambú	Bembé	Bembé-2	Yoruba	Tonada	Asaadua	Sorsonet	Bemba	Ashanti
Soli		1	2	7	3	6	2	9	4	5
Tambú			1	6	2	5	1	8	3	4
Bembé				5	1	4	2	7	2	3
Bembé-2					4	1	5	2	3	2
Yoruba						3	1	6	1	2
Tonada							4	3	2	1
Asaadua								7	2	3
Sorsonet									5	4
Bemba										1
Ashanti										
Σ	39	31	32	35	23	29	27	51	23	25

Fig. 48.14a,b Toussaint's analysis of the patterns in Fig. 48.13 as a SplitsTree **(a)** and swap distance matrix **(b)** (after [48.106])

48.3.5 The Challenges of Interdisciplinarity

It seems, then, that many of the most innovative new directions in analytical ethnomusicology are interdisciplinary. So now the question becomes how to do work in a multidisciplinary project that speaks faithfully to each field involved. Whether or not we wish it were so, each field to some extent speaks its own language and has its own assumptions and priorities. And while scholars like *Victor Grauer* [48.102] synthesize ideas from quite divergent fields without the help of collaborators, Steven Brown, in his presentation at the 2012 AAWM conference, opined that most of us cannot hope to become experts in all of these branches; we must instead build working partnerships with specialists in other fields. Indeed, part of the reason that Savage et al.'s study of Taiwanese folksongs was so insightful was that, of the five authors contributing to the brief seven-page report, one worked in musicology, one in evolutionary genetics, one in psychology, neuroscience and behavior, and the final two in the medical profession in Taiwan. Specialists from divergent fields working toward a common analytical goal can lead to insights in all these different disciplines, and I believe that this open collaborative approach is the future of analytical ethnomusicology.

48.3.6 What Happens Now?

With all these opinions and methodologies at our disposal, the role of an ethnomusicologist now is to select, from an ever-expanding toolkit, those approaches best suited to the individual genre or piece under examination. Each may draw into focus (and conversely, obscure from view) certain aspects of the music, and thus the researcher may wish to attempt analysis of a given piece or genre through several approaches. Some of the earlier methods – general description or trait listing or classification – are perhaps more broadly applicable. Others, like Schenkerian, paradigmatic, or phylogenetic analysis will only be useful for the examination of certain genres or pieces or datasets. The researcher may even need to invent her/his own method, inspired by some of those explored here and always informed by the music theory (oral or otherwise) of her/his teachers and collaborators within the culture under examination. And, as we have learned from many of the more recent studies in analytical ethnomusicology, interdisciplinarity and an openness to collaboration with scholars in other fields – from anthropology to biology to computer science – will be what leads to some of our deepest analyses, our most exciting discoveries, and our most insightful new questions.

References

48.1 J. Kerman: How we got into analysis, and how to get out, Crit. Inq. **7**(2), 311–331 (1980)

48.2 A. Schneider: Comparative and systematic musicology in relation to ethnomusicology: A historical and methodological survey, Ethnomusicology **50**(2), 236–258 (2006)

48.3 M. McLean: *Pioneers of Ethnomusicology* (Llumina, Coral Springs 2006)

48.4 F. Densmore: *Chippewa Music* (Bulletin 45, Government Printer, Washington 1910)

48.5 F. Densmore: *Music of the Indians of British Columbia*, Vol. 27 (Da Capo, New York 1943)

48.6 A. Ellis: On the musical scales of various nations, J. Soc. Arts **33**, 485–527 (1885)

48.7 J.C. Fillmore: A woman's song of the Kwakiutl Indians, J. Am. Folk. **6**(23), 285–290 (1893)

48.8 J.C. Fillmore: The harmonic structure of Indian music, Am. Anthropol. **1**, 297–318 (1899)

48.9 B. Gilman: Hopi songs, J. Am. Ethnol. Archaeol. **5**, v–226 (1908), whole issue

48.10 E.M. von Hornbostel: African negro music, Afr. J. Int. Afr. Inst. **1**(1), 30–62 (1928)

48.11 J. Stanyek: Forum on transcription, Twent.-Century Music **11**, 101–161 (2014)

48.12 D. Christensen: Erich M. von Hornbostel, Carl Stumpf and the institutionalization of comparative musicology. In: *Comparative Musicology and Anthropology of Music: Essays on the History of Ethnomusicology*, ed. by B. Nettl, P.V. Bohlman (Univ. of Chicago Press, Chicago 1990) pp. 201–209

48.13 E.M. von Hornbostel: Die Probleme der vergleichenden Musikwissenschaft, Z. Int. Musikges. **7**, 85–97 (1905), Reprinted with English translation E.M. von Hornbostel: The problems of comparative musicology. In: *Hornbostel Opera Omnia*, ed. by K.P. Wachsmann, D. Christensen, H.-P. Reinecke (Martinus Nijhoff, The Hague 1975) pp. 249–270

48.14 E.M. von Hornbostel: Phonographierte Indianermelodien aus Britsch-Columbia. In: *Boas Anniversary Volume* (Stechert, New York 1906), Reprinted with English translation E.M. von Hornbostel: Indian melodies from British-Columbia recorded on the phonograph. In: *Hornbostel Opera Omnia* ed. by K.P. Wachsmann, D. Christensen, H.-P. Reinecke (Martinus Nijhoff, The Hague 1975) pp. 299–322

48.15 E.M. von Hornbostel: Phonographierte tunesische Melodien, Sammelbd. Int. Musikges. **8**, 1–43 (1906), Reprinted with English translation E.M. von Hornbostel: Tunisian melodies recorded on the phonograph. In: *Hornbostel Opera Omnia*, ed. by K.P. Wachsmann, D. Christensen, H.-P. Reinecke (Martinus Nijhoff, The Hague 1975) pp. 323–380

48.16 B. Nettl: *Nettl's Elephant: On the History of Ethnomusicology* (Univ. Illinois Press, Urbana 2010)

48.17 H. Helmholtz: *On the Sensations of Tone as a Psychological Base for the Theory of Music*, 3rd edn. (Longmans, Green, London 1895), translated by A.J. Ellis

48.18 R. Wallaschek: *Primitive Music: An Inquiry Into the Origin and Development of Music, Songs, Instruments, Dances and Pantomimes of Savage Races* (Longmans, Green, New York 1893)

48.19 R. Stone: *Theory for Ethnomusicology* (Prentice-Hall, Upper Saddle River 2008)

48.20 M. Hood: Music, the unknown. In: *Musicology*, ed. by F. Harrison, M. Hood, C. Palisca (Prentice-Hall, Englewood Cliffs 1963) pp. 215–326

48.21 C. Stumpf: *The Origins of Music* (Oxford Univ. Press, Oxford 2012), translated by D. Trippett

48.22 C. Sachs: *History of Musical Instruments: The Rise of Music in the Ancient World East and West* (Norton, New York 1940)

48.23 A. Schneider: *Musikwissenschaft und Kulturkreislehre: Zur Methodik und Geschichte der vergleichenden Musikwissenschaft* (Bad Godesberg, Bonn 1976)

48.24 C. Sachs: *Geist und Werden der Musikinstrumente* (D. Reimer, Berlin 1929)

48.25 E.M. Hornbostel: The ethnology of African sound instruments, Africa **6**, 129–157 (1933)

48.26 E.M. von Hornbostel: Über ein akustiches Kriterium für Kulturzusammenhänge, Z. Ethnol. **43**, 601–615 (1911)

48.27 P.E. Savage, S. Brown: Toward a new comparative musicology, Anal. Approaches World Music J. **2**(2), 148–197 (2013)

48.28 M. Stokes: Ethnomusicology IV: Contemporary Theoretical Issues, 10: Music theory and analysis. In: *New Grove Dictionary of Music and Musicians*, http://www.oxfordmusiconline.com.libproxy.mit.edu/subscriber/article/grove/music/52178pg4#S52178.4.10 (Oxford Music Online)

48.29 M. Hood: Ethnomusicology. In: *Harvard Dictionary of Music*, 2nd edn., ed. by W. Apel (Harvard Univ. Press, Cambridge 1969) pp. 298–300

48.30 A. Merriam: *Ethnomusicology of the Flathead Indians* (Aldine, Chicago 1967)

48.31 A. Merriam: *The Anthropology of Music* (Northwestern Univ. Press, Evanston 1964)

48.32 J. Blacking: Review of the anthropology of music, Curr. Anthropol. **7**(2), 218 (1966)

48.33 C. Seeger: *Studies in Musicology 1935–1975* (Univ. California Press, Berkeley 1977)

48.34 B. Nettl: *Folk and Traditional Music of the Western Continents* (Prentice-Hall, Englewood Cliffs 1965)

48.35 T. Rice: Toward the remodeling of ethnomusicology, Ethnomusicology **31**(3), 469–488 (1987)

48.36 L. Tilley: *Reyong Norot Figuration: An Exploration into the Inherent Musical Techniques of Bali*, MA Thesis (University of British Columbia, Vancouver 2003)

48.37 J.-J. Nattiez: Simha Arom and the return of analysis to ethnomusicology, Music Anal. **12**(2), 241–265 (1993), translated by C. Dale

48.38 N. England, R. Garfias, M. Kolinsky, G. List, W. Rhodes: Symposium on transcription and analysis: A Hukwe song with musical bow, Ethnomusicology **8**(3), 223–277 (1964)

48.39 G. Solis: Thoughts on an interdiscipline: Music theory, analysis, and social theory in ethnomusicology, Ethnomusicology **56**(3), 530–554 (2012)

48.40 A. Lomax: *Cantometrics* (Univ. of California, Berkeley 1976)

48.41 A. Lomax: *Folk Song Style and Culture* (American Association for the Advancement of Science, Washington D.C. 1968)

48.42 A. Seeger: Theory and method: Ethnography of music. In: *Ethnomusicology: An Introduction*, ed. by H. Myers (Macmillan, London, Houndmills 1992)

48.43 B. Nettl: *The Study of Ethnomusicology: Thirty-One Issues and Concepts, New Edition* (Univ. Illinois Press, Urbana, Chicago 2005)

48.44 M. Kolinski: Classification of tonal structures. In: *Studies in Ethnomusicology*, Vol. 1, ed. by M. Kolinski (Folkways, New York 1961) pp. 38–76

48.45 M. Kolinski: The general direction of melodic movement, Ethnomusicology **9**, 240–264 (1965)

48.46 M. Kolinski: The structure of melodic movement: A new method of analysis. In: *Studies in Eth-*

nomusicology, Vol. 2, ed. by M. Kolinski (Oak Publications, New York 1965) pp. 95–120
48.47 D. Christensen: Inner tempo and melodic tempo, Ethnomusicology **4**, 9–13 (1960)
48.48 M. Kolinski: The evaluation of tempo, Ethnomusicology **3**(2), 45–57 (1959)
48.49 M. Kolinski: Review of The Ethnomusicologist by Mantle Hood, Yearb. Int. Folk Music Counc. **3**, 146–160 (1971)
48.50 C. Lévi-Strauss: *Structural Anthropology* (Basic Books, New York 1963)
48.51 C. Morris: *Foundations of the Theory of Signs* (Univ. of Chicago Press, Chicago 1938)
48.52 G. Herzog: Speech-melody and primitive music, Musical Q. **20**(4), 452–466 (1934)
48.53 R. Barthes: *Elements of Semiology* (Jonathan Cap, London 1967), translated by A. Lavers and C. Smith
48.54 N. Ruwet: Methodes d'analyze en musicology, Rev. Belg. Musicol. **20**, 65–90 (1966), later translated by Mark Everist as: N. Ruwet: Methods of analysis in musicology, Music Anal. **6**(1/2), 11–36 (1987)
48.55 S. Arom: *African Polyphony and African Polyrhythm: Musical Structure and Methodology* (Cambridge Univ. Press, Cambridge 1991)
48.56 N. Chomsky: *Syntactic Structures* (Mouton, The Hague 1957)
48.57 H. Maclay: Overview. In: *Semantics*, ed. by D. Steinberg, L. Jakobovits (Cambridge University Press, Cambridge 1971) pp. 157–182
48.58 J.D. Sapir: Diola-Fogny funeral songs and the native critic, Afr. Lang. Rev. **8**, 176–191 (1969)
48.59 J. Sundberg, B. Lindblom: Generative theories in language and music description, Cognition **4**(1), 99–122 (1976)
48.60 V. Chenoweth, D. Bee: Comparative-generative models of a New Guinea melodic structure, Am. Anthropol. **73**, 773–782 (1971)
48.61 S. Feld: Linguistic models in ethnomusicology, Ethnomusicology **18**(2), 197–217 (1974)
48.62 A.D. Patel: *Music, Language and the Brain* (Oxford Univ. Press, Oxford 2008)
48.63 F. Lerdahl, R. Jackendoff: *A Generative Theory of Tonal Music* (MIT Press, Cambridge 1983)
48.64 J. Becker: Response to "Consilience Revisited", Ethnomusicology **56**(1), 112–117 (2012)
48.65 M. Tenzer (Ed.): *Analytical Studies in World Music* (Oxford Univ. Press, Oxford 2006)
48.66 J. Stock: The application of Schenkerian analysis to ethnomusicology: Problems and possibilities, Music Anal. **12**(2), 215–240 (1993)
48.67 T. Rice: *Ethnomusicology: A Very Short Introduction* (Oxford Univ. Press, Oxford 2014)
48.68 M. Tenzer: *Gamelan Gong Kebyar: The Art of Twentieth-Century Balinese Music* (Univ. of Chicago Press, Chicago 2000)
48.69 M. Tenzer, J. Roeder (Eds.): *Analytical and Cross-Cultural Studies in World Music* (Oxford Univ. Press, Oxford 2011)
48.70 M. Perlman: *Unplayed Melodies: Javanese Gamelan and the Genesis of Music Theory* (Univ. California Press, Berkeley 2004)
48.71 T. Turino: *Music as Social Life: The Politics of Participation* (Univ. of Chicago Press, Chicago 2008)
48.72 T. Rice: Ethnomusicological theory, Yearb. Tradit. Music **42**, 100–134 (2010)
48.73 H. Myers (Ed.): *Ethnomusicology: An Introduction* (Norton Grove, New York 1992)
48.74 D.A. Prentice, D.T. Miller: Pluralistic ignorance and the perpetuation of social norms by unwitting actors, Adv. Exp. Soc. Psychol. **28**, 161–209 (1996)
48.75 D. Locke: *Drum Gahu: A Systematic Method for an African Percussion Piece* (White Cliffs Media, Crown Point 1987)
48.76 L. Tilley: *Kendang Arja: The Transmission, Diffusion, and Transformation(s) of an Improvised Balinese Drumming Style*, Ph.D. Thesis (University of British Columbia, Vancouver 2013)
48.77 M. Herndon: Analysis: The herding of sacred cows?, Ethnomusicology **18**(2), 219–262 (1974)
48.78 J. Stock: New directions in ethnomusicology: Seven themes toward disciplinary renewal. In: *The New (Ethno)musicologies*, ed. by H. Stobart (The Scarecrow, Lanham 2008)
48.79 I. Monson: Riffs, repetition, and theories of globalization, Ethnomusicology **43**(1), 31–36 (1999)
48.80 M. Perlman: The social meaning of modal practices: Status, gender, history, and pathet in central javanese music, Ethnomusicology **42**(1), 45–80 (1998)
48.81 L. Tilley: Dialect, diffusion, and balinese drumming: Using sociolinguistic models for the analysis of regional variation in Kendang Arja, Ethnomusicology **58**(3), 481–505 (2014)
48.82 W. Sudirana: *Gamelan Gong Luang: Ritual, Time, Place, Music, and Change in a Balinese Sacred Ensemble*, Ph.D. Thesis (University of British Columbia, Vancouver 2013)
48.83 A.C. McGraw: *Musik Kontemporer: Experimental Music by Balinese Composers*, Ph.D. Thesis (Wesleyan University, Middletown 2005)
48.84 A.C. McGraw: *Radical Traditions: Reimagining Culture in Contemporary Balinese Music* (Oxford Univ. Press, Oxford 2013)
48.85 B. Brinner: *Knowing Music, Making Music: Javanese Gamelan and the Theory of Musical Competence and Interaction* (Univ. of Chicago Press, Chicago 1995)
48.86 R. Widdess: Schemas and improvisation in Indian music. In: *Language, Music and Interaction*, ed. by R. Kempson, C. Howes, M. Orwin (College Publications, London 2013) pp. 197–209
48.87 N. Cook: Computational and comparative musicology. In: *Empirical Musicology: Aims, Methods, Prospects*, ed. by E. Clarke, N. Cook (Oxford Univ. Press, Oxford 2004)
48.88 E. Gómez, P. Herrera, F. Gómez-Martin: Computational ethnomusicology: Perspectives and challenges, J. New Music Res. **42**(2), 111–112 (2013)
48.89 G. Tzanetakis, A. Kapur, W.A. Schloss, M. Wright: Computational ethnomusicology, J. Interdiscip. Music Stud. **1**(2), 1–24 (2007)
48.90 D. Conklin: Antipattern discovery in folk tunes, J. New Music Res. **42**(2), 161–169 (2013)

48.91 O. Cornelis, J. Six, A. Holzapfel, M. Leman: Evaluation and recommendation of pulse and tempo annotation in ethnic music, J. New Music Res. **42**(2), 131–149 (2013)

48.92 J. Six, O. Cornelis, M. Leman: Tarsos, a modular platform for precise pitch analysis of western and non-western music, J. New Music Res. **42**(2), 113–129 (2013)

48.93 R. Dannenberg, W. Birmingham, G. Tzanetakis, C. Meek, N. Hu, B. Pardo: The MUSART testbed for query-by-humming evaluation, Comput. Music J. **28**(2), 34–48 (2004)

48.94 K.A. Lindsay: *Rhythm Analyzer: A Technical Look at Swing Rhythm in Music*, Masters Thesis (Southern Oregon University, Ashland 2006)

48.95 J. Bilmes: *Timing is of the Essence: Perceptual and Computational Techniques for Representing, Learning, and Reproducing Timing in Percussive Rhythm*, Masters Thesis (MIT, Cambridge 1993)

48.96 C.-R. Lin, N.-H. Liu, Y.-H. Wu, A. Chen: Music classification using significant repeating patterns. In: *Proc. Int. Conf. Database Syst. Adv. Appl. DASFAA'04, Jeju Island*, ed. by Y.-J. Lee, J. Li, K.-Y. Whang, D. Lee (Springer, Berlin, Heidelberg 2004) pp. 506–518

48.97 G. Toussaint, M. Campbell, N. Brown: Computational models of symbolic rhythm similarity: Correlation with human judgments, Anal. Approaches World Music J. **1**(2), 380–430 (2011)

48.98 M. Tenzer: Generalized representations of musical time and periodic structures, Ethnomusicology **55**(3), 369–386 (2011)

48.99 M. Tenzer: Temporal transformations in cross-cultural perspective: Augmentation in baroque, carnatic, and balinese music, Anal. Approaches World Music J. **1**(1), 1–23 (2011)

48.100 J. Becker: *Deep Listeners: Music, Emotion, and Trancing* (Indiana Univ. Press, Bloomington 2004)

48.101 P. Savage, S. Brown: Comparative Musicology, http://www.compmus.org

48.102 V.A. Grauer: Echoes of our forgotten ancestors, World Music **48**(2), 5–58 (2006)

48.103 J. Stock: Clues from our present peers? A response to Victor Grauer, World Music **48**(2), 73–91 (2006)

48.104 S. Brown, P.E. Savage, A.M.-S. Ko, M. Stoneking, Y.-C. Ko, J.-H. Loo, J.A. Trejaut: Correlations in the population structure of music, genes and language, Proc. R. Soc. B **281**, 1–7 (2013)

48.105 P.E. Savage, E. Merritt, T. Rzeszutek, S. Brown: CantoCore: A new cross-cultural song classification scheme, Anal. Approaches World Music J. **2**(1), 87–137 (2012)

48.106 G. Toussaint: Classification and phylogenetic analysis of African ternary rhythm timelines. In: *Proceedings of BRIDGES: Mathematical Connections in Art, Music and Science* (University of Granada, Granada 2003) pp. 25–36

48.107 M.J. O'Brien, J. Darwent, R.L. Lyman: Cladistics is useful for reconstructing archaeological phylogenies: Palaeoindian points from the southeastern united states, J. Archaeol. Sci. **28**, 1115–1136 (2001)

49. Musical Systems of Sub-Saharan Africa

Simha Arom

The following chapter is a brief overview of Sub-Saharan music traditions based on our current knowledge. It seeks to fill a gap that exists in the studies done thus far because, to the best of our knowledge, no inventory of the essential parameters that go into the constitution of this region's musical systems has been made. This is probably due to the fact that ethnomusicologists specializing in African music as well as many others give priority to the cultural or anthropological context in which music practices have a function, rather than considering the subject itself – music – as a system.

In light of this situation, and while I am aware of the weaknesses and imperfections such an approach will have difficulty avoiding, it nevertheless seems worthwhile to make an attempt.

This chapter is not intended to be exhaustive but is rather, beyond its specific content, an attempt to provide a tool for study that is easy to use and practical to all those who are interested in and curious about the *grammar* that underlies the Sub-Saharan African music traditions: scholars, teachers, students, not to mention those on the creative end, notably jazz musicians and composers, and more generally all people interested in music from beyond Europe and those interested in the *manufacturing* of African folk music.

In my mind, the value of this work extends beyond the text to include collections of references and the selective but very rich *bibliography*. These collections are divided into the following categories: general characteristics of African music, the acquisition of musical knowledge, taxonomy, scales, time organization, form and structure, variations, polyphonic techniques in general, hocket and polyrhythm. Each one lists authors whose work covers these categories. Readers can then consult the bibliography for further study of an author or subject presented in the text.

Traditional Sub-Saharan music can generally be described as follows:

- It is transmitted *orally* with no written support.
- It is collective: the community as a whole ensures its perennity.
- It is most often *anonymous and undated.* We don't know *who* created it or *when*.
- It is *functional* – or more precisely, *circumstantial*. It is not conceived for use outside its sociocultural context.
- It is not the subject of abstract speculations by the people who practice it, so the theory that underlies the music is essentially *implicit*.

For *general characteristics* of African music, see [49.1–13].

The diversity of traditional music coincides with the diversity of the ethnic groups and subgroups as reflected in their specific languages and dialects. Each ethnic group or subgroup has its own dialect or manner of speaking that distinguishes it from all others. This same diversity exists in music. There are as many musical idioms as there are ethnic communities or languages and dialects in Sub-Saharan Africa.

The acquisition of the basis of musical knowledge goes hand-in-hand with language learning, ie, it occurs empirically. In the first phase, passive newborn babies absorb the music that surrounds them; as their motor functions develop, children progressively join in the various musical activities, first by clapping their hands in rhythm, then by singing, and finally by trying to play an instrument (for more concerning *music learning in traditional societies*, see [49.5, 12, 14–22]).

The consistency of African traditional music is manifested within each ethnic community, firstly through its *functionality* – in other words the social and/

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_49

or religious circumstances with which it is intimately linked – through the groups that are required for its performance, and lastly through the elements that constitute its systemic framework.

It seems important to immediately point out that functionality and musical systems are closely linked within ethnic communities. The relationships between the repertoires and the circumstances can be of three types.

Each circumstance that requires musical accompaniment corresponds to a particular repertoire, characterized by a specific number of songs, a predetermined instrumental ensemble and a set of rhythmic or polyrhythmic formulae that belong only to the music associated with this circumstance. Each repertoire is thus the equivalent of a musical category distinct from all other categories (on the subject of *categorization*, see [49.10, 12, 14, 18, 21, 23–37]).

One given situation may draw on several repertoires. Several repertoires may use the same rhythmic or polyrhythmic formulae. Within the same language group, people can share repertoires with the same name and functions. In some regions, the repertoires and the contexts with which they are associated are part of a broad dissemination network. Two or several populations with different languages share one or several repertoires with names and performance contexts that are the same.

The constituent elements of this music include the following parameters: pitch, time structure, timbre, form, structure, performance procedure and technical process. We will examine them one by one.

The *pitch system* is obviously related to the musical scales, which vary widely. They range, according to region or ethnic group, from the anhemitonic pentatonic system to the equiheptatonic scale, pentatonic intervals of whole or equidistant steps, tetratonic, hexatonic, diatonic scales and still others, without overlooking those that have movable degrees (depending on whether the melodic movement is upward or downward). In many cases, the same community uses various types of scales that are respectively linked to different repertoires (for more regarding *scale systems*, see [49.6, 9, 11, 12, 24, 25, 27, 31, 33, 38–58]).

The *structuring of time* is based on music that is almost always measured, i. e., within which the durations have strictly proportional relationships.

The great majority of forms of African music are based on a strict *periodicity*, determined by the occurrence of similar events at identical positions in a temporal cycle.

Periods are most often subdivided into a constant number of isochronic – i. e., equidistant – pulses, which constitute the *organic* standard with respect to which all of the durations are organized both horizontally – in the case of monody –, and vertically, within polyphonic and/or polyrhythmic structures. The *aksak* rhythm is the only exception to this rule. Borrowed from Ottoman music theory, this Turkish term means *limping*, i. e., *irregular*. It designates a rhythmic system in which pieces or sequences *performed in a lively tempo* are based on the continuous repetition of a module characterized by the uninterrupted repetition of alternating binary and ternary rhythmic cells (such as $2+3$, $2+2+3$, etc.). These cells are most often grouped in odd numbers. These groupings and the manner in which they are arranged determine the form, the structure and the articulation of the *aksak*. In Sub-Saharan Africa, *aksak* is used by the Mofu (Cameroon), the Bulia (Democratic Republic of Congo and Republic of Congo) and the Ju/'Hoansi-San (Namibia).

It can be seen that this time organization differs considerably from the Western conception, in that it most often ignores the intermediate level of the *measure*, i. e., a regular accented scheme based on the alternation of a *strong* beat with one or several *weak* beats.

The pulse is subdivided into smaller values, according to a binary or ternary principle. These are the *minimum operational values* through which all durations – in both vertical and horizontal terms – are organized (for more concerning the *structuring of time*, see [49.2–5, 10, 12, 14, 18, 23, 30, 31, 33, 34, 37, 40, 41, 46, 48–50, 54, 59–87]). However, in certain geocultural areas, much of this music contains two simultaneous pulses in a ratio of $2:3$, one involving two minimum values, the other three. This is the *hemiola* principle.

Timbre refers to the *tools* that contribute to the materialization of the music, i. e., the voice and the instruments.

Certain repertoires require only male voices, while others make exclusive use of women's voices. Others still are reserved for only children's voices. Collective music most often combines both voices and instruments.

The combinations of instrumental timbres may be put into three categories, described below in descending order of frequency of use.

- Different types of instruments played together (a xylophone accompanied by two slit drums, a rattle and a pellet, for instance).
- Various instruments of the same type, e.g., two or three xylophones, accompanied by one or several different percussion instruments, such as idiophones and membranophones (as above).
- Lastly, groups in which all of the instruments are of the same type, but of different sizes. This is often the case for ensembles of horns or whistles, which can

include up to twenty instruments, such as xylophone and drum ensembles.

Let us remember that in a traditional environment, the instrumental group, i. e., its specific composition, considered in terms of the instruments that it includes is most often closely associated with a particular circumstance or set of circumstances.

With regard to *forms and structures*, it is important to draw a distinction between music that allows for no modification and music that is renewed and recreated with each performance. The former is very rare and almost always associated with events of a historical nature or part of primordial liturgies. It is *frozen* music, as opposed to the great majority of forms that allow, within extremely strict periodic, metric and rhythmic structures, a degree of freedom left to the performers, thereby allowing them to produce numerous variations, and in some cases genuine improvisations (for more regarding *forms, structures and patterns*, see [49.2, 4–6, 10–12, 14, 18, 23, 34–37, 41, 46, 49, 54, 56, 60, 62, 74, 80, 88–99]).

A second distinction must be made between music that is strictly cyclical, i. e., repetitive, which can be referred to as *closed*, and forms of music that can be called *open* because its performance is based on the irregular alternation of diverse juxtaposed periodic structures, leading to very substantial works. But regardless of whether they are open or closed, these forms of music – with rare exceptions – always leave the performers a wide range of freedom within which their spontaneous creativity can be expressed (for more regarding variation, see [49.2–5, 9, 12, 14, 20, 21, 23, 40, 42, 45, 46, 49, 50, 63, 64, 66, 68, 71, 74, 77, 83, 96, 100–108]).

The *modalities of performance* refer to the distribution of roles within a musical event, for example:

- Between a vocal soloist and a choir
- Between a singer and an instrument on which the singer accompanies himself or herself
- For more diversified groups, between the sung parts and the parts of the instrument that accompany them.

In the first case, the most frequent situation is a regular alternation between the two protagonists: either the choir exactly repeats the melodic unit sung by the soloist, in which case we say that the singing is *antiphonal*, or the choir completes it, making the music *responsorial* in nature.

In general, a singer who accompanies himself or herself on a melodic instrument also divides the period of the song into several segments, some assigned to their own voice, which the instrument of course supports, and other segments which respond to it. An ongoing dialogue develops between these two elements, instilled with the same spirit.

Such a distribution of roles is not random. On the contrary, it stems from a perfect rationale based on the *principle of complementarity*. By corroborating the coherent segmentation of the musical discourse, the performance procedures constitute a precious clue for the discovery of its underlying structure.

With respect to the *technical processes*, we must make a distinction between *monody* and *plurivocality*, i. e., polyphony in all of its aspects.

Monody can be vocal and/or instrumental. When the singing is duplicated by a melodic instrument, it is frequent that the melodic and rhythmic coincidence between the two is not absolute. This sporadic phenomenon – which is often deliberate – generates the most rudimentary form of plurivocality called *heterophony*.

Polyphony in the strict sense involves one of the following processes:

- *Overlapping/tiling* (*tuilage*) This refers to the sporadic overlapping of two musical sources for which the realization follows a principle of alternation between two protagonists (soloist and choir, two soloists, or two choirs) in antiphonal and responsorial singing, which is very common in Africa. There is tiling when one of the protagonists intones before the intervention of the preceding one is finished.
- *Drone* This refers to one or several notes that act as a base for one or several melodies performed simultaneously.
- *Parallel movement* This is very widespread in vocal music; it proceeds mostly with either intervals of fifths and fourths, or with sixths and thirds.
- *Divergent motion* This can be either *homophonic*, when the rhythmic articulation is the same for all of the parts present, or *contrapuntal*, when the various parts are rhythmically independent. The use of counterpoint, while relatively widespread in instrumental music, is quite rare in vocal music. It is used however by Pygmy and San people (in Central Africa, Namibia, and Botswana) as well as by the Dorze (Ethiopia) and the Bateke (Gabon) (for more on *polyphony in general*, see [49.1, 6, 8, 11–14, 18, 20, 23, 29–32, 35, 42–46, 49, 50, 54, 60, 66, 69, 73, 78, 83, 85, 97, 101, 102, 104, 106, 109–131]).
- *Hocket* This can be used in three ways:
  - *Strictly vocal*, or *a cappella*
  - *Instrumental*, with instruments of the same type (horns or whistles), each of which usually produces the same sound at a predetermined pitch

  - *Vocal-instrumental*, where each of the performers alternates between sung sounds and instrumental sounds (concerning the *hocket technique*, see [49.11, 12, 14, 23, 29, 32, 40, 42, 46, 78, 96, 102, 108, 116, 127, 129, 132–134]).
- *Polyrhythm* Most African folk music is supported by instrumental ensembles composed of instruments that are strictly percussive in which the pitch relationships are not relevant. They provide the periodic and the metric framework for the superimposed melodic events and the music that these ensembles play is based on the *crossing of diverse rhythmic patterns in which the reciprocal accents are offset.* The result of this is an extremely dense – but always coherent – entanglement of antagonistic rhythms and through this a polyrhythm in which the matrix is as dense as it is complex (for more regarding *polyrhythm*, see [49.2, 3, 5, 9, 11, 12, 14, 18, 23, 34, 40, 42, 45, 46, 49, 50, 54, 60–62, 64–66, 68, 71, 73, 74, 76, 80, 83, 88, 92, 94, 96, 101, 106, 107, 112, 114, 135–139]).

Over the course of time, within each geocultural area, and often each ethnic community, musical heritage has followed a process of evolution through innovation and/or borrowing. This explains the extraordinary richness and the enormous variety of the types of music found on the African continent.

While the depositaries of these forms of music do not make them the subject of abstract speculation, all traditional musical idioms in Africa, inasmuch as they always obey a set of rules, are well and truly covered by a theory and thus constitute a *system* in the full meaning of the term.

Acknowledgments. I express my gratitude to Sylvie Le Bomin for agreeing to read this text and for her judicious remarks.

References

49.1 E.M. von Hornbostel: African Negro music, Africa **1**(1), 30–62 (1928)

49.2 A.M. Jones: The study of African musical rhythms, Bantu Stud. **9**, 295–319 (1937)

49.3 A.M. Jones: *African Music in Northern Rhodesia and Some Other Places* (The Rhodes-Livingstone Museum, Livingstone 1958)

49.4 A.M. Jones: On transcribing African music, Afr. Music **2**(1), 11–14 (1958)

49.5 A.M. Jones: *Studies in African Music*, Vol. 2 (Oxford Univ. Press, London 1959)

49.6 G. Kubik: *Mehrstimmigkeit und Tonsysteme in Zentral- und Ostafrika* (Böhlaus, Wien 1968)

49.7 G. Kubik: *Theory of African Music* (Univ. of Chicago Press, Chicago 2010), 2 Vol.

49.8 A. Lomax: Folk song style, Am. Anthropol. **61**(6), 927–954 (1959)

49.9 A.P. Merriam: Characteristics of African music, J. Int. Folk Music Counc. **11**, 13–18 (1959)

49.10 A.P. Merriam: The African idiom in music, J. Am. Folk. **75**(2), 120–130 (1962)

49.11 J.H. Kwabena Nketia: Les langages musicaux de l'Afrique subsaharienne. Étude comparative, Rev. Music. **288/289**, 7–42 (1972)

49.12 J.H. Kwabena Nketia: *The Music of Africa* (Victor Gollancz, London 1975)

49.13 G. Rouget: La musique en Afrique noire. In: *Encyclopédie de la musique Fasquelle*, Vol. III (Fasquelle, Paris 1961) pp. 939–940

49.14 S. Arom: *African Polyphony and Polyrhythm. Musical Structure and Methodology* (Cambridge Univ. Press, Cambridge 1991)

49.15 G.T. Basden: *Among the Ibos of Nigeria* (J.B. Lippincott, Philadelphia 1921)

49.16 J. Blacking: *The Role of Music amongst the Venda of Northern Transvaal* (International Library of African Music, Johannesburg 1957)

49.17 S.D. Cudjoe: The techniques of Ewe drumming and the social importance of music in Africa, Phylon **14**, 280–229 (1953)

49.18 V. Dehoux: *Les 'Chants à penser' des Gbaya de Centrafrique* (Selaf, Paris 1986)

49.19 P. Gbeho: Beat of the master drum, West Afr. Rev. **22**, 1263–1265 (1951)

49.20 S. Le Bomin: Étudier une musique instrumentale par ses processus d'apprentissage, Rev. Musicol. **90**(2), 179–192 (2005)

49.21 A.P. Merriam: *The Anthropology of Music* (Northwestern Univ. Press, Evanston 1964)

49.22 A.P. Merriam: The Bala musician. In: *The Traditional Artist in African Society* (Indiana Univ. Press, Bloomington 1973) pp. 250–281

49.23 S. Arom: The use of play-back techniques in the study of oral polyphonies, Ethnomusicology **20**(3), 483–519 (1976)

49.24 S. Arom: A synthesizer in the Central African bush. A method of interactive exploration of musical scales. In: *Für György Ligeti. Die Referate des Ligeti-Kongresses Hamburg 1988* (Laaber-Verlag, Laaber 1991) pp. 163–178

49.25 S. Arom, N. Fernando, F. Marandola: An innovative method for the study of African scales. In: *Proc. SMC'07, 4th Sound Music Comput. Conf., Lefkada*, ed. by C. Spyridis (National and Kapodistrian Univ. of Athens, Athens 2007) pp. 107–116

49.26 S. Arom, N. Fernando, S. Fürniss, S. Le Bomin, F. Marandola, J. Molino: La catégorisation des patrimoines musicaux dans les sociétés de tradition orale. In: *Catégories et catégorisation. Per-*

spectives interdisciplinaires, ed. by F. Alvarez-Pereyre (Peeters, Paris 2009) pp. 273–313
49.27 N. Fernando: Study of African scales: A new experimental approach for cognitive aspects. In: *Approaches to African Musics*, ed. by E. Cámara de Landa, S. Martinez García (Universidad de Valladolid, Valladolid 2007) pp. 57–72
49.28 N. Fernando: La construction paramétrique de l'identité musicale, Cah. Ethnomusicol. **20**, 39–66 (2007)
49.29 N. Fernando: *Patrimoines musicaux de l'Extrême-Nord du Cameroun* (Selaf, Paris 2011)
49.30 S. Fürniss: Musiques aka et baka: Une parenté de référence, J. Afr. **82**(1/2), 107–136 (2012)
49.31 S. Fürniss, E. Olivier: Pygmy and Bushman music: A new comparative study. In: *Central African Hunter-Gatherers in a Multidisciplinary Perspective: Challenging Elusiveness*, ed. by K. Biesbrouck (CNWS, Leiden 1999) pp. 117–132
49.32 C. Joyeux: Étude sur quelques manifestations musicales observées en Haute-Guinée Française, Rev. Ethnogr. Tradit. Pop. **5**(8), 170–212 (1924)
49.33 S. Le Bomin, F. Bikoma: *Musiques Myènè – de Port-Gentil à Lambaréné (Gabon)* (Sépia, Paris 2005)
49.34 S. Reich: Gahu. A dance of the Ewe tribe in Ghana. In: *Writings about Music* (New York Univ. Press, New York 1974) pp. 29–37
49.35 A. Schaeffner: *Origine des Instruments de Musique. Introduction ethnologique à l'histoire de la musique instrumentale* (Payot, Paris 1936)
49.36 A. Schaeffner: Situation des musiciens dans trois sociétés africaines. In: *Les Colloques de Wégimont III* (Société d'édition 'Les Belles Lettres', Paris 1960) pp. 33–49
49.37 B. Schmidt-Wrenger: Musique des Tshokwe du Zaïre. (brochure accompagnant le disque éponyme). In: *Coll. 'Enregistrements de musique africaine'*, Vol. 11 (Musée Royal d'Afrique Centrale, Tervuren 1975)
49.38 S. Arom: Le "syndrome" du pentatonisme africain, Music. Sci. **1**(2), 139–161 (1997)
49.39 S. Arom, F. Voisin: Theory and technology in African music. In: *The Garland Encyclopaedia of World Music*, Vol. 1, ed. by R. Stone (Garland, New York 1998) pp. 254–270
49.40 C. Ballantine: The polyrhythmic foundation of Tswana pipe melody, Afr. Music **3**(4), 52–67 (1965)
49.41 P. Berliner: *The Soul of Mbira* (Univ. California Press, Los Angeles 1978)
49.42 R. Brandel: Africa. In: *Harvard Dictionary of Music*, 2nd edn., (Belknap of Harvard Univ. Press, Cambridge 1970) pp. 17–24
49.43 A. Gide: *Le Retour du Tchad. Carnets de route* (Gallimard, Paris 1928)
49.44 Y. Grimaud: *Note sur la musique des Bochiman comparée à celle des Pygmées Babinga* (Peabody Museum, Harvard Museum et Musée de l'Homme, Cambridge, Paris 1957)
49.45 Y. Grimaud: Note sur la musique vocale des Bochiman !Kung et des Pygmées Babinga. In: *Les Colloques de Wégimont III* (Société d'édition 'Les Belles Lettres', Paris 1960) pp. 105–126
49.46 P.R. Kirby: *The Musical Instruments of the Native Races of South Africa* (Oxford Univ. Press, London 1934)
49.47 P.R. Kirby: Buschmann- und Hottentottenmusik. In: *Die Musik in Geschichte und Gegenwart*, Vol. 2 (Bärenreiter, Kassel 1952) pp. 502–510
49.48 M. Kolinski: Review of A.M. Jones: *Studies in African Music*, Music. Q. **16**(1), 105–110 (1960)
49.49 G. Kubik: Musikgestaltung in Afrika, Neues Afr. **5**, 195–200 (1961)
49.50 G. Kubik: Harp music of the Azande and related peoples in the Central African Republic, Afr. Music **3**(3), 37–76 (1964)
49.51 G. Kubik: Transmission et transcription des éléments de musique instrumentale africaine, Bull. Int. Comm. Urgent Anthropol. Ethnol. Res. **11**, 47–61 (1969)
49.52 F. Marandola: L'apport des nouvelles technologies à l'étude des échelles musicales d'Afrique centrale, J. Afr. **69**(22), 109–119 (1999)
49.53 F. Marandola: The study of musical scales in Central Africa: The use of interactive experimental methods. In: *Computer Music Modeling and Retrieval*, ed. by U.K. Wiil (Springer, Berlin 2004) pp. 34–41
49.54 G. Rouget: Un chromatisme africain, L'Homme **1**(3), 1–15 (1961)
49.55 G. Rouget: Sur les xylophones équiheptatoniques des Malinké, Rev. Music. **55**(1), 47–77 (1969), avec la collaboration de J. Schwarz
49.56 O. Tourny: Using the same to make different. Analysis of a traditional musical repertoire based on centonisation, Music. Sci. **13**, 181–200 (2009)
49.57 H. Tracey: Towards an assessment of African scales, Afr. Music **II**(1), 10–20 (1958)
49.58 H. Tracey: Measuring African scales, Afr. Music **4**(3), 73–77 (1969)
49.59 V. Kofi Agawu: *African Rhythm: A Northern Ewe Perspective* (Cambridge Univ. Press, New York 1995)
49.60 S. Akpabot: African instrumental music, Afr. Arts **5**(1), 63–64 (1971)
49.61 M. Beclard-D'Harcourt: Notes relatives à la transcription des phonogrammes. in C. Joyeux: "Étude sur quelques manifestations musicales observées en Haute-Guinée Française", Rev. Ethnogr. Tradit. Pop. **5**(18), 173–202 (1924)
49.62 J. Blacking: Patterns of Nsenga Kalimba music, Afr. Music **2**(4), 26–43 (1961)
49.63 J. Blacking: Deep and surface structure in Venda music, Yearb. Int. Folk Music Counc. **3**, 91–108 (1971)
49.64 A.M. Dauer: Afrikanische Musik und völkerkundlicher Tonfilm. Ein Beitrag zur Methodik der Transkription, Res. Film **5**(5), 439–456 (1966)
49.65 V. Dehoux, N. Fernando, S. Le Bomin, F. Marandola: Itinéraire rythmique du Cameroun à la Centrafrique, Cah. Musiques Tradit. **10**, 81–105 (1997)
49.66 N.M. England: Bushman counterpoint, J. Int. Folk Music Counc. **19**, 58–66 (1967)

49.67 Z. Estreicher: Une technique de transcription de la musique exotique, Bibliothèques et Musées de la Ville de Neuchâtel (Rapport), Neuchâtel (1957) pp. 67–92

49.68 Z. Estreicher: Le rythme des Peuls Bororo. In: *Les Colloques de Wégimont IV* (Société d'édition 'Les Belles Lettres', Paris 1964) pp. 185–228

49.69 S.M.X. de Golberry: *Fragmens d'un voyage en Afrique*, Vol. 2 (Treuttel et Würtz, Paris 1802)

49.70 F. Harrison: *Time, Place and Music. An Ethnomusicological Observation c. 1500 to c. 1800* (Frits Knuf., Amsterdam 1973)

49.71 A.M. Jones: African rhythm, Africa **24**, 26–47 (1954)

49.72 A.M. Jones, L. Kombe: *The Icila Dance Old Style. A Study in African Music and Dance of the Lala Tribe of Northern Rhodesia* (Longmans, Green and Co., Cape Town 1952)

49.73 P.R. Kirby: The musical practices of the /?Auni and ≠Khomani Bushmen, Bantu Stud. **10**, 373–431 (1936)

49.74 J. Koetting: Analysis and notation of West African drum ensemble music. In: *Selected Reports in Ethnomusicology*, Vol. I/3 (Univ. of California, Los Angeles 1970) pp. 116–146

49.75 P. Kolbe: *Description du Cap de Bonne-Espérance ...* (Chez Jean Catuffe, Amsterdam 1741)

49.76 M. Kolinski: A cross-cultural approach to metro-rhythmic patterns, Ethnomusicology **17**(3), 494–506 (1973)

49.77 G. Kubik: Xylophone playing in Southern Uganda, J. R. Anthropol. Inst. **94**(2), 138–159 (1964)

49.78 S. Le Bomin: *Musiques Bateke – Mpa Atege (Gabon)* (Sépia, Paris 2004)

49.79 D. Locke: Principles of offbeat timing and cross-rhythm in Southern Ewe dance drumming, Ethnomusicology **26**(2), 217–246 (1982)

49.80 A. Lomax: *Folk Song Style and Culture* (American Association for the Advancement of Science, Washington D.C. 1968)

49.81 A.P. Merriam: African music rhythm and concepts of time-reckoning. In: *Ethnomusicological Theory and Method*, ed. by K. Kaufman Shelemay (Garland Publishing, New York 1990) pp. 293–311

49.82 J.H. Kwabena Nketia: *Folk Songs of Ghana* (University of Ghana, Legon 1963)

49.83 H. Pantaleoni: The three principles of timing in Anlo dancing drumming, Afr. Music **5**(2), 50–57 (1972)

49.84 M. Park: *Voyage dans l'intérieur de l'Afrique* (Dentu et Casteret, Paris 1800), 2 Vol.

49.85 D. Rycroft: Nguni vocal polyphony, J. Int. Folk Music Counc. **19**, 88–103 (1967)

49.86 R.M. Stone: In search of time in African music, Music Theory Spectr. **7**, 139–148 (1985)

49.87 R.A. Waterman: African influence on the music of the Americas. In: *Acculturation in the Americas* (Univ. Chicago Press, Chicago 1952) pp. 207–218

49.88 V. Kofi Agawu: "Gi Dunu", "Nyekpadudo" and the study of West African rhythm, Ethnomusicology **30**(1), 64–81 (1986)

49.89 V. Kofi Agawu: Variation procedures in North Ewe song, Ethnomusicology **34**(2), 221–243 (1990)

49.90 S. Akpabot: *Form, Function, and Style in African Music* (Macmillan Nigeria, Lagos 1998)

49.91 S. Arom: De la chasse au piège considérée comme une liturgie, World Music **16**(4), 3–19 (1974)

49.92 S. Arom: Time structure in the music of Central Africa: Periodicity, meter, rhythm and polyrhythmics, Leonardo J. Int. Soc. Arts Sci. Technol. **22**, 91–99 (1984)

49.93 L. Ekwueme: Structural levels of rhythm and form in African music: with particular reference to the West Coast, Afr. Music **5**(4), 27–35 (1975)

49.94 R. Kauffman: African rhythm. A reassessment, Ethnomusicology **24**(3), 393–415 (1980)

49.95 A. King: Employments of the 'Standard Pattern' in Yoruba music, Afr. Music **2**(3), 51–54 (1960)

49.96 J.H. Kwabena Nketia: The Hocket-technique in African music, J. Int. Folk Music Counc. **14**, 44–52 (1962)

49.97 J.H. Kwabena Nketia: Multi-part organization in the music of the Gogo of Tanzania, J. Int. Folk Music Counc. **19**, 79–88 (1967)

49.98 G. Rouget: A propos de la forme dans les musiques de tradition orale. In: *Les Colloques de Wégimont* (Elsevier, Bruxelles 1956) pp. 132–144

49.99 O. Tourny: *Le Chant Liturgique Juif Éthiopien. Analyse Musicale d'une Tradition Orale* (Peeters-Selaf., Louvain-Paris 2009)

49.100 M.S. Eno Belinga: *Littérature et musique populaire en Afrique Noire* (Ed. Cujas, Paris 1965)

49.101 R. Brandel: *The Music of Central Africa. An Ethnomusicological Study* (Martin Nijhoff, The Hague 1961)

49.102 R. Brandel: Polyphony in African music. In: *The Commonwealth of Music in Honour of Curt Sachs* (The Free Press, New York 1965) pp. 26–44

49.103 V. Erlmann: Model, variation, and performance: Ful'be praise song in Northern Cameroon, Yearb. Tradit. Music **17**, 88–112 (1986)

49.104 S. Fürniss: Aka polyphony: Music, theory, back and forth. In: *Analytical Studies in World Music*, ed. by M. Tenzer (Oxford Univ. Press, Oxford, New York 2006) pp. 163–204

49.105 L. Godsey: The use of variation in Birifor funeral music. In: *Studies in African Music*, Selected Reports in Ethnomusicology, Vol. V (University of California, Los Angeles 1984) pp. 67–80

49.106 R. Günther: *Musik in Rwanda. Ein Beitrag zur Musikethnologie Zentral-Afrikas*, Annales, Sciences Humaines, Vol. 50 (Tervuren Koninklijk Museum Voor Midden-Afrika, Tervuren 1964)

49.107 M. Hood: *The Ethnomusicologist* (Mc Graw-Hill, New York 1971)

49.108 J. Koetting: Hocket concept and structure in Kasena flute ensemble music. In: *Selected Reports in Ethnomusicology*, Studies in African Music, Vol. 5 (Univ. of California, Los Angeles 1984) pp. 161–172

49.109 S. Arom, N. Fernando, F. Marandola: A cognitive approach to Bedzan Pygmies vocal polyphony and Ouldeme instrumental polyphony: Methodology and results. In: *Proc. 5th Trienn. ESCOM Conf.*, ed. by R. Kopiez (Inst. for Research in Music Educa-

tion, Hanover Univ. of Music and Drama, Hanover 2003) pp. 457–466, description: 1 CD-*R*

49.110 G. Bosman: *Voyage de Guinée, Contenant une Description Nouvelle et Très Exacte de Cette Côte ...* (Antoine Schouten, Utrecht 1705)

49.111 T.E. Bowdich: *Mission from Cape Coast to Ashantee* (John Murray, London 1819)

49.112 R. Brandel: Music of the giants and the pygmies of the Belgian Congo, J. Am. Musicol. Soc. **5**(1), 16–28 (1952)

49.113 P. Collaer: Notes sur la musique d'Afrique Centrale, Probl. Afr. Cent. **26**, 267–271 (1954)

49.114 A.M. Dauer: Research films in ethnomusicology. Aims and achievements, Yearb. Int. Folk Music Counc. **1**, 226–231 (1969)

49.115 S. Fürniss: La conception de la musique vocale chez les Aka: terminologie et combinatoires de paramètres, J. Afr. **69**(2), 147–162 (1999)

49.116 A. Gide: *Voyage au Congo. Carnets de route* (Gallimard, Paris 1927)

49.117 Y. Grimaud: Étude analytique de la danse 'Choma' des Bochiman !Kung (Polyrythmie). In: *Les Colloques de Wégimont IV* (Société d'édition 'Les Belles Lettres', Paris 1964) pp. 171–183

49.118 G.-M. Haardt, L. Audouin-Dubreuil: *La croisière noire. Expédition Citroën Centre-Afrique* (Plon, Paris 1927)

49.119 F. Hornburg: Phonographierte Afrikanische Mehrstimmigkeit, Musikforschung **3**(2), 120–142 und 161–176 (1950)

49.120 C. Joyeux: Notes sur quelques manifestations musicales observées en Haute-Guinée, Rev. Music. **10**(2), 49–58 (1910)

49.121 R. Kauffman: Multipart relationships in Shona vocal music. In: *Studies in African Music*, Selected Reports in Ethnomusicology, Vol. 5 (University of California, Los Angeles 1984) pp. 145–160

49.122 P.R. Kirby: A study of Negro harmony, Music. Q. **16**(3), 404–430 (1930)

49.123 P.R. Kirby: The reed-flutes ensembles of South Africa. A study in South African native music, J. R. Anthropol. Inst. G. B. Irel. **63**, 313–388 (1933)

49.124 R. Knight: The style of Mandinka music. A study in extracting theory from practice. In: *Studies in African Music*, Selected Reports in Ethnomusicology, Vol. 5 (Univ. of California, Los Angeles 1984) pp. 3–66

49.125 G. Kubik: Transcription of Mangwilo xylophone music from film strips, Afr. Music **3**(4), 35–51 (1965)

49.126 J.-B. Labat: *Relation Historique de l'Ethiopie Occidentale ...* (C.J.B. Delespine le fils, Paris 1732)

49.127 A.A. Mensah: The polyphony of Gyil-gu, Kudzo and Awutu Sakumo, J. Int. Folk Music Counc. **19**, 75–79 (1967)

49.128 A. Schaeffner: La découverte de la musique noir, Monde Noir **8/9**, 205–218 (1950)

49.129 A. Schaeffner: *Les Kissi, une société noire et ses instruments de musique* (Hermann, Paris 1951)

49.130 G. Schweinfurth: *Au coeur de l'Afrique 1868–1871* (Hachette, Paris 1875), 2 Vol.

49.131 K.P. Wachsmann: *Tribal Crafts of Uganda. Part Two: The Sound Instruments* (London Univ. Press, London 1953)

49.132 S. Arom: The music of the Banda-Linda horn ensembles. Form and structure, Sel. Rep. Ethnomusicol. **5**, 173–193 (1984)

49.133 G. Herzog: Canon in West African xylophone melodies, J. Am. Musicol. Soc. **2**(3), 196–197 (1949)

49.134 C.E. Schmidt: Interlocking techniques in Kpelle music, Selected Reports in Ethnomusicology, Stud. Afr. Music **5**, 195–216 (1984)

49.135 A.M. Jones: African drumming. A study of the combination of rhythms in african music, Bantu Stud. **8**(1), 1–16 (1934)

49.136 G. Kubik: Oral notation of some West and Central African time-line patterns, Rev. Ethnol. **3**(2), 169–176 (1962)

49.137 C. Sachs: *The Rise of Music in the Ancient World East and West* (W.W. Norton, New York 1943)

49.138 C. Sachs: *Rhythm and Tempo. A Study in Music History* (W.W. Norton, New York 1953)

49.139 C. Sachs: *The Wellsprings of Music* (Martinus Nijhoff, Amsterdam 1962)

50. Music Among Ethnic Minorities in Southeast Asia

Håkan Lundström

In the countries of mainland Southeast Asia there are several ethnic minority groups, particularly in the mountainous inland and in forest areas. There is much variation between the customs of these peoples but it is also possible to see similarities on a metalevel. In this chapter strong traits in the village-based music culture of the ethnic minorities are presented, in some cases on purely historical grounds. These traits are in many cases paralleled in the tradition of the majority peoples. Against this background follows a discussion on musical change and matters of sustainability.

The geographic and cultural location of mainland Southeast Asia, bordering the influential old cultures of India and China, is evident in its traditional music as well as in many other aspects of life. Generally the modal systems and the use of microintervals, as in the Vietnamese classical tradition, are seen as influences from India. Certain musical instruments and music dramas appear to have a Chinese background like the *tuong* in Vietnam. In the most southern part there are also influences from the Middle East.

On the other hand, Southeast Asia, including insular Southeast Asia and the southern parts of China, has since early times been a cultural area with a number of distinct practices that have spread to the courts of imperial China, Korea and Japan. Musical instruments made from bamboo are abundant, including tube zithers, from simple bamboo tubes with one string to modern refined zithers. Gongs and xylophones, in combination with the musical organization called colotomic, are present in Thai and Laotian ensemble music and all over East and Southeast Asia. Another factor considered original for the area is the free reed, a construction present in mouth organs like the Laotian *khene*. To this can be added the manner in which poetry and music have met in partly improvised vocal music, as in Laotian *mo lam* and Vietnamese *quan ho*.

These are some of the musical practices in Southeast Asia that have resulted from people moving from one place to another and from political power. They have been integrated through history and developed into what are now recognized as characteristics of Southeast Asian music. This is not to forget the influence of European music from colonial times on – chiefly Christian music, school music and classical music – or various forms of popular music in the global age and the emergence of hybrid musical styles.

This chapter will deal with the music of the many ethnic minorities in Southeast Asia, some of which are lowland farmers of irrigated rice, while others grow dry rice on mountainsides and mountaintops. Al-

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_50

though the terms majority and minority peoples are not unproblematic, they will be used here for practical reasons. The focus will be mainly the highlanders in the mountainous inland areas, including Yunnan in southernmost China, which is culturally related. These are the peoples that sometimes have been referred to as *hill tribes* or *mountain people* belonging to four dominant language families, namely the Sino-Tibetan, Austroasiatic, Tai-Kadai and Hmong-Mien languages. Most are traditionally villagers for whom dry rice farming, hunting and fishing have been the main source of food.

While the published sources are limited, *The Garland Encyclopedia of World Music* contains an overview [50.1] as does *Neues Handbuch der Musikwissenschaft* [50.2]. There are only a few monographs on music of minorities [50.3–6], musical instruments [50.7, 8] and verbal arts including songs [50.9–14]. The majority of the available material is in the form of articles in various journals and ranges from general descriptions of minority music cultures to studies of songs and specific musical instruments. There are a fair number of audio recordings, often with valuable comments. There are very few studies that attempt to present the music as integrated in a cultural context so the best sources in this regard are actually anthropological studies, which often include information on music. Beside some old iconographic sources and ancient Chinese writings, information about more recent history can be found in publications by explorers – mainly French – from the latter part of the 19th century and the beginning of the 20th century.

The music cultures of the ethnic minorities demonstrate a wide variety of musical expressions and musical instruments. Traditionally most of these peoples live in villages in the mountainous inland areas and are mainly rice farmers and hunters. This is reflected in their music, which is often related to major seasons during the farming year, particularly the harvesting season. The music also reflects the religious life. Ceremonies directed to the ancestor spirits or to the rice soul are especially important, like the buffalo ceremony when a water buffalo is slaughtered as a sacrifice to certain spirits. In such ceremonies gongs and drums usually play a central role.

The social uses of the music are evident in ensembles where each member can produce only one or very few tones. This is the case in the gong ensembles, but also in certain flute ensembles. Much of the singing is done as alternating singing, often between a man and a woman or a group of men and a group of women, often when courting or as pastime during festivities. Typically the singing uses a basic melody frame that is lengthened or shortened depending on the words. It is a kind of poetic recitation that often leaves room for variation and improvisation.

The examples considered in this chapter stem from a few of these cultures and while they are representative in a general sense it must be mentioned that even though the musical traditions may seem similar on the surface the amount of variety is astounding, even between neighboring villages of the same minority. Many of these minority cultures and their languages are now counted as endangered. The musical practices that are mentioned in this chapter will serve as a background for a discussion of change and sustainability.

50.1 Singing Manners

One of the earliest transcriptions of a song – in French translation – is a sung dialog between a boy and a girl recorded in north Laos in an area inhabited mainly by Kammu (Khmu or the similar) people. As is still done today, the boy starts by praising the girl's beauty while she politely replies belittling herself as being ugly [50.15]. A couple of decades later another song was published in the author's own transcription of the Kammu language. This was the time of year when the forest was to be burnt in order to make fields and a ceremony was held that included bronze drums, gongs and cymbals. One group of boys and one group of girls then sang in alternation and the author says [50.16, p. 197]:

> *I have in vain tried to learn the meaning of these two songs. Oddly enough everybody knows them by heart and no one can explain them. Could it be a foreign language or an ancient form of today's Kammu which now has become incomprehensible?*

Though the words may actually be incomprehensible to some extent even to the singers and indeed very difficult to translate, this particular song still exists and can be found in variants [50.14, p. 181] or reconstructions [50.13, pp. 5–6].

These are examples of the few but important cases where historical records can provide a background for the study of contemporary practices. In the case of alternating singing there is definitely continuity: the practice is still very widespread in this area and found among practically all the minorities. Alternating singing occurs in a number of situations when there are two or more

people present, who take turns singing either as a response to a sung statement, or in continuation through songs linked to the previous song by some sort of association. It could be a sung dialog between two persons of the same sex or between a male and a female, particularly in a courting situation. As quoted above, alternating singing may also be performed by two groups of singers responding to each other. At parties where more people are present alternating singing may take the form of chains of linked songs. Courting singing is not always a private matter between a boy and a girl, but can also be a more ceremonial situation where the parent generation is present as is the case among the Lisu in northern Thailand [50.17, p. 51ff.]. A well-known ceremonial courting tradition with alternating singing belongs to the Hmong New Year, spring and harvest festivals in Yunnan, China, and with some variations among Hmong in Southeast Asia [50.12].

50.1.1 Monomelodic Styles

The words of the songs may be preexisting orally transmitted poems – normally with an amount of variation – or they can be totally or partially made up on the spot. The singing is normally based on one basic melody, a melodic template that can be altered with regard to pitches and length as different words are sung. Such a template usually has a rather distinct starting formula as well as ending formulae at the end of the phrases, while the middle part is where variation will occur. Often it starts high with a word meaning *oh*, *hey* or similar and then ends lower, but there are variations to this (Fig. 50.1).

People may sing like this when they wander alone in the forest or in the fields, but the most special singing situation is when there is a party in a village, in connection with certain ceremonies or when somebody visits the village. People sit down in a circle and sip rice wine through straws from a common large earthenware pot. At these times people sing to each other and a song is expected to get an answer. The poems deal with many things. Loneliness during travels and yearning and love songs are common categories, but a very high proportion deal with human relations. This kind of singing did not appeal musically to early explorers or musicologists, who found it monotonous and primitive, as in this case referring to *Kha* in Laos [50.18, p. 46]:

> *The singing of the jar is limited to the 'tribal' phrase, often short and monotonous. In no way do they re-echo the magnificent and important ensembles, choirs and drums . . .*

This reflects the fact that singing cannot be understood by analyzing the music alone, but that the language aspect and the context must also be taken into account. In so doing this kind of singing turns out to be a complex creative activity.

Taking the Kammu in northern Laos as an example, in principle each dialect area has its own melodic template, each village its own variant of it and each individual a personal variant of the village template. Basically the melodies of neighboring villages are rather similar, but the farther away you get the more different the melodies. Looked at from an outside perspective these melodies may at first seem very much alike, but to the insider all the nuances are perceptible. Parts of the template include vocalices – particularly in the beginning and end of phrases – and other parts contain the words of the poem or song. In some dialect areas – like Kammu tonal dialects in the western part of north Laos with two tones – the vocalice parts are short, while in the nontonal Kammu dialects to the east large portions of the template consists of vocalices. There Kammu women may play the vocalice part on the flute while also sounding their voice. In the parts where the words

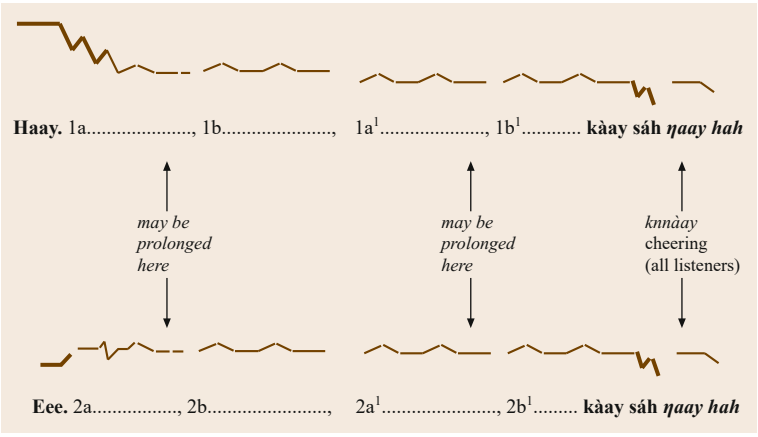


Fig. 50.1 Approximate graph of the west (Yùan) Kammu musical template including introductory and final formulae (in *bold*)

Table 50.1 Situational framework for singing in Kammu culture

Social	Sex/age: Status: Relation:	Man/woman/young/old Higher/equal/lower Friend/relative/formal
Spatial	Local level: General level:	Village/fields/forest Individual/village/area/country
Chrono-logical	Taboos:	Calendrical/seasonal/life cycle

are sung they use only the voice and then again return to the voice-and-flute playing of the next melisma [50.19].

In a singing dialog one is expected to praise and beautify the person(s) the song is directed to and to deprecate and belittle oneself and one's own belongings. Several songs exist in pairs of praise and belittling functions, particularly for beginning or ending of a singing dialog. The songs do not necessarily build up to a continuous and straight conversation, but rather resemble a chain of telegram-like units, often also including symbolic messages. This singing has a strong social meaning and taking part in the singing means expressing one's own identity and reinforces one's cultural belonging.

This manner of singing is called *tɔ́əm* and the poem that is performed is called *trnə̀əm*. Each song can be understood as recreated in the performance situation by the combination of traditional poems, stock words or phrases and occasionally instantly improvised words and a melodic template. This is a creative process involving *musicopoetic* (musical, poetic, linguistic) competence as well as *situational* (social, spatial, chronological) competence since the singer must choose words, polite phrases and even melodic template with regard to the context: Where does the singing take place? What is the occasion? Who is present? The aim is a functional and aesthetically satisfactory performance [50.6] (Table 50.1).

A common denominator in alternating singing is the use of one basic melody or tune: a melodic template that is varied according to the words. This can be referred to as a *monomelodic* practice and it is very common among most peoples. The Lisu in northern Thailand have different templates for singing in courting, festival and soul-calling contexts [50.17, pp. 56–59]. The songs of the Mnong Gar of Vietnam are reported to have a high degree of variation from one singer to another and evidently also by the same singer. One song can be used in several situations. To sing is called *tong* and to sing in exchange is called *tâm tong* (*tâm* = exchange) [50.20–22]. This practice is also found in other places, notably the *shange* (or *mountain songs*, a term often used to designate monomelodic exchange songs among various peoples) of southern Jiangsu in the Shanghai area in China that have *tune regions* and *textual regions*. Texts have a *supraregional* distribution whereas tunes are limited to smaller geographic units that to some degree coincide with dialect areas. Some singers know up to ten such tunes, usually with a strong preference for one of them. Singers sometimes attach the name of their village or county to it [50.23, pp. xii, 1298–1131, 267]. This is very similar to the Kammu of northern Laos and certainly several other minorities. It also parallels practice in traditions of the Laotian major population where local styles of *lam* or *khap* singing are are often named by city, village or area [50.24, pp. 96–100].

50.1.2 A Monomelodic Organization of Vocal Genres

Many – or most – ethnic minorities in Southeast Asia have vocal genres similar to the Kammu *tɔ́əm*. Among the examples in the literature are studies of songs among Sino-Tibetan peoples such as the love dialog songs among the Hmong Blanc in northern Thailand [50.12] and Lisu singing [50.4], [50.17, p. 56ff.]. The songs of the Lawa in northern Thailand are sung to one melody, which is adjusted to the rhythms and intonation of the Lawa language [50.25]. In insular Southeast Asia similar types of singing exist, for example in Sabah [50.26, pp. 50–51]. This mono-melodic singing is often described as not really singing and characterized by terms like *recitation*, *chanting* or *heightened speech*, which indicates that these English language terms are culture specific just like the terms *song* or *singing* are. Therefore they are actually ethnocentric and not applicable in all music cultures.

This becomes clear when one looks closer at monomelodic singing. In the same Kammu village tradition there are several genres, each with its individual name, its individual melodic template and its individual use context. Though some poems, *trnə̀əm*, are specific to certain situations many of them can be performed – or recreated – in any of these genres by a person who knows the techniques involved. The genres thus function as different representations of the *trnə̀əm* in different contexts. These are represented in Table 50.2.

While the terms for each genre are quite specific, Kammu people would normally differ between *singing* and *talking* in a general sense. For instance, *tɔ́əm* would be considered *singing* and *hrlɨ́ɨ* would be described as more like talking. One must accept that general terms like these are not very specific, neither in Kammu nor in English. In a scientific context they are therefore not very useful and could actually mislead a researcher trying to understand a specific music culture. In order to

Table 50.2 Vocal genres in Yùan Kammu tradition

Vocal genre	Melody	Situation	Performed by
Táɔm	Template, longdrawn, falling	Partying, in village	Male/female
Hrlíi flattering words	Template built almost totally on word tones (Yùan Kammu languages have two word tones)	In village when no party, in fields	Children, youths, male/female
Hrwə̀	Template built on a short melodic motif	Fields, forests	Adolescents, male/female
Húuwə̀	Like *hrwə̀* with recurring refrain	Fields, forests	Adolescents, male/female
Yàam weeping	Template built on a recurring melodic motif	Fields, forests	Mainly female, also children

avoid this the term *vocal expression* is useful for describing a performance that is neither song nor speech – or a little of both – and *vocal genre* may be used for the subcategories of vocal expressions [50.6, pp. 15–16].

Aside from a number of other categories of vocal expressions belonging to rituals, ceremonies or other situations, the vocal genres of the Kammu may be seen as a system of monomelodic organization. Evidence in written sources points toward the existence of similar organizations in other minority cultures. Among the White Hmong in Thailand there are different singing genres and each has its own fixed melody [50.12, p. 14]. Similar systems may lie behind two reports concerning the Mon-Khmer-speaking groups Jeh and Rengao in Vietnam [50.27], [50.28, pp. 153–163]. It is likely that the lack of literary references to similar category systems stems from the fact that this special feature has not been particularly studied. Therefore more research is needed in this area in order to reach an understanding of vocal tradition.

50.1.3 Tone Languages and Music

Many of the languages in Southeast Asia are tone languages, which means that anything from two up to seven tones are used to decide the meaning of spoken words. In most of the traditional singing the language tones are also reflected in the pitches that are sung, though this can be done in many different ways. Normally the beginnings and ends of phrases are determined by the melodic template, (*music centration*), while the rest of the template may be dominated by the word tones (*language centration*).

The execution of the word tones may make two performances built on the same template sound quite different. It is therefore possible to perform a song by whistling or on a leaf, a flute or a fiddle so that a listener who knows the language, the culture and the context can understand it. This means that instrumental performances may sometimes actually be specific songs performed on an instrument. This capacity is commonly used in the courting situation where mouth harps or other instruments may serve as voice disguisers or voice surrogates. In this way a secret message can be communicated without actually pronouncing it with one's natural voice.

Hmong girls in Sapa, close to the Chinese border in northern Vietnam, usually play the mouth harp in order to call a boyfriend. The player thinks the words of a song and plays them on the instrument so that a listener can recognize the song and understand its contents – and often recognize who is actually sending the message. Mouth harps that may be made from bamboo or from laminated bronze in the shape of a bamboo leaf produces a buzzing low tone and the player can alter the harmonics by changing the form of the mouth cavity. In this way a whistling melody built on harmonics can be created. This is a widespread practice and in Kammu and Lamet tradition the mouth harp is primarily a boy's instrument used by boys for serenading outside a girl's door. The girl can recognize who it is from the playing style [50.13, p. 381ff.], [50.29, p. 100], [50.30].

Some wind instruments are equipped with a free reed of bamboo or metal fastened to the side of the instrument. The player will cover the section with the reed with his lips and it will produce a characteristic sweet and mellow sound when blown. One such instrument is a side-blown horn from a gaur or a water buffalo with a free reed of copper fitted onto its side. When it is played the opening is covered by one hand and a small hole at the narrow end of the horn is alternatively open or covered by the thumb of the other hand. It was used for signaling when a larger animal – like a deer or a gaur – had been killed in hunting. The side-blown horn with a free reed is widely spread among ethnic minorities in Southeast Asia and it often has ceremonial or ritual functions. Among the Mnong people in Vietnam it is played to call, inform and entertain during community activities such as the buffalo ceremony, when a buffalo is sacrificed to the ancestor spirits.

Similar constructions can be used on side-blown flutes as well, but the reed may also be placed near the top end of the flute and kept inside the mouth cavity when the flute is played so that it is held aslant. Though this construction can be found elsewhere, for instance among the Tai people, it is often referred to as the *Hmong flute* since its mellow sound is considered char-

Fig. 50.2 Detail of the top end of a free-reed Hmong flute. A narrow tongue has been cut out from the rectangular brass piece. When the flute is played it is held so that the lips cover the whole area and the reed that produces the tone can swing freely. (Photo: H. Lundström)

acteristic of the Hmong people (Fig. 50.2). This kind of flute is often used for love songs performed so that the melody reflects the different tones of the language and a listener can understand the words from listening to the flute (cf. [50.31]).

In the mouth organ several free-reed pipes of different lengths are placed together in a wood, bamboo or calabash container that the player blows into. It can sound more than one pitch at a time and usually one or more pipes also serve as drones. The mouth organ exists in a number of different shapes and is unique to Southeast Asia and East Asia. In the *khene* of Laos, the pipes are arranged in two flat rows parallel to each other and protrude below the wind chamber.

The Hmong variety of the *khene* (*qeej*, pronounced *geng*) has six rather long free-reed pipes in a wooden wind chamber. It is played by males in combination with movement and has many functions, particularly for love exchange and for communicating with spirits at ceremonies. In funeral ceremonies the *khene* is used for a set of melodies for the different steps of the ritual. In this ritual context the *khene* speaks by playing the words of the ritual songs using the seven word tones of the language. Like with other instruments it is not a simple illustration of the word tones. The words are rather transposed into the idiomatic resources of the instrument and may not be easily understood by a listener [50.32]. At certain times after a funeral the song of guidance to the dead person's soul is performed on the *khene*. This is an instruction to the soul explaining how to get to the land of the dead [50.33] (see also for Kammu ritual *Svantesson* et al. [50.34]). The Samre and other peoples in the Cardamom mountains in northwest Cambodia also use mouth organs at funerals as well as other ceremonies and often as accompaniment to singing [50.35].

50.2 The Sounds of Bamboo and Metal

Bamboo that grows in the area from insular Southeast Asia to Japan in the East is a cheap, strong and versatile material of great importance in many aspects of life. It is used for house building, for constructing tools and kitchenware as well as for musical instruments. Another material that has played important roles in this area is manmade, namely metal and various alloys of metal. The knowledge necessary to produce metal objects is known through archaeological findings and is believed to have started in the so-called Dongson culture that existed in the most northern part of today's Vietnam more than 2000 years ago, which is famous for its bronze drums.

50.2.1 Bamboo and Musical Instruments

Since it is hollow, a simple piece of bamboo emits a sound when bounced on the ground, when struck or when clapping one's hand close to the opening and can be fine-tuned by adjusting its length to the air it contains to get the best resonance. In some places stamping tubes are used as single tubes bounced against the ground, for instance at Kammu house-building feasts, but they are also made in series of different sizes tuned so that they produce different pitches when bounced against the ground, a wooden log or a stone. Each individual may handle one or two tubes, producing as many pitches. Among the Senoi and Temiar in Malaysia such ensembles of stamping tubes are used at trance-dancing ceremonies where a medium, often female, serves as a lead singer and brings words that are repeated by a chorus [50.5].

The sound of stamping tubes is rather warm and soft as it is created by the shock when the tube hits the ground or log, which sets the air inside the tube in motion. A similar effect occurs when clapping one's cupped hands in front of the opening of a tube, much in the same manner as producing sound by clapping the

hands together in front of the mouth cavity. A tube of a particular length will produce a distinct pitch. Several tubes of different sizes can then produce different pitches like the *k'long put* of the Sedang, Jarai and Bahnar in Vietnam. Among the Brau one man holds two long bamboo tubes cross-wise and one female at each opening creates the sound. Two of them clap their hands together alternately in front of either opening of the same tube. One of the two beats a rhythmic pattern while the player at the opposite end of the same tube alternates between dampening that sound by hitting that opening with the palm of her hand or permitting it to sound. This is *ding buk*, played in connection with clearing trees for making fields (Fig. 50.3).

Bamboo tubes can be further fine-tuned by cutting off a piece aslant so that it gets the shape of a large goose quill – or a bamboo spear for that matter – with one end serving like a tongue and the other as a resonating chamber. Such tubes occur as single tubes struck with a beater of some kind or as concussion tubes made from a pair: one large tube and one shorter, like in the *kltɔ̀ɔŋ* of the Kammu that can be found all over the area. Its sound is generally more percussive than that of the stamping tubes. The *ko kleh* of the Sedang in Vietnam consists of three such tubes of different sizes tied together in a frame and struck with a short wooden stick, producing both rhythm and pitches. When such tubes of different sizes – and pitches – are placed lying down on two parallel ribs or on the player's outstretched legs they become a xylophone. A xylophone may also be made by suspending them hanging on parallel ropes like the *trung* of the Bahnar in Vietnam.

Fig. 50.3 The *ding buk* of the Brau people. Three women sound the tubes by clapping their hands in front of the opening. The fourth (*far left*) dampens the sound of that tube. This photo has a foklorized touch to it. Minorities are also often depicted in pastoral situations like this one in scientific documentations. (Photo: Vietnamese Institute for Musicology)

The Kammu *tàaw-tàaw* is an idiophone made from a piece of thin bamboo, which is split and cut so that two tongues are produced, which vibrate when struck and thus produce sound. The instrument is held in one hand and the split end is struck against the wrist of the other hand. It produces a buzzing sound, which serves as a drone when partials are produced, much like a mouth harp. A fingerhole on the handle side, handled by the right hand thumb, and sometimes a second hole on the other side of the tube, operated by the first or second finger, increase the sound possibilities of the drone and consequently also the partials. Further variation is obtained if the open handle-end of the tube is dampened by pressing it to the chest or thigh. This instrument is typically used by Kammu women while walking along forest paths as the sound is thought to scare off wild animals [50.36, p. 199].

Pitches may also be produced by cutting out a string from the rind of the bamboo, which produces an idiochordic bamboo zither. The string is kept stretched by means of a stick serving as bridge at either end. In the middle of the tube, under the string, there may be a small sound hole. The tube is open at both ends. If there is a node at one end, a small square opening will be cut through the wall there and the sound quality can be changed when dampening it by pressing the end of the tube to the chest. There may be more than one string or strings of other material may be fastened to the tube. It may be plucked by the fingers like the *kərantuŋ* of the Temiar in Malaysia and equipped with resonating chambers made from calabash like the *goong* among peoples in the North Central Highland in Vietnam.

Bamboo is perfectly suited for a variety of musical instruments. Many of the instruments mentioned here can be easily made in situations when there is much material at hand like during house building or when working in the fields, and they can easily be thrown away unless they have a ritual meaning of some kind. The variety of constructions and of sonorous qualities is very large. Some of these variants, like zithers and tubes, are considered to have had important roles in the evolutionary processes that have led to instruments that are particularly common in the majority cultures, like tube zithers and xylophones [50.37].

50.2.2 Bamboo Ensembles

Around the year 1900 a French traveler described how he was greeted by bamboo orchestras when approaching Kammu villages in northern Laos and was accompanied by villagers into the village. He even reproduced a photograph of such a group, which is one of the few documentations of minority music from that time. In this passage we get a vivid picture of an ensemble of

idiophones, where the quill-shaped bamboo idiophones were played according to a repeated rhythmic pattern supported by stamping tubes and other bamboo instruments [50.38, pp. 159–160]:

> *Very complex this fanfare. I count six different rhythm parts and most refined at that. Movement of four beats. Three players strike a crotchet at the second half of the fourth beat and two other ones on horseback at the end of the first and the beginning of the second. Not one single time do they falter. The counterrhythm of the alto is likewise perfectly reproduced. The ensemble is harmonic and of correct rhythm . . .*

This is a fine early description of a so-called *colotomic pattern* that functions as an organizing principle in this music. One of the instruments used at that time was the series of concussion tubes that the Kammu call *kltɔ̀ɔŋ*. In the Yùan area usually five pairs of *kltɔ̀ɔŋ* of different sizes are played together. The members of the ensemble strike their individual *kltɔ̀ɔŋ* according to a certain rhythmic pattern. The result is not only a percussion orchestra but also a melodic motif, which varies with each new pattern. The typical pattern of the Rmcùal village – also used for stamping tubes – is remembered by a mnemonic song. Each pair of tubes in the ensemble has a name from lowest pitch to the highest, meaning: beginner, followers (two pairs), stopper and never-stopping (= bourdon). In this manner each person knows the individual position in the ensemble and each person knows the mnemonic song (Figs. 50.4 and 50.5).

The *kltɔ̀ɔŋ* has a number of uses played in ensemble or individually. The overriding function ascribed to them is that of calling and leading a soul of a person or an animal from one point to another, for instance leading the soul of an important visitor or of a slain animal into the village. Another use was leading human souls back to the village in the case when one had concluded that an illness had been caused by one or more of the person's souls going astray [50.39]. Similar concussion tubes are known from many peoples in the area and also from peoples in Borneo and the Philippines with similar uses.

The technique of making music where each person plays only one pitch is sometimes called the *hocket* technique, a term that stems from the musicological term hoquetus that refers to similarly constructed music in the Western medieval church tradition. In Southeast Asia it may be used for instruments other than percussion, like the *dinh tut* of the Edê people in Vietnam that consists of five bamboo pipes of different sizes. The bottom end is closed and the head is cut flatly. It is performed by five females each of whom plays one pipe. It is usually performed in the field or in such rites as the new rice celebration. It is believed that the sound will wake up the rice mother's soul, which will help the rice to grow quickly.

50.2.3 Gongs and Cymbals

Various gongs and cymbals exist among most of the ethnic minorities in the area. In the western part of Southeast Asia the bossed gong dominates and is generally played in combination with a pair of cymbals, while in the eastern part ensembles made up totally or to a large extent by gongs dominate.

Thus gongs and cymbals form a basic unit in much of the music for festive occasions among the Karen and other minorities in Burma [50.40] as well as the Kammu in Laos. The large kettle gongs – or bronze drums as they are commonly called – are gongs with very wide rims. They have a long history and exist in several distinct types [50.7]. Up to the mid-20th century kettle gongs were still manufactured in Burma particularly by the Karen people [50.41]. Karen, Shan, Lolo and Kammu are among those groups who still use

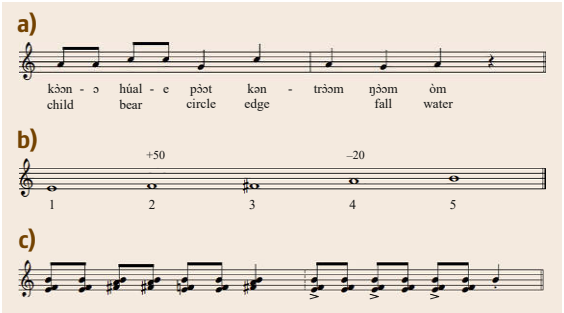


Fig. 50.4 **(a)** Mnemonic song for playing *kltɔ̀ɔŋ*, Rmcùal village, Laos. Translation: The bear cub went around the edge and fell into the water. **(b)** Approximate tuning of *kltɔ̀ɔŋ*. **(c)** *Kltɔ̀ɔŋ* pattern of Rmcùal village, Laos

Fig. 50.5 A set of four pairs of *kltɔ̀ɔŋ* of the Kammu. The photo was taken in northern Thailand in 1979 not in a ceremonial context but as a demonstration. (Photo: D. Tayanin)

kettle gongs to some extent. When played the kettle gongs are often suspended so that the surface is vertical.

Kettle gongs represent large spiritual and economic value. Most of them are inherited goods that belong to individual families and a family's economic status is measured by the number of kettle gongs they own. In Kammu belief the kettle gongs belong to an underground spirit called *róoy yàk* who controlled the underground metals and was dangerous [50.42]. The ancestor spirits are believed to protect the living from evil spirits who could cause sickness, and the living-house – particularly the fireplace – is considered to be the home of the ancestors. The sound of the kettle gongs will call the ancestor spirits to the village and they are thought to be present during the ceremonies where kettle gongs are used.

If a kettle gong is handled with care, i. e., if it is played only on the proper occasions and if the correct sacrifices and prayers are made before using it, it will bring luck and prosperity. If, on the other hand, it is mistreated, the spirits will be angered and the kettle gong will turn into a dangerous object that can produce famine, sickness or even death. There are many stories about kettle gongs being possessed by evil spirits.

A kettle gong of the type commonly used has figurines of frogs on the rim of its face and on its side small figurines depicting elephants, snails, cicadas or rice loafs. Most of these figures as well as other patterns are symbols of fertility and wealth. Depending on these figurines each kettle gong is considered to have certain powers. The main situation for using the kettle gongs is at funerals. The occurrence of a death is first signaled by beating a kettle gong fast and loudly. Villagers and people in neighboring villages working in the fields or hunting in the forest can understand which family is signaling by combining the direction of the sound with the sound of a particular kettle gong.

Fig. 50.6 Cymbals (*left*), kettle gong or bronze drum (*center*) and bossed gong (*right*) being played by Kammu in Yunnan, China, in 1986. The bronze drum is suspended so that its surface is vertical. In the center the star or sun pattern can be seen and two circles have been painted around it. In this case the same person beats the bronze drum and the gong (more often they are played by two persons). Judging from the people watching in the background this seems to be a demonstration or possibly recording situation. (Photo: courtesy of Li Daoyong)

Another important situation for using kettle gongs is buffalo ceremonies, when one or more buffaloes are sacrificed in the case of serious illness, at house-building feasts and at harvest feasts. On these occasions the kettle gongs are played slowly and in an elaborate way, and other instruments like gongs, cymbals and wooden drums may join (Fig. 50.6). In Kammu tradition the kettledrum, which is the largest of the instruments with the longest reverberation, will play a slow pattern consisting of a number of long beats and with a final cadence of three faster beats called *yàal* (slow) and *kmtʎʌn* (fast), respectively. If larger gongs are present they will play shorter beats and a special rhythmic pattern coordinated with the final faster beats. The cymbals are beaten in rhythm, every other beat open with a roll and every other closed. This is actually a slower variant of the same colotomic pattern as that of the bamboo ensemble discussed above. Colotomic patterns of this type with an ending cadence are common in Southeast Asia [50.43].

50.2.4 Gong Ensembles

In the east part of Southeast Asia, particularly in the border area of the Central Highlands of Vietnam, Laos and Cambodia, large ensembles of gongs are common. Gongs are also often played in ensemble with drums and bamboo percussion instruments as well as with flutes and singing. The gong ensembles vary in size but can consist of a large number of gongs. The community thus owns a set of gongs that belong together and range from the largest to the smallest one. Both flat and bossed gongs occur. In the ensemble each player holds one gong that hangs on a string held in the left hand or strung around the shoulder and produces one pitch. The players often move in a long line or in a circle. In each piece they play they are coordinated by a certain pattern. The playing involves striking the gong with the right-hand fist or with a mallet – both soft and hard mallets occur – combined with various damping techniques.

The gongs are played on ceremonial occasions, particularly buffalo ceremonies in which a water buffalo is sacrificed to spirits when it is decided that they need to be placated, for instance in case of severe sickness. The *Muong* people play the gongs at the New Year's festival (according to the lunar calendar) and on other

Fig. 50.7 A gong ensemble of Sedang with flat gongs being struck with beaters on the inside (*four players to the left*) and bossed gongs struck on the boss on the front side by bare fist or beater respectively. The photo was taken in a communal house, a so-called *rong* with a very high pointed roof, near Kontum in central Vietnam, during a recording session in 2007. (Photo: H. Lundström)

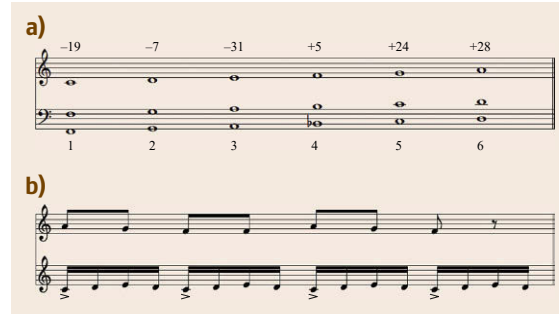


Fig. 50.8a,b A transcription of an ensemble of six gongs of the Ma in Lam Dong Province, Vietnam. **(a)** For each gong the pitches of the fundamental are shown, along with the first partial of one octave higher and the second partial of another fifth higher. **(b)** The melody pattern and the ostinato pattern. After booklet accompanying the CD *Vietnam. Musiques des montagnards* [50.45, p. 70]

happy occasions like celebrating the building of a new house or greeting distinguished guests. In other cases, as with the Sedang, the gong ensemble is used at buffalo ceremonies (Fig. 50.7). Colotomic structures in which the largest gong provides the basic slow pattern and the increasingly smaller gongs are played faster are used by the Bunong in Mondolkiri in northeast Cambodia (Fig. 50.8) [50.44, p. 34].

A gong will produce a fundamental tone and partials. In many cases it is the second partial (the fifth in the octave above the fundamental tone) that dominates the sound and creates melodic movement that varies with the pattern that is used. In an ensemble of only gongs the largest ones often produce an ostinato motif while the smaller gongs produce a melody. As was the case with the above-mentioned Kammu bamboo ensemble the pattern may be remembered by a mnemonic song.

The practice of building music on a combination of colotomic patterns and the hocket technique is widespread in Southeast Asian majority cultures as well. This manner of ensemble playing is well known from the *gamelan* orchestras of Indonesia to the *gagaku* court music of Japan. There is an interesting relationship between bamboo and metal ensembles, which in an evolutionary perspective has been taken as a historical relationship, starting with bamboo and developing into metal ensembles. This is no simple matter, however, for there are actually instances where metal ensembles are replaced by bamboo ensembles for economical or practical reasons.

50.3 Music and Village Life

In traditional village life people who make music and the music they make exist in a framework where not only social conventions but also calendar rules, taboos and religious practices decide when it is appropriate or inappropriate to perform a certain kind of music. There are auspicious days for various activities and inauspicious days that concern everybody in their daily lives. Taboos may be general for a whole village or specific for a particular family or person. There are examples of taboos against music and singing for up to three years in a family where a death has occurred. Other taboos may have the function of reinforcing social conventions or avoiding angering some spirit. Though there is much variation in detail certain religious beliefs are rather general on the level of principle. The dominating belief systems are usually summarized by the terms spirit cult or animism, which means that most living beings and objects are considered to have a spiritual aspect [50.46]. In most cases there are also influences from Hindu and Buddhist practices and thought and in some areas also Christian beliefs.

50.3.1 The Spiritual Context

There is a hierarchy among the spirits depending on what power they are believed to have with which

they can influence the living. Those given great importance are placated through ceremonies of various kinds that often include sacrifices – or gifts – consisting of certain plants or objects and also blood sacrifice, where the water buffalo nowadays is the highest one. In this way the village spirits, ancestor spirits and certain malevolent spirits are treated and balance is upheld. Most spirits are not necessarily evil but nevertheless dangerous for the living. For instance ancestor spirits, who are generally believed to protect the living, may cease doing so if somebody did something wrong or didn't care for them properly. Then people would fall ill and perhaps even die. If this happens, a buffalo ceremony, which in turn is surrounded by a number of taboos, could be staged on a well-chosen day. Similarly, living beings are believed to have souls and there is a hierarchy of souls as well. Apart from human souls the souls of rice and of large game are especially important. These also demand proper action from people. Beside its soul a living animal may also be considered a representation of a certain spirit.

The spiritual aspect is thus very present and in many cases this also involves music. Villagers in general must know the most important of this framework of calendar rules, taboos and religious beliefs in order to handle daily life. In the case of larger ceremonies that involve the whole village, for example harvest feasts, the village elders, village headman and sometimes a master of rituals must be consulted. In most areas there are also shamans, who are considered to be able to communicate directly with spirits and who are called upon in cases of illness. The shaman séances normally involve music in one way or another. Among the Kammu a special set of songs and music on gongs and cymbals send the shaman off to the spirit world and call the shaman back. The shaman will sing with a special voice believed to be that of the spirit in question and in the curing process the shaman will sing certain magic songs or spells. The Hmong shaman travels to the spirit world on a symbolic horse and besides chanting uses gong and rattling metal sounds, like a rattle-sword made from a sword and metal rings [50.39].

50.3.2 Praising the Rice Soul

All matters relating to the farming of rice are of extreme importance since a successful harvest guarantees basic food for the coming year, whereas a small harvest due to weather conditions or other circumstances could cause a village to starve unless other sources of food could compensate, or unless there is wealth enough to buy food. A number of rites are performed in order to call the rice soul to the fields and to make it want to stay there. The details of the ceremonies or rites differ from people to people and also between villages of the same people, but the overall functions have much in common.

Those minorities that live on mountain slopes, like the Kammu people, will make their fields on the mountainside a bit away from the village. An area will be burnt to make a field and the Kammu use a rotation system so that they return to the same place a number of years later. In that manner the forest can recover. Once the field is cleared and the rice has been sown and begun to grow, young people will stay in field houses to watch over the rice and protect it from various animals. In this season there is time for singing, learning songs, and making music on instruments from material at hand. It is also the season when people in Rmcùal and surrounding villages may present their wife-taking clan with certain objects in order to please the rice soul. These objects are a fragrant herb, *crwàaŋ*, and materials for a decorated pole, *cóh*, to be placed by the field house, and three bamboo clappers, called *pɔ́h*, that are placed at the top, the bottom and the middle of the field. The pole and the clappers are made from different sizes of bamboo, including the plated decorations and tassels. The bamboo clappers are connected by ropes to a pole just beside the field house, from where they can be operated (Fig. 50.9).

The large clapper on top, *pɔ́h mòk*, is made from giant bamboo and the smaller one at the bottom, *pɔ́h yíaŋ*, from a large bamboo variety. Their main sounding part is a bamboo culm in which a slit has been cut lengthwise. When one of the ropes is pulled, it lifts the split bamboo of the clapper it is connected to, so that the slit opens and then emits a sound when it closes. In this way, the three clappers can be played in ensemble from the pole in the center of the field. A bamboo clapper can produce several different sounds. The loudest and most resonant one is achieved when the clapper is pulled up with a sudden jerk so that it claps together in upright position. A softer sound is obtained when the clapper is arrested in the middle position. Each of these sounds can be followed by weaker sounds when the two halves of the clapper are allowed to bounce two or more times. When it falls back into its original position against the anvil (*téñ-téñ*), the weakest sound is heard: a thump that cannot be heard from a distance. The third clapper, *pɔ́h làk-làk*, consists of a piece of bamboo with a lengthwise slit placed on top of a pole. A narrow waist is cut so that the slit opens easily. This clapper has a high-pitched rattling sound.

The large and small clappers are played in similar ways, but the large one is too heavy to be played

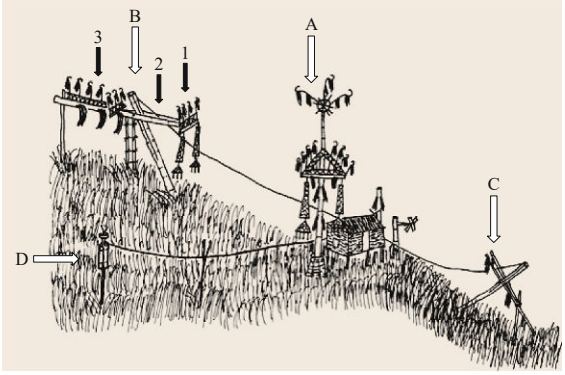


Fig. 50.9 Outline of the placement of the Kammu bamboo clapper family for the rice soul in a field. A) decorated pole, *cóh*; B) large clapper, *pɔ́h mòk*; C) small clapper, *pɔ́h yíaŋ*; D) third clapper, *pɔ́h làk-làk*; 1) forehead, *ktáh*; 2) carrying sling, *prnɔ̀*; 3) hump, *pɔ̀*. (Drawing: D. Tayanin)

rapidly. Once a steady tempo has been achieved, two rhythm patterns are played in alternation, called *yàal* (slow) and *kmtʌ́ʌn* (fast). The third clapper will rattle along in a rather fast tempo. This is exactly how the kettle gong, gong and cymbals are played in the village when there is a ceremony of the kind where the ancestor spirits are called upon. The rice soul thus gets a bamboo version of this ensemble where the largest clapper corresponds to the kettle gong, the smaller clapper to the gong and the smallest third clapper to the cymbals.

There are still other meanings of the clapper ensemble which serves as a symbolic buffalo sacrifice. The front part of the largest clapper is called forehead (*ktáh*) and the hind part hump (*pɔ̀*): this refers to the same body parts of a buffalo, which has a ritual role as sacrificial animal. The part behind the forehead is called carrying sling (*prnɔ̀*), which refers to a carrying sling for babies that in turn implies the rice soul, which is thought of as a small child: the rice child (*kɔ́on ŋɔ́*) (Fig. 50.9).

While the bamboo clappers are set up some members of the recipient family return to the village to prepare food, and then bring the food and wine to the field for a feast. Two cups of wine are offered to the rice soul and a good singer from the recipient family sings in praise of the clappers. A singer in the giver's family sings in reply, praising the rice of the recipient family's fields and belittling the gift [50.47, p. 100ff.].

While the clappers have important social and spiritual meaning they also serve to scare off birds and other animals that might harm the rice. The season when the rice grows is in many places characterized by the sounds of various sound sources designed to serve as scarecrows. Many of them are powered by wind or by water and they range from simple constructions like the humming propeller seen on Fig. 50.9 just beside the field house (*táalɛ́ɛ lòm*) to the complex machines of tuned bamboo tubes like the Sedang *tang koa* [50.18, p. 56ff.], [50.48, pp. 24–25].

Table 50.3 Music during the Kammu farming year, village Rmcùal, Yùan area

Activity	Music	Ascribed meaning
Burning season		
Clearing the fields	Side-blown horn	Calling and the rice soul from the forest to the new field
Sowing season		
Sowing	Music taboo	Loud sounds would anger the *rice child*
Village ceremony	Concussion tubes, drum, gong and cymbals	Receiving the *rice child* to the village common house
Weeding and growing season		
Calling rain	Song; ground harp or ground friction drum	Arousing the dragon spirit to make it rain
Cursed field ceremony	Bumble pipes (by steam over fireplace), gongs and cymbals (side-blown horn taboo)	Main purpose: driving out area spirit that owns a part of the field
Pig ceremony	Music taboo (kettle gong, gong and cymbal) except singing during the party that follows	Calling ancestor spirit to ask protection against evil spirits considered active in this season. Gongs and drums would make the ancestors think a buffalo was sacrificed
Clappers raised in the field	Bamboo clappers; songs	Pleasing the rice soul and making it stay in the field
Harvesting season		
Harvest	Taboo side-blown horn, kettle gong, gongs, cymbal	Loud sounds would anger the *rice child*
Harvest feast in the village	Drum, gongs, cymbals; songs	Greeting the rice soul
Year-end season		
Begging	Song	Getting rid of waste spirits: youths go begging for crops that will be thrown away to the west outside the village

50.3.3 The Farming Year

The greeting of the rice soul is one of a cycle of ceremonies during the farming year that starts with the initiation of the clearing of the fields and ends with the New Year festivities when the old year is finished off (Table 50.3). Most of these ceremonies have a relation to a certain kind of music or musical instrument, sometimes even in the form of taboo against music.

The ascribed meanings of these activities are totally within the spiritual world. This is the insider (or emic) perspective. In an outsider (or etic) perspective one might think of other explanations. It is quite obvious that spiritual beliefs and rules help to regulate human life. Even more so, the cycle of ceremonies during the farming year signals important seasons and the passages from one season to another and also serves as a calendar. Sowing and harvesting are particularly important and critical phases and in the case of the Kammu this is marked by the taboo against music. The different village cultures have variants of this cycle of ceremonies and music [50.21, 49–52]. Many of the beliefs and practices also have parallels among the majority peoples [50.53–56].

50.4 Village Music and Modern Society

The basic units of the minority communities are the villages and the clans in them. The music, rituals and beliefs serve to hold the villages together. The collectively shared systems of music making and its organization therefore play important roles in making the music culture stable over time. Maybe the strongest symbol of the village community as a unit is the wooden drums that in many cases are collectively owned, used in collective ceremonies and kept in the village common house. Thus the wooden slit drum of the Wa people in south China is surrounded by a number of rituals [50.57, 58]. In some areas gongs or bronze drums have a similar role.

50.4.1 Change in Village Cultures

That village cultures are stable does not mean that they are static. Music traditions are known to have gone out of use, for instance the Kammu in north Laos used to have a wooden slit drum much like that of the Wa but it has long ago been replaced by a long wooden drum with two skins [50.59]. In other cases new practices have been integrated into the existing systems or added to them, like music, instruments and songs from neighboring peoples, including the majority peoples. There is a more or less pronounced hierarchy between the minority peoples living in a specific area that may make those who want to raise their status adapt to the culture and life of those who are considered higher in status. In the long run this may lead to hybrid forms of culture and to changed experienced identity [50.60].

These things are part of the continuous change, but change can be more drastic. Historically, for instance, it is believed that the Kammu were once the dominant people in Laos, but that about one thousand years ago they were forced to move up the mountainsides by a migration of the Thai people from the north, who are now the majority population. As recently as the beginning of the 19th century conflicts in south China made many Hmong move into northern Laos and Vietnam where they generally made their villages and swiddens on the mountaintops.

The Kammu village Rmcùal that has been used as reference several times in this chapter is actually an example of rather dramatic change in recent times. In fact it no longer exists, so all information about its music and traditions is historical. After a period of depopulation the last family left the village in 1999 and actually the whole area where Rmcùal was is now deserted. The reasons for this are many. Wars – particularly the internal war in Laos in the 1960s and early 1970s – made Kammu people in the area become soldiers on different sides in the conflict. Young men left their villages and eventually there were too few people in the area to uphold a necessary production of rice. Gradually people started moving to new settlements in the lowlands or to suburbs of nearby towns, like Nam Tha. This has meant that the music that was related to work and to social organization has lost its base and become abandoned or recontextualized. On the other hand, people have better access to education, healthcare and trade centers [50.61]. This development is not limited to this particular area, but also occurs elsewhere [50.62, pp. 4–5].

Generally speaking, while the systems of music organization that have been exemplified in this chapter have started to dissolve, those kinds of music that can be more easily recontextualized are music relating to important ceremonies, like funerals, and social singing. Thus the recited prayer that guides the soul of a dead person to a place of rest located in China is still being used among Hmong in the US and those who live in exile, have started cultural associations that also promote

musical instruments, like the Hmong mouth organ or the gongs and cymbals of the Kammu [50.63].

Wars and political changes have affected minorities in the whole of Southeast Asia and continue to do so when it comes to national policies and legislation concerning land rights. Generally the uphill farmers who use swidden techniques are believed to cause deforestation and soil erosion, which has led to reforestation programs and resettlement programs that have changed their practical circumstances thoroughly.

50.4.2 Survival in Modern Society

In this chapter the music of the ethnic minorities has so far been treated as belonging to their own separate universes on the village level. Though there is interaction with neighboring peoples the music of villages of the same people tend to share the same instruments, instrument names, vocal genres and genre names over long distances. Similarly, even though there is interaction with the majority peoples, the ethnic villages have more in common with other villages of the same ethnicity than with the majority in their vicinity. This holds true as long as the village traditions are rather intact. The radical changes these communities have undergone, particularly since the mid-20th century, are changing this picture and is a development that can be expected to continue. Therefore the music of people of minority background must also be seen as part of the whole national music culture in the country where they live. Still, the music cultures of the ethnic minorities cannot be seen as subcultures of the national music culture, but rather as separate music cultures.

In Vietnam music of ethnic minorities was incorporated in the development of a new national music that began in the north in the mid-1950s and nationally after the Vietnam War in 1975. In line with politics in the Soviet Union and China the aim was to build a revolutionary music culture. This would be achieved by eliminating aspects of the old society considered negative, transforming customs considered backward and building a new music culture [50.64, p. 93]. This is usually referred to as *neotraditional music* and consists of reformed traditional music, political songs and modern compositions [50.64]. The latter are often composed with motifs and rhythms inspired by ethnic music. The neotraditional music includes modified versions of the bamboo xylophone *trung* and the bamboo tubes *k'long put* along with modified versions of Vietnamese instruments and is performed by conservatory-trained musicians in official situations, nationally and abroad (Fig. 50.10) [50.65]. On a symbolic level it represents a nation consisting of all its ethnic groups of which the majority people Viet (or Kinh) is one. The procedure is selective and those instruments and musical pieces that have an ethnic background were moved from their traditional contexts and placed in a new context.

Traditional Vietnamese music, for instance the music of the zither *danh tranh*, was moved from intimate music situations to stage and TV performances involving less – or no – improvisation [50.64]. Similarly, ethnic music performed by musicians of minority background has, much after the Chinese model, also become staged, particularly at festivals and at competitions. This is another example of selective recontextualization since the stage and television contexts have their own sets of practices and criteria that the music is adapted to. Staging is closely related to tourism, whether music is literally performed on a stage or in a village specializing in tourism [50.66–69].

In 2008 the *Space of gong culture* was inscribed on UNESCO's list of Intangible Cultural Heritage of Humanity. The gong traditions of the Central Highland in Vietnam became the first ethnic music to get this status. Other countries in East Asia have been active with cultural heritage designations for some time, particularly so Japan and in the 2000s China [50.70]. In Vietnam this has been used in order to safeguard and revitalize music that was not previously cared for or – as in the case of the vocal chamber music *ca trù* – discouraged

Fig. 50.10 An ensemble from the Lam Dong province in the Central Highlands of Vietnam performing at the national festival for traditional instruments in Da Lat, Vietnam, 2014. It appears to be a local neotraditional ensemble with modified instruments consisting of Viet musicians representing the minorities of the province. *From left to right*: suspended gongs, suspended bamboo xylophone, bamboo xylophone with resonating tubes, festive decorated pole (Fig. 50.9); behind the pole a variant of the *klong put* bamboo instrument, drum, stone xylophone, *klong put* bamboo instrument. (Photo: E. Wettermark)

or even banned [50.71]. While this status will serve as a support for those who carry on the tradition it also leads to an increased focus on that certain activity by media and tourist agencies [50.69, 72].

Since its start in the 1950s the Vietnamese Institute for Musicology has conducted fieldwork studying and documenting folk music, including music of the ethnic minorities [50.73]. The recordings are archived and are gradually being digitized. Documentation is also organized in collaboration with local culture centers and many recordings that are published on CD, VCD or DVD formats are aimed at local use. Documented recordings are an important asset once a music has been discontinued and revival movements may occur. The fieldwork normally consists of field study, documentation, archiving and publishing. In many cases knowledge is lost in ethnic villages and often only the oldest people know the music. One has noticed loss not only of musical knowledge in general but also loss of knowledge for making musical instruments: selecting material, treating the material and handling the instruments. In recent years fieldwork has also been combined with education in order to revitalize the transmission process: senior folk performers recognized by the communities will teach people of the whole community with priority on the young generation [50.74, pp. 39, 57–59]. These transmission classes aim at empowering people to revitalize their music culture as a living culture rather than preserving its music as frozen objects. Similar initiatives may be taken from within the culture, which is the case among the Tampuan in Ratanakiri Province, northeastern Cambodia, where songs in the traditional style that deal with modern topics are being created and used [50.75].

The Vietnamese example demonstrates that the music culture of an ethnic minority group or village cannot only be seen as a separate unit but also must be seen in relation to the national music culture of which it is a part and to the processes going on there. This is crucial when thinking of the future of this music. The circumstances differ in detail between the countries in the area but are similar on a general level [50.76]. The processes that are mainly conducted by *external* agents like politicians, producers, researchers and development workers are complex and need to be problematized in order to understand their effects [50.62]. There are many and often conflicting views on how these various agents relate to the future of the music cultures of the minorities and these conflicting views can also be found within the group of ethnomusicologists. The old relation to traditional music that advocates preservation with arguments of authenticity stands against the view that the music cultures should be left alone to handle the situation and that loss of music traditions is a normal and unavoidable thing. There are those who see stage shows, competitions and tourism as leading to folklorization that is opposed to authenticity, while others prioritize the fact that these are some of the few areas where performers can make a financial gain.

It is a fact, though, that all these circumstances, which are both opportunities and threats to the minority music cultures, exist and will continue to exist for the foreseeable future. The situation is not unique to Southeast Asia but can be seen in a wider perspective. In the international research project *Sustainable futures for traditional musics* matters of sustainability in 11 different contexts in the world were studied by a comparative method in order to increase knowledge of these processes [50.77, 78]. As the understanding of the situation in Southeast Asia may benefit from studies in a wider geographic perspective, likewise interdisciplinary approaches are important, not least with linguistics, where endangered languages have been an issue for some time and are also debated [50.79, 80]. Music loss among endangered musics has much in common with language loss – actually the two are closely connected particularly in the area of singing, which in many cases demands a full command of the language.

This chapter has included examples of traditional village-based music cultures of the Southeast Asian minority peoples and it has dealt with the present-day situation after substantial changes in life patterns and a degree of recontextalization into the main national music cultures. What lies in the future is a large and important field for musicological research that takes into account both the traditional and the present conditions, includes local culture politics, and also relates to global musicological and interdisciplinary research.

50.A Appendix: Recordings

50.A.1 CDs

- *Baishibai. Songs of the Minority Nationalities of Yunnan*. PAN 2038CD (Pan Records, Leiden 1995)
- *Bamboo on the Mountains. Kmhmu Highlanders from Southeast Asia and the United States* SFW 4056 (Smithsonian Folkways Recordings, Washington 1999)
- *Karenni. Music from the Border Areas of Thailand and Burma*. Pan Records Ethnic Series, PAN 2040CD (Pan Records, Leiden 1994)
- *Music from Vietnam 3. Ethnic Minorities*. CAP 21479 (Caprice Records, Stockholm 1995)
- *Music from Vietnam 5. Minorities from the Central Highland and Coast.* CAP 21674 (Caprice Records, Stockholm 2003)
- *Music from Villages in Vietnam 1–2*. (Vietnamese Institute of Musicology, Hanoi 2009)
- *Music of Laos. Khmou', Oï, Brao, Lao, Phou-noï. Kui, Lola, Akha, Hmong and Lantene traditions*. W 260118 (Maison des Cultures du Monde, Paris 2004) inedit
- *Vietnam. Music of the Montagnards*. CNR 2741085/6 (Le Chant du Monde, Arles 1997)
- Vietnam Institute for Musicology: Series of CD, VCD and DVD recordings

50.A.2 LPs

- *Cambodge. Musique "Samrê" des Cardamomes recueillie par Jacques Brunet*. LD 112 (Disques Alvarès, Paris 1969)
- *Musique des tribus Chinoises du Triangle d'Or*. ARN 33535 (Arion, Paris 1980)
- *Musique Mnong Gar du Vietnam*. Collection Musée de l'Homme, OCR 80 (Disques Ocora, Paris, recorded in 1958)
- *Musiques du Cambodge des forêts*. Anthologie de la Musique des Peuples (1976)
- *Musiques du Viet-Nam*. Disques BAM LD 434, n.d.
- *The Protomalayans of Malacca*. An anthology of South-East Asian Music, BM 30 L 2563 (Bärenreiter-Musicaphon, Kassel, recorded in 1963)
- *The Negrito of Malacca*. An anthology of South-East Asian Music, BM 30 L 2562 (Bärenreiter-Musicaphon, Kassel, recorded in 1963)
- *The Senoi of Malacca*. An anthology of South-East Asian Music, BM 30 L 2561 (Bärenreiter-Musicaphon, Kassel, recorded in 1963)

References

50.1 T. Miller, S. Williams (Eds.): *Southeast Asia*, The Garland Encyclopedia of World Music, Vol. 4 (Garland, New York 1998)

50.2 H. Oesch: Aussereuropäische Musik Teil 2. In: *Neues Handbuch der Musikwissenschaft*, Vol. 9, ed. by C. Dahlhaus (Laaber/Athenaion, Laaber/Wiesbaden 1987)

50.3 E. Mareschal: *La musique des Hmong* (Musée Guimet, Paris 1976)

50.4 H.P. Larsen: *Lisuernes musik [The music of the Lisu]*, M.A. Thesis (Univ. of Copenhagen, Copenhagen 1981)

50.5 M. Roseman: *Healing Sounds from the Malaysian Rainforest: Temiar Music and Medicine* (Univ. of California Press, Berkeley 1991)

50.6 H. Lundström: *I will send my Song. Kammu Vocal Genres in the Singing of Kam Raw* (NIAS Press, Copenhagen 2010)

50.7 A.J. Bernet Kempers: *The Kettledrums of Southeast Asia. A Bronze Age and its Aftermath*, Modern Quarternary Research in Southeast Asia, Vol. 10 (A.A. Balkema, Rotterdam/Brookfield 1988)

50.8 T.N. Thanh: *Musical Instruments of Vietnam's Ethnic Minorities (A Partial Introduction)* (The Gioi Publishers, Hanoi 1997)

50.9 J. Boulbet: *Dialogue lyrique des Cau Maa'*, Bulletin de l'École Française d'Extrême-Orient (École Française d'Extrême-Orient, Paris 1972)

50.10 J. Dournes: *Florilège jörai*, La Collection Sudestasie Bibliotheque (Editions Sudestasie, Paris 1987)

50.11 J. Dournes: *Florilège sré*, La Collection Sudestasie Bibliotheque (Editions Sudestasie, Paris 1990)

50.12 J. Mottin: *55 Chants d'Amour Hmong Blanc* (The Siam Society, Bangkok 1980)

50.13 F. Proschan: *Kmhmu Verbal Art in America: The Poetics of Kmhmu Verse*, UMI (Univ. of Texas, Ann Arbor/Austin 1989)

50.14 H. Lundström, D. Tayanin: *Kammu Songs. The Songs of Kam Raw* (NIAS Press, Copenhagen 2006)

50.15 P. Lefèvre-Pontalis: Chansons et fêtes du Laos. In: *Collections de contes et chansons populaires 21–22* (Ernest Leroux, Paris 1896)

50.16 H. Roux, V.C. Tranh: Les Tsa Khmu, Bull. l'École Fr. d'Extrême Orient **27**, 169–222 (1927)

50.17 H.P. Larsen: The music of the Lisu of northern Thailand, Asian Folk. Stud. **43**, 41–62 (1984)

50.18 G. de Gironcourt: Recherches de géographie musicale en Indochine, Bull. Soc. Études Indoch. **17**(4), 1–174 (1943)

50.19 G. Jähnichen: The last of their kind: Khmu flute songs. In: *Studia Instrumentorum Musicae Pop-*

ularis II (New Series), ed. by G. Jähnichen (MV-Wissenschaft, Münster 2011) pp. 133–147
50.20 G. Condominas: Chansons Mnong Gar, France-Asie **87**(Autumn), 648–656 (1953)
50.21 G. Condominas: *We have eaten the forest. The story of a Montagnard village in the Central Highlands of Vietnam* (Penguin, London 1977)
50.22 G. Condominas: *Mnong Gar Music of Vietnam*. Disques Ocora: Collection Musée de l'Homme, Anthology of Proto-Indochinese music 1 (n.d.)
50.23 A. Schimmelpenninck: *Chinese Folk Songs and Folk Singers. Shan'ge Traditions in Southern Jiangsu* (CHIME Foundation, Leiden 1997)
50.24 C.J. Compton: *Courting Poetry in Laos: A Textual and Linguistic Analysis*, Center for Southeast Asian Studies Special Report 18 (Northern Illionois Univ., DeKalb 1979)
50.25 K. Wenk: Gesänge und Opfersprüche der Lawa in Nordthailand, Nachr. Ges. Nat. Völkerkd. Ostasiens **85/86**, 198–218 (1962)
50.26 I. Skog: *North Borneo Gongs and the Javanese Gamelan. Studies in Southeast Asian Gong Traditions*, Studies in Musicology, Vol. 2 (Stockholm Univ., Stockholm 1993)
50.27 A.P. Cohen: Jeh music. In: *Notes from Indo-China*, ed. by M. Gregerson, D. Thomas (SIL Museum of Anthropology, Dallas 1980) pp. 85–97
50.28 M.J. Gregerson: *The Rengao of Vietnam: An Ethnography of Texts*, PhD Thesis (Univ. of Texas, Arlington 1991), UMI no. 9217631
50.29 K.G. Izikowitz: *Lamet, Hill Peasants in French Indochina*, Etnologiska Studier 17 (Etnografiska Museet, Göteborg 1951)
50.30 L. Ó Briain: *Hmong Music in Northern Vietnam: Identity, Tradition and Modernity*, E-thesis (Univ. of Sheffield, Sheffield 2012)
50.31 Music from Villages in Vietnam 1 (Vietnamese Institute of Musicology, Hanoi 2009), track 14
50.32 C. Falk: The Dragon taught us: Hmong stories about the origin of the free reed pipes Qeej, Asian Music **35**(1), 17–56 (2004)
50.33 N. Tapp: Qha Ke (guiding the way) from the Hmong Ntsu of China, 1943, Hmong Stud. J. **9**, 1–36 (2008)
50.34 J.-O. Svantesson, Kàm Ràw (D. Tayanin), K. Lindell, H. Lundström: *Dictionary of Kammu Yùan Language and Culture* (NIAS Press, Copenhagen 2014)
50.35 J. Brunet: *Cambodge. Musique "Samrê" des Cardamomes*. LD 112 (Disques Alvarès, Paris 1969), booklet to LP recording
50.36 K.G. Izikowitz: *Över Dimmornas Berg [Across Foggy Mountains]* (Albert Bonniers, Stockholm 1944)
50.37 T. Grame: Bamboo and music: A new approach to organology, Ethnomusicology **VI**(1), 8–14 (1962)
50.38 A. Raquez: *Pages Laotiennes* (F.-H. Schneider, Hanoi 1902)
50.39 J. Lemoine: Commentary: The (H)mong Shamans' power of healing: Sharing the esoteric knowledge of a great (H)mong Shaman, Hmong Stud. J. **12**, 1–36 (2011)
50.40 G. Douglas: Performing ethnicity in Southern Shan State, Burma/Myanmar: The Ozi and Gong traditions of the Myelat, Ethnomusicology **57**(2), 185–206 (2013)
50.41 R.M. Cooler: *The Karen Bronze Drums of Burma: Types, Iconography, Manufacture, and Use* (Brill, Leiden, New York, Köln 1995)
50.42 H. Lundström, D. Tayanin: Kammu gongs and drums I: The kettlegong, gongs, and cymbals, Asian Folk. Stud. **40**(1), 65–86 (1981)
50.43 J. Becker: Percussive patterns in the music of mainland Southeast Asia, Ethnomusicology **12**(2), 173–191 (1968)
50.44 J. Brunet: Les instruments de musique de la province de Mondolkiri, Études Cambodg. **1**, 31–34 (1965)
50.45 Q.H. Tran, H. Zemp: *Notes on the Music (Booklet accompanying the CD Vietnam. Music of the Montagnards)*. Le Chant du Monde CNR 2741085/6 (1997) pp. 68–73, 92–109
50.46 D. Tayanin: *Being Kammu. My Village, My Life* (Cornell Univ./Southeast Asia Program, Ithaca/New York 1994)
50.47 K. Lindell, H. Lundström, J.-O. Svantesson, D. Tayanin: *The Kammu Year. Its Lore and Music*, Scandinavian Institute of Asian Studies, Studies on Asian Topics, Vol. 4 (Curzon, London, Malmö 1982)
50.48 P. Collaer: Hydraulisches Schlagröhrenspiel aus Bambus. Gebiet der nationalen Minderheit der Sedang im mittleren Vietnam, Musikgesch. Bild. I: Musikethnol. **3**, 24 (1979)
50.49 P. Bitard: Rites agraires des Kha Braou, Bull. Soc. Études Indoch. **27**(1), 9–18 (1952)
50.50 M.M.J. Kemlin: Rites agraires des Ruengao, Bull. l'École Fr. d'Extrême-Orient **8**, 493–522 (1908)
50.51 M.M.J. Kemlin: Rites agraires des Ruengao, Bull. l'École Fr. d'Extrême-Orient **9**, 131–158 (1909)
50.52 A. Maurice: Trois fêtes agraires Rhadé, Bull. l'École Fr. d'Extrême-Orient **45**(1), 186–207 (1951)
50.53 C. Archaimbault: Les rites agraires dans le Moyen-Laos, France-Asie **16**, 160–163, 1185–1194, 1274–1283 (1959)
50.54 E. Porée-Maspero: *Rites agraires des Cambodgien 1* (Mouton, Paris 1962)
50.55 E. Porée-Maspero: *Rites agraires des Cambodgien 2* (Mouton, Paris 1964)
50.56 E. Porée-Maspero: *Rites agraires des Cambodgien 3* (Mouton, Paris 1969)
50.57 T. Obayashi: The wooden slit drum of the Wa in the Sino-Burman border area, Beitr. Japanol. **3**(2), 72–88 (1966)
50.58 Y. Zhao: A tentative research on the ethnic culture of the wooden drum of Wa ethnic group. In: *Dynamics of Ethnic Cultures Across National Boundaries in Southwestern China and Mainland Southeast Asia: Relations, Societies, and Languages*, ed. by Y. Hayashi, G. Yang (Ming Muang, Chiang Mai 2000) pp. 233–245
50.59 H. Lundström, D. Tayanin: Kammu gongs and drums II: The long wooden drum and other drums, Asian Folk. Stud. **40**(2), 173–189 (1981)
50.60 G. Evans: Ethnic change in highland Laos. In: *Laos. Culture and society*, ed. by G. Evans (Silkworm, Chiang Mai 1999) pp. 125–147

50.61 O. Évrard: Following Kàm Ràw's trail. In: *Hunting and Fishing in a Kammu Village. Revisiting a Classic Study in Southeast Asian Ethnography*, ed. by D. Tayanin, K. Lindell (NIAS Press, Copenhagen 2012) pp. 1–28

50.62 F. Bourdier: Indigenous populations in a cultural perspective. The paradox of development in Southeast Asia, Anthropos **103**, 1–12 (2008)

50.63 L. Schein: Hmong/Miao transnationality: Indentity beyond culture. In: *Hmong/Miao in Asia*, ed. by N. Tapp, J. Michaud, C. Culas, G.Y. Lee (Silkworm, Chiang Mai 2004) pp. 273–290

50.64 T.H. Le: *Dán Tranh Music of Vietnam. Traditions and Innovations* (Australia Asia Foundation, Melbourne, Tokyo 1998)

50.65 M. Arana: *Neotraditional Music in Vietnam*, The Journal of Vietnamese Music (Nhac Viet, Kent, Ho Chi Minh City 1999)

50.66 C. Rubenstein: The cultural show: Is it culture or what and for whom?, Asian Music **23**(2), 1–62 (1992)

50.67 S. Touhy: The sonic dimensions of nationalism in modern China: Musical representation and transformation, Ethnomusicology **45**, 107–131 (2001)

50.68 G. Jähnichen: Quoting minorities in Viet performing arts – A short history of musical subordination. In: *Music and Minorities in Ethnomusicology: Challenges and Discourses from Three Continents*, ed. by U. Hemetek (Institut für Volksmusikforschung und Ethnomusikologie, Vienna 2012) pp. 13–24

50.69 Q.H. Tran: Westernization and modernization of the gongs of the highlanders of Central Vietnam: Are they good for the development of their music in the globalization of World music? In: *Music and Minorities in Ethnomusicology: Challenges and Discourses from Three Continents*, ed. by U. Hemetek (Institut für Volksmusikforschung und Ethnomusikologie, Vienna 2012) pp. 113–118

50.70 K. Howard (Ed.): *Music as Intangible Cultural Heritage. Policy, Ideology, and Practice in the Preservation of East Asian Traditions* (Ashgate, Surrey, Burlington 2012)

50.71 E. Wettermark, H. Lundström: Ca trù. In: *Sustainable Futures for Music Cultures: An Ecological Perspective*, ed. by H. Schippers, C. Grant (Oxford Univ. Press, New York 2016)

50.72 K. Howard (Ed.): *Music as Intangible Cultural Heritage. Policy, Ideology, and Practice in the Preservation of East Asian Traditions*, SOAS Musicology Series (Ashgate, Farnham, Burlington 2012)

50.73 H.L. Dang: *Vien am Nhac* (Vietnam Musicology Institute, Hanoi 2000)

50.74 Vietnamese Institute for Musicology: *Sixth Symposium of the ICTM Study Group on Music and Minorities. Second Symposium of the ICTM Study Group on Applied Ethnomusicology, 19–30 July 2010. Program and Abstracts* (Vietnamese Institute for Musicology, Hanoi 2010)

50.75 T.W. Saurman: Singing for survival in the highlands of Cambodia: Tampuan revitalization of music as cultural reflexivity. In: *Music and Minorities in Ethnomusicology: Challenges and Discourses from Three Continents*, ed. by U. Hemetek (Institut für Volksmusikforschung und Ethnomusikologie, Vienna 2012) pp. 95–104, Klanglese 7

50.76 T. Miller, S. Williams: The impact of modernization on traditional musics. In: *Southeast Asia*, The Garland Encyclopedia of World Music, Vol. 4, ed. by T. Miller, S. Williams (Garland, New York 1998) pp. 113–143

50.77 H. Schippers: Applied ethnomusicology and intangible cultural heritage: Understanding 'ecosystems' of music as a tool for sustainability. In: *Oxford Handbook of Applied Ethnomusicology*, ed. by S. Pettan, J.T. Titon (Oxford Univ. Press, New York 2015) pp. 134–157

50.78 H. Schippers, C. Grant (Eds.): *Sustainable Futures for Music Cultures: An Ecological Perspective* (Oxford Univ. Press, New York 2016)

50.79 C.F. Grant: *Strengthening the Vitality and Viability of Endangered Music Genres: The Potential of Language Maintenance to Inform Approaches to Music Sustainability*, PhD Thesis (Griffith Univ., Brisbane 2012)

50.80 J.-O. Svantesson, N. Burenhult, A. Holmer, A. Karlsson, H. Lundström: *Language Documentation and Description*, Vol. 10 (School of Oriental and African Studies, London 2012)

51. Music Archaeology

Ricardo Eichmann

Music archaeology is a rather recent field of academic studies that emerged in the 19th century and provides access to musical cultures of ancient civilizations. Modern research, characterized by expanding interdisciplinary methodological approaches, has its roots in the 1970s. Music archaeology explores ancient music and archaeoacoustic phenomena embedded in archaeological contexts and objects, iconographic representations, and written sources. After a long period of processing and classifying such data, cultural turns in the past two decades have increasingly influenced the interpretation of archaeological music evidence.

The sources for reconstructing the sound and musical culture produced by ancient civilizations are accessed by classical humanities disciplines, particularly archaeology, and collected, classified and interpreted from different perspectives [51.1, 2]. Conditioned by the research goals and methodologies of the humanities, archaeology investigates human sociocultural evolution and the underlying technical, economic and social innovations [51.3]. Toward this end music archaeology cooperates closely with disciplines in the natural sciences and enhances their methods and procedures. This productive collaboration yields robust, verifiable data about the living conditions of ancient societies and their culture. Ideally, music archaeology is pursued in interdisciplinary cooperation between archaeologists and musicologists, whereby the methods applied depend in each case on the particular issues and the nature of the source material. Sources associated with music and found in an archaeological context were systematically researched by individual scholars already in the 19th century. The work of the musicologist Carl Engel on *the music of the most ancient nations*, dating from 1864, is outstanding for this time [51.4]. His investigations included an analysis of archaeological relics and findings on the ancient advanced civilizations of Western Asia and Egypt. Later, lasting impulses for interdisciplinary archaeological music research relating to Western Asia and Egypt emanated from such musicologists as *C. Sachs* [51.5, 6] and *H. Hickmann* [51.7] in the early decades of the 20th century. A synthesis on prehistoric music in Europe was available in 1934 [51.8]. Also deserving of mention is an internationally acclaimed series on the iconographic history of music (*Musikgeschichte in Bildern*) edited by *Heinrich Besseler*, *Max Schneider* and *Werner Bachmann*, which also considered the *music of antiquity* [51.9] and comprised volumes on Greece, Etruria and Rome, Egypt, Mesopotamia, Central Asia, India, and America.

Continuous research on music in antiquity is, however, only discernible since the late 1970s. Since that time, an increasingly expanding, interdisciplinary field of work has been established [51.2, 10–12]. Today, the field is usually called music archaeology (Musikarchäologie, archéologie musicale) or archaeomusicology, sometimes archaeo-music (Archäomusikologie, archéo-musicolgie), depending on the disciplinary environment with which the authors or initiators of the working teams most closely associate themselves; musicology or archaeology. In several places the name of the field is controversial (see, e.g., [51.2, 12]). In this article the term *music archaeology*, which was introduced by researchers already at an early date [51.13], is understood as the designation for a specialization within archaeological research, analogous to *social archaeology*, *economic archaeology* and *industrial archaeology*. Other than several brief

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_51

surveys of developments over the past 40 years (e.g., [51.1, 2, 14]), a comprehensive, systematic history of archaeological music research worldwide is not yet available.

Numerous multidisciplinary research groups address different geographical settings and historical periods or different topics of musical culture across a spectrum ranging from the fabrication of musical instruments to aspects of musical life. In this context, basic research in the field of music archaeology has also led to the production of systematic catalogs of relevant relics and findings as well as monographs on specialized topics investigated in historical depth.

Usually embedded in an archaeological context, music archaeology is also a focus of the emerging field of archaeoacoustic research, which is also concerned with nonmusical acoustic phenomena [51.15, 16]. This is reasonable, since it cannot be determined whether many sound-generating objects, such as horns, whistles, bells and drums, were ever used to produce music, or just had a signaling function. This is also the case for metal artifacts from prehistoric European hoards, which stimulate archaeoacoustic investigation because of their sound qualities [51.17], and for many ancient American whistles that served to imitate animal sounds [51.18, 19].

A comprehensive manual of the theory and practice of music archaeology does not yet exist, although efforts in this direction are being undertaken in the international research community [51.2, 10, 20], whose members regularly interact at conferences held with increasing frequency. An overview of past international conferences and proceedings that provide information about music archaeology activities and research fields worldwide is to be found on the websites of music archaeology research teams (e.g., [51.21–24]).

51.1 Methods

Music archaeology's primary sources are cultural environments significant for their musical contexts. They are typically studied by archaeologists during stratigraphic excavations using modern methods of securing, classifying and documenting remnants of human activity. These include physical survey techniques based on geomagnetics, electrical resistivity imaging and ground-penetrating radar to locate buried structures or portable objects; excavation techniques including find-related fine-scale stratigraphic excavations to contextualize relics; documentation methods including geographic information systems and satellite-based measurement systems to determine the precise georeferenced location of relics and findings; techniques for salvaging finds, such as en-bloc recovery of fragile objects for later examination under laboratory conditions; three-dimensional (3-D) scanning technology, computer tomography and x-ray imaging to visualize objects still in an en-bloc recovery state and to determine the constructive features of artifacts; 3-D printing techniques to reproduce artifacts; restoration techniques to stabilize fragile objects; chemical investigation of material remnants in order to reconstruct surface treatments and the formation of patina; microscopy to detect traces of the fabrication and use of objects; technologies from the natural sciences to determine the age of relics and findings, including optically stimulated luminescence, radiocarbon analysis, dendrochronology and x-ray fluorescence spectroscopy; archaeobotanical, archaeozoological and archaeometallurgic methods for identifying materials; the study of isotopes suitable as markers for tracing provenance; and finally, anthropology, pathology and genetic technology to gain insights about the living conditions of individuals and the demographic characteristics of a society.

Also of importance are written records that permit conclusions about past musical life as well as iconographic representations of music scenes, musicians and musical instruments. The significance of this kind of source material depends on the philological, historical and semiotic assessment of the particular text or image evidence. Everyday knowledge is seldom adequate for interpreting texts and images in terms of music archaeology. As a rule, knowledge of ancient languages is essential for studying ancient text passages and existing translations, particularly if the texts of ancient authors have not been translated or annotated by philologists trained in musicology or music archaeology. The case is similar for the image repertoire of the past, which makes use of iconographic codes that seldom correspond to those of today. The iconographic features of an ancient civilization can only be reliably interpreted against the background of their cultural context. This requires an application of analytic archaeological methodologies to art that presuppose a firm knowledge of the artistic handicraft output and stylistics of a community. It must also be recognized that documented source materials always represent only a very small fraction of a past culture. For ancient Egyptian culture it can be asserted, as *Alexandra von Lieven* [51.25, p. 99], emphasizes,

> *that most of our material comes from tombs, ... material from younger epochs is much more abundant*

than from older periods (this is enhanced by the fact that older material is stratigraphically lower and thus more vulnerable to a high ground water level), that the continued use of traditional material can distort our view of later epochs, that texts and depictions contain only parts of the reality, and that possible oral traditions are completely lost to us.

Computer-based analyses of characteristics are used to establish typologies that can clarify technological, chronological and chorological issues. *G. Kolltveit* elaborated a paradigmatic typology of European Jew's harps, which can be identified since the 13th century [51.26, pp. 65–153]. Iconographic sources are also suitable for studying musical instruments in large geographical areas, although the meaning of images sometimes remains vague because of stylistic differences, one example being the lyres that emerged in ancient Western Asia starting from the fourth millennium BCE [51.27].

The question of the extent to which archaeological music research can take into account the insights of ethnomusicology has been and continues to be discussed [51.1, 2, 28, 29]. In this context, methodological and concrete cultural aspects that have an influence on archaeological music research have to be distinguished. From an ethnographic standpoint they primarily include terminology, classification [51.30], forms of presentation, playing techniques, and music theory. From an ethnological perspective they comprise theoretical reflections about the sociocultural relevance of the music and how analogies and symbols are handled [51.31, p. 177]. According to the latter orientation, musical phenomena are considered to be products of social and cognitive processes, just like other cultural features of a society, and perhaps only meaningful within very narrowly limited spheres, not to be regarded as the result of some kind of purposeful evolution [51.3], [51.2, p. 5], [51.31]. The extent to which analogies arising from an ethnographic perception can be legitimately used for music archaeological interpretations depends on how the ethnoarchaeological research approaches are assessed [51.2, pp. 6–7] and in any case requires verifiable justification.

Ethnomusicological analogies can be regarded as a chance to better understand the use of sound in specific cultures but it can be risky to use ethnomusicological analogies without having studied the proper archaeological context. An intensive ethnographic survey of analogies can make the reconstruction of ancient instruments more plausible. [51.20, p. XIII]

In comparison to ethnomusicology, which is primarily concerned with relatively brief periods of time (*short-term perspective*), the research field of music archaeology extends back over several millennia (*long-term perspective*). This is advantageous for archaeological music research because it allows gradual changes to be studied over long periods of time. It also discourages an ethnocentric view of music in favor of an independent appreciation of the most varied sound cultures [51.20, p. XIII]. Whether and to what extent universally valid musical absolutes gained from ethnomusicological perspective can be regarded as a prerequisite for broader music archaeological interpretations is a matter of debate [51.32]. This issue has research potential.

The methods and procedures employed by archaeologists are also used to investigate sources relevant for music history from more recent contexts, e.g., when details of written or oral transmission are to be verified with other approaches, or are not available [51.33]. Archaeological music concepts have been employed in modern history, whereby the idea of *excavation* should be understood more in a metaphorical sense and as referring to research on ancient texts [51.34].

As is the case for modern archaeology, today music archaeology can only obtain verifiable, fruitful results in multidisciplinary and interdisciplinary research partnerships. Besides the classical humanities (philology, history, archaeology, art history), these need to draw on musicology (organology, music theory), ethnology (historical ethnomusicology, oral traditions, ethnography) and the natural sciences (physics, survey methods, acoustics, chemistry, materials science, patina studies) [51.2, p. 4, Fig. 1].

51.2 Research Topics

In practice, music archaeology is dominated by research topics related to the excavation, identification and organological study of musical instruments and their sociocultural context. One of the strengths of archaeo-organological research is the macroscopic and microscopic analysis of surface traces related to the production, use, damage or ground deposition (taphonomy) of a musical instrument. To this can be added the reconstruction of the successive stages of instrument construction (*chaîne opératoire*), starting with the supply of materials and tools, extending to the various stages of crafting, and ending with a ready-to-play instrument. This makes it possible to draw conclusions about the complexity and cognitive preconditions for

the production process [51.35], to the benefit of numerous artifact studies. Examples include the interpretation of Paleolithic aerophones [51.36, 37] and the tone reservoir and function of an Egyptian necked bowl lute, which was produced, as well as subsequently repaired, in late antiquity [51.38]. As with other empirical cultural sciences, music archaeology is thus in a position to access the knowledge embedded in artifacts and to interpret it against the conceptual background of a revived *material turn* [51.39]. The possibility of reading artifacts or contexts like a text has increased dramatically due to modern natural science methodologies and procedures, and allows the reconstruction of the life history of an individual musical instrument [51.38]. Linking such material studies with spatial contexts creates fertile archaeoacoustic studies [51.40]:

> *From a material culture perspective there are at least three key areas to which archaeology can contribute: the instrument and other accoutrements on which the music is made; the place the music is made, performed or remembered; and the place that inspired that music to be made or which is recalled or described in song.*

For example, the resonant characteristics of Paleolithic caves containing wall paintings are being examined from this perspective [51.16].

Musical instruments are reconstructed in the context of experimental archaeology [51.41–44]. The insights gained in the process about the production and handling of the instruments can be used to interpret organological details and playing techniques. For example, in the case of wooden objects the design of a workpiece can be influenced by the particular tools that are available for chopping, sawing, carving, and finishing, by the properties of the wood (hardness, grain), and finally by the skills of the craftsperson. In practice, different goals may prevail that influence how the copy of an ancient musical instrument is made [51.45]. For example, there are copies produced for museums that look like the original in all details, but cannot necessarily be played. There are also archaeological music reconstructions, ideally based on reactivated ancient design and construction techniques, that can be used for experimental music making. There are also instruments produced with modern tools simply for the purpose of putting the operating principles to the test, and reconstructions primarily based on an interpretation of ancient images and the subjective aesthetic sensibilities of their makers [51.46]. The European Music Archaeology Project (2013–2018) intended to present in an exhibition reconstructions of musical instrument from European and neighboring regions [51.23].

51.3 Musical Practice

In contrast to ethnomusicology, which is concerned with the live music of recent societies and documented with the help of recording equipment, music archaeology has at best only isolated sounds with which to work, generated by either well-preserved or reconstructed musical instruments. The oldest unmistakable and complex musical instruments are over 40 000 years old and were found during systematic excavations in the Swabian Jura [51.47]. They are aerophones with up to five finger holes and made of bird bones (swan, vulture) and mammoth tusk. The latter material especially requires great technical skill to form. The discovery site of these instruments testifies to an early Upper Paleolithic context (Aurignacien) undoubtedly connected with modern humans (*Homo sapiens sapiens*). Artifacts found in a Neanderthal context that could have been musical instruments are ambiguous [51.48].

Taking into account paleoanthropological, neurophysiological and ethnological insights about Upper Paleolithic humans and hunter-gatherer societies, it becomes evident that early aerophones were components of a well-developed cultural communication system and served to strengthen the social cohesion of those societies. Images of hand-in-hand human figures, in some cases stylistically abstract and reaching back to earliest Neolithic times, can be interpreted as round dancers [51.49]. Round dances are likewise effective cultural techniques for increasing group cohesion and alliance formation [51.50].

Besides aerophones, several other instruments had been developed all over the world at the time, such as rattles, drums, bullroarers, pipes, trumpets and perhaps lithophones. However, lasting enrichment of the musical instrument inventory took place in the Old World with the development of complex stringed instruments (harps, lyres, lutes), which can be traced back to the second half of the fourth millennium BCE. It is striking that the so far earliest stringed instruments (bowed harps) arose at a time characterized by the emergence of early city-states in Mesopotamia. Harps and lyres were used on official occasions and at banquets or in the context of cultic rituals. Already in antiquity, they were a subject of music-theoretical discourse (e.g., tuning instructions). There is later evidence for spike bowl lutes

from the late third millennium BCE that frequently also occur in everyday scenes. They were first played by the singers of epic poetry in a socially elite environment, and in the second and first millennia BCE appear more frequently in the hands of socially less privileged people, such as warriors and entertainers, and in rare cases in the hands of a copulating couple [51.51].

An analysis of iconographic sources permits gender-relevant studies according to which ancient Western Asian lutes were depicted exclusively in male hands up until Hellenistic times. Along with other instruments like the box lyre, this instrument was brought to Egypt and Asia Minor in the 17th and 16th centuries BCE, where it was not exclusively played by men, but primarily by women (frequently lightly dressed dancers and women engaged in sex) [51.52–54]. Burial goods also supply evidence for this practice. It is noteworthy that lutes were not depicted in pre-Hellenistic Greece, suggesting that they were not highly valued or used. Following the line of these observations, the cultural transition characterized by Hellenistic culture could have led to a change in musical practice. Only in the extensively Hellenized world ranging from North Africa to the Middle East does the lute seem to have overcome the social barriers of prior times and neighboring regions. The question, *If the culture changed did the music change as well?* [51.28, p. 121], has not yet been systematically addressed from an archaeological music perspective. Another example related to this problem comes from Chinese music archaeology. According to Bo Lawergren, *the sliding playing technique so characteristic for modern* qin *playing was unavailable before 150 BCE*. At that time *only open strings could have been played*, which provided steady pitches [51.55, p. 297].

> *Both playing techniques (open strings and sliding technique) can be regarded as the expression of different musical styles at least: The old style may have forced the musician to a strict musical performance, while the later style may have allowed more freedom in the reproduction of the repertoire. If we follow Lawergren, the change of playing techniques took place in the early Han dynasty (206 BCE–220 CE), which is characterised by political conceptions different from those of their predecessors.* [51.20, p 11]

The elites of advanced ancient civilizations went to immense effort to produce sounds. There are examples from Mesopotamia (Ur), where already in the third millennium BCE female musicians and their instruments, including a lyre with a bull's head application of pure gold, were interred as *burial objects* to join a royal woman in her tomb. In second and first millennia BCE China, large carillon arrays composed of bronze bells in various dimensions were used to furnish royal graves. During the same period in Scandinavia, cast bronze lurs, typically played in pairs and used ritually, were produced of a quality unachieved today [51.56]. Huge numbers of resources and great expertise were invested in the creation of objects that produced sounds.

In the ancient Near East music was used to communicate with the gods, from whom one expected a good future. King Hammurabi (18th century BCE) accordingly employed musicians as part of successful political management in the sense of influencing the future for the better [51.57]. Music in ancient Greece was associated with cultic practices, similar to the situation in the ancient Near East and Egypt. Musical harmony was valued as the acoustic expression of an ideal state (Plato, *Politeia*). There was a similar understanding in ancient China, where the success of a state was associated with musical traditions (Lü Buwei, Spring and Autumn, third century BCE). By contrast, the eastern *barbarians* were unfamiliar with music, according to the Romans. The sounds they produced recalled something more like noise or a din. At that time a lack of music was a sign of inferiority [51.58]. Today, it is self-evident that this depends on the cultural point of view.

51.4 Music Theory

In the early written records of Western Asia and Egypt, especially from the third to first millennia BCE, there are numerous documents that provide very precise information about the use of musical instruments [51.59], the training of professional musicians, their musical repertoire and their social standing [51.57, 60–63]. Descriptions of rituals make it possible to draw conclusions about musical performance in practice and the role of musicians in cultic ceremonies [51.64]. Comparable sources from the 16th century CE report on the musical practices of indigenous populations of the Americas [51.18, 65].

In the ancient Near East archaeological music information sheds light on the fundamentals of music theory and tonal structures, which permits reconstruction of a heptatonic diatonic tonal system.

> *Konrad Volk wrote that Mesopotamians probably anticipated the ancient Greek heptatonic diatonic modal system. It is conjectural, however, to*

> *say how precisely these scales and modes were built. According to Stefan Hagel, the fine tuning of scales was based on a just intonation (not on a Pythagorean intonation or on tempered tuning). He does not agree with a major influence of music theory from the Near-East to Greece:* A comparison with ancient Greek music suggests a largely independent development of musical form from at least as early as the first half of the second millennium. *Leon Crickmore, however, does not exclude a possible connection:* The evidence cited from archaeology, musicology and the history of mathematics indicates the likelihood of the existence of a musical and mathematical tradition [...] lasting at least from nineteenth century Mesopotamia until fourth century Greece. Recent interpretations of the cuneiform tablets encouraged by the International Conference of Near Eastern Archaeomusicology (ICONEA) have led to alternative explanations of ancient Mesopotamian heptatonic scales and their fine tuning. [51.66, pp. 32], [51.67–72]

Although no such detailed written information on music theory exists from pre-Roman Egypt [51.25], excavations have brought to light numerous musical instruments, some of them in excellent condition [51.5, 7, 73, 74]. Instruments that convey an idea of the intervals that can be produced using the tone reservoir are of particular interest. This is the case for the *Coptic lutes* (a kind of *pandura*) of late antiquity, which were carved out from a single block of wood and originally equipped with wooden semifrets that were either glued on or fitted into a groove [51.38]. The instruments had three strings, whereby two strings formed a chord (probably tuned in unison), so only two string units were available for playing. Each of these string units had its own fret scale, variously divided by the frets. Depending on how the bridge position is reconstructed, precise intervals can be determined. The instruments were clearly designed for a limited range of modes. Impressions caused by string-wrapped full frets have been preserved on spike bowl lutes from Pharaonic times dating from the mid-second millennium BCE.

Studies of the fret series for these lutes and of the tone reservoir for end-blown flutes from the same cultural context permits reconstruction of heptatonic scales without semitones that build on a pentatonic scale. Three-quarter tones (circa 150 cents) and *neutral* thirds (circa 350 cents) are a characteristic of these scales, clearly distinguishing them from those of ancient Western Asia represented by Mesopotamian cuneiform sources of the second and first millennium BCE [51.66, 75–77]. However, traces of corrections made to the fret position between *neutral* thirds and minor thirds confirm that it was possible to alternate between different scale types in Egyptian musical performance. Such findings on musical instruments have not yet been systematically and exhaustively analyzed from a musicological perspective.

Whereas surviving records about ancient Mesopotamian music represent a theoretical foundation not yet confirmed by preserved musical instruments, ancient Egyptian musical instruments do reflect what is known of musical practice [51.77, p. 60]. This does not necessarily mean that the two advanced civilizations favored different types of scales, since many musicians were deported from Western Asia to Egypt in the second millennium BCE. If they not only brought with them Western Asian musical instruments, which first occurred in Egypt in the 17th and 16th centuries BCE, but also their own Western Asian musical repertoire, then it is possible that the fundamentals of music theory in both areas were quite similar. Further research and new discoveries of musical instruments from Western Asia are needed to clarify the matter.

Other studies on the tone reservoirs of different musical instruments concern the music cultures of other regions of the world, especially China and Central America (see, e.g., [51.78] (lithophones, gongs), [51.79] (ocarinas), [51.80] (aulos), [51.81] (end-blown flutes), and [51.82] (panpipes)). Using end-blown flutes, ocarinas, panpipes, lithophones and gongs, it was possible to ascertain interval sizes, scales, standardized tunings and tuning fluctuations that depend on how the instrument is played. Flutes from Neolithic cemeteries in Jiahu (China) allow inferences to be made, for example, about an increase in the tone reservoir in the course of settlement history (seventh and sixth millennia BCE) and an evolution from pentatonic to heptatonic scales [51.81].

51.5 Ancient Sounds

The music once produced with instruments found in an archaeological context cannot be recreated. Instructions for performing musical compositions (lyrics and instrumental accompaniment) seem to have appeared for the first time in the second half of the second millennium BCE in the locally limited region of Ugarit (Syria) [51.59, 83–85]. Some 60 notation fragments from the Greek/Roman cultural area have been dated to

the period between the third century BCE and the third and fourth centuries CE. They comprise information about instrumental music, lyrics and rhythm [51.86]. Although such records convey an idea of the melodic lines, other musical characteristics like pace, accent and dynamics are unknown. Because of the nature of its source material music archaeology can hardly deal with the structure of the music forms of ancient societies. The point of departure for research is therefore in most cases related to the tone reservoir, scale types and tonal character of the musical instruments, as well as the sociocultural function of music and musicians, whereby the actors and social relevance of musical performance are of primary interest.

One field of music archaeology involves working closely with instrument makers to reconstruct instruments and their tone reservoir in order to play them experimentally [51.87, 88]. In such cases the main goal is not related to an attempt to produce ancient music, which would be an impossible task. It is rather to demonstrate the sound potentials of musical instruments and to present new compositions and improvisations that make use of the ancient tone reservoir and the original conception of how an instrument should sound. To do this, a wide-based research approach is required that thoroughly investigates the acoustic, organological and technical framework. In the case of aerophones, both different blowing techniques and different fingering techniques have to be considered. Accordingly, the function of aerophones reclaimed from Upper Paleolithic contexts is currently a matter of controversy. It has not yet been clarified whether, to use modern terminology, we are dealing with flutes, oboes, clarinets or trumpets [51.48, 89, 90]. It is also not known whether Irish bronze horns and southern Scandinavian lurs were played with the help of circular breathing. There is currently no reliable basis for reconstructing playing techniques [51.56]. In the case of string instruments, what can be produced as a sound reconstruction depends on the tuning and characteristics of the strings, and these have, with rare exceptions, not been preserved. In general, it can be said that the sound volume of ancient string instruments is noticeably lower compared with modern acoustic instruments. Such instruments were probably not primarily used for instrumental performance but rather to accompany vocal music. This is undoubtedly a fruitful realm for future archaeoacoustic research. These few examples illustrate that any conclusions drawn in the field of experimental music archaeology require verifiable justification and documentation.

Reconstructing instruments and recording experimental music with the assistance of music archaeologists are not only suitable didactic methods for teaching and for educational programs in museums; they also have a place in living-history scenes in the entertainment media [51.87, pp. 295–296]. This is obviously an extremely speculative pursuit, but on the other hand, involvement with instruments and notation fragments from the distant past can spur the creativity of everyone participating in the process.

51.6 Conclusion

To take stock of the current situation, music archaeology is today an increasingly multidisciplinary and interdisciplinary field of research, so far primarily concerned with processing archaeological relics related to music and typological classification, and with reconstructing and explaining how sound-generating objects function. Archaeological music studies that pay more attention to the cultural context have become more common in the past two decades, thanks to different cultural turns (*spatial and material turn*). Nevertheless, there continue to be studies that extract music-related sources from their cultural context and look at them in isolation. This can lead to clarification of numerous detail problems, but frequently leaves untouched significant sociocultural changes in ancient music history. There is great potential in this area for future music archaeological research.

Acknowledgments. The comments of Adje Both and Lars-Christian Koch helped to improve previous German versions of this Chapter. Susan Giegerich translated the final version into English. Philip V. Bohlman reviewed the final English version for musicological terminology. However, all responsibility remains with the author.

References

51.1 E. Hickmann: Music archaeology – An introduction. In: *Saiteninstrumente im archäologischen Kontext. Stringed Instruments in Archaeological Context, Studien zur Musikarchäologie I*, Orient-Archäologie, Vol. 6, ed. by E. Hickmann, R. Eichmann (Marie Leidorf, Rahden/Westf. 2000) pp. 1–4

51.2 A.A. Both: Music Archaeology: Some methodological and theoretical considerations, Yearb. Tradit. Music **41**, 1–11 (2009)

51.3 C. Renfrew, P.G. Bahn: *Archaeology: Theory, Methods and Practice* (Thamaes and Hudson, London 2016)

51.4 C. Engel: *The Music of the Most Ancient Nations; Particularly of the Assyrians, Egyptians, and Hebrews* (John Murray, London 1864), repr. Berlin (2014)

51.5 C. Sachs: *Die Musikinstrumente des alten Ägyptens (Staatliche Museen zu Berlin, Mitteilungen aus der Ägyptischen Sammlung 3)* (Curtius, Berlin 1921)

51.6 C. Sachs: *The Rise of Music in the Ancient World. East and West* (Norton, New York 1943)

51.7 H. Hickmann: *Catalogue genéral des antiquités égyptiennes du Museé du Caire Nos. 69201–69852* (Imprimerie de l'Institut Français d'Archaeology Orientale, Kairo 1949)

51.8 O. Seewald: *Beiträge zu Kenntnis der steinzeitlichen Musikinstrumente Europas* (Schroll, Wien 1934)

51.9 H. Besseler, M. Schneider: Vorwort. In: *Ägypten. Musikgeschichte in Bildern II: Musik des Altertums* (Deutscher Verlag für Musik, Leipzig 1961) p. 5

51.10 E. Hickmann: An Introduction. In: *Challenges and New Objectives for Music Archaeology, Studien zur Musikarchäologie VI*, Orient-Archäologie, Vol. 22, ed. by A.A. Both, R. Eichmann, E. Hickmann, L.-C. Koch (2008) pp. XV–XXV

51.11 E. Hickmann: Archaeomusicology. In: *The New Grove Dictionary of Music and Musicians*, Vol. 1 (Macmillan, London 2000) pp. 848–854

51.12 E. Hickmann: s.v. Musikarchäologie. I. Begriffsbestimmung, Aufgabenfeld. In: *Musik in Geschichte und Gegenwart 6*, ed. by F. Blume, L. Finscher (Bärenreiter, Kassel 1997) pp. 929–935

51.13 Z. Estreicher: Ein Versuch der Musikarchäologie, Schweiz. Musikztg. **88**, 348–355 (1948)

51.14 R. Till: Sound archaeology. An interdisciplinary perspective. In: *Archaeoacoustics: The Archaeology of Sound: Publication of Proceedings from the 2014 Conference in Malta*, ed. by L.C. Eneix (The OTS Foundation, Myakka City 2014) pp. 23–32

51.15 C. Scarre, G. Lawson (Eds.): *Archaeoacoustics* (McDonald Institute for Archaeological Research, Cambridge 2006)

51.16 L.C. Eneix (Ed.): *Archaeoacoustics: The Archaeology of Sound: Publication of Proceedings from the 2014 Conference in Malta* (The OTS Foundation, Myakka City 2014)

51.17 C. Berends: *Klänge der Bronzezeit. Musikarchäologische Studien über bronzezeitliche Hortfunde Mitteldeutschlands*, Universitätsforschungen zur Prähistorischen Archäologie, Vol. 187 (Dr. Rudolf Habelt, Bonn 2010)

51.18 A. Both: Versunkene Klangwelten Altamerikas. Der Natur eine Stimme verleihen. In: *Musikarchäologie. Klänge der Vergangenheit*, Archäologie in Deutschland Sonderheft 7, ed. by R. Eichmann, L.-C. Koch (Wissenschaftliche Buchgesellschaft, Darmstadt 2015) pp. 51–61

51.19 M. Stöckli, A. Both (Eds.): *Flower World. Music Archaeology of the Americas/Mundo Florido*, Arqueomusicología de las Américas, Vol. 1 (Ēkhō, Berlin 2012)

51.20 R. Eichmann: Musical perceptions – past and present. On ethnographic analogy in music archaeology – an introduction. In: *Studien zur Musikarchäologie VII*, Orient-Archäologie, Vol. 25, ed. by R. Eichmann, E. Hickmann, L.-C. Koch (Marie Leidorf, Rahden/Westf. 2010) pp. XII–XIV

51.21 International Study Group on Music Archaeology: http://www.musicarchaeology.org

51.22 International Council for Traditional Music, Study Group on Music Archaeology: http://www.ictmusic.org/group/music-archaeology

51.23 European Music Archaeology Project: http://www.emaproject.eu

51.24 Klankbord: *Newsletter for Ancient and Medieval Music*, http://www.klankbordsite.nl/

51.25 A. von Lieven: Music archaeology – Music philology. Sources on Egyptian music and their inherent problems. In: *Musikarchäologische Quellengruppen: Bodenurkunden, mündliche Überlieferung, Aufzeichnung/Music Archaeological Sources: Finds, Oral Transmission, Written Evidence. Studien zur Musikarchäologie IV*, Orient-Archäologie, Vol. 15, ed. by E. Hickmann, R. Eichmann (Marie Leidorf, Rahden/Westf. 2004) pp. 99–105

51.26 G. Kolltveit: *Jew's Harps in European Archaeology* (Unipub AS, Oslo 2004)

51.27 Å. Norborg: *Ancient Middle Eastern Lyres*, Musikmuseets skrifter, Vol. 25 (Almqvist Wiksell, Stockholm 1995)

51.28 B. Nettl: Some questions on the relationship of music archaeology and ethnomusicology: Informal comments on constructing the past from the present. In: *Musikarchäologische Quellengruppen: Bodenurkunden, mündliche Überlieferung, Aufzeichnung/Music-Archaeological Sources: Finds, Oral Transmission, Written Evidence. Studien zur Musikarchäologie IV*, Orient-Archäologie, Vol. 15, ed. by E. Hickmann, R. Eichmann (Marie Leidorf, Rahden 2004) pp. 117–124

51.29 S. Psaroudakes: Archaeomusicology and ethnomusicology in dialogue, Eulimene **4**, 189–197 (2003)

51.30 E.M.V. Hornbostel, C. Sachs: Systematik der Musikinstrumente. Ein Versuch, Z. Ethnol. **46**, 553–590 (1914), English edition: Classification of musical instruments. Galpin Soc. J. **14**, 3–29 (1961)

51.31 D.A. Olsen: The ethnomusicology of archaeology: A model for the musical/cultural study of ancient material culture, Sel. Rep. Ethnomusicol. **8**, 175–192 (1990)

51.32 L.-C. Koch: Klang und Kultur. Musikethologische Erkenntnisse als Grundlage für musikarchäologisches Arbeiten. In: *Musikarchäologie. Klänge der Vergangenheit*, Archäologie in Deutschland Sonderheft 7, ed. by R. Eichmann, L.-C. Koch (Wissenschaftliche Buchgesellschaft, Darmstadt 2015) pp. 12–21

51.33 R. Eichmann, L.-C. Koch, D.-R. Meyer: A preliminary archaeological report on two lutes of the Surbahar type (India) and an ethnomusicological perspective. In: *Challenges and Objectives in Music Archaeology. Studien zur Musikarchäologie VI*, Orient-Archäologie, Vol. 22, ed. by A.A. Both, R. Eichmann, E. Hickmann, L.-C. Koch (Marie Leidorf, Rahden/Westf. 2008) pp. 547–564

51.34 P.V. Bohlman: Das Sammeln der Singenden. Über die Archäologie des Gesangswettbewerbs im Zeitalter des europäischen Nationalismus. In: *Musikarchäologie. Klänge der Vergangenheit*, Archäologie in Deutschland Sonderheft 7, ed. by R. Eichmann, L.-C. Koch (Wissenschaftliche Buchgesellschaft, Darmstadt 2015) pp. 85–90

51.35 M.N. Haidle, J. Bräuer: Special issue: Innovation and the evolution of human behavior from brainwave to tradition – How to detect innovations in tool behavior, PaleoAnthropology **2011**, 144–115 (2011)

51.36 F. D'Errico: Just a bone or a flute? The contribution of taphonomy and microscopy to the identification of prehistoric pseudo-musical instruments. In: *The Archaeology of Sound: Origin and Organisation. Studien zur Musikarchäologie III*, Orient-Archäologie, Vol. 10, ed. by E. Hickmann, A.D. Kilmer, R. Eichmann (Marie Leidorf, Rahden/Westf. 2002) pp. 89–90

51.37 G. Lawson, F. d'Errico: Microscopic, experimental and theoretical re-assesment of Upper Palaeolithic bird-bone pipes from Isturitz, France: Ergonomics of design, systems of notation and the origins of musical traditions. In: *The Archaeology of Sound: Origin and Organisation. Studien zur Musikarchäologie III*, Orient-Archäologie, Vol. 10, ed. by E. Hickmann, A.D. Kilmer, R. Eichmann (Marie Leidorf, Rahden/Westf. 2002) pp. 119–142

51.38 F. Calament, R. Eichmann, C. Vendries: *Le luth dans l'Égypte byzantine. La tombe de la „Prophetesse" d'Antinoë au musée de Grenoble*, Orient-Archäologie, Vol. 26 (Marie Leidorf, Rahden/Westf. 2012)

51.39 D. Bachmann-Medick: Cultural Turns, Version 1.0, In: *Docupedia-Zeitgeschichte*, Vol. 29, 3.2010 http://docupedia.de/zg/Cultural_Turns?oldid=84593

51.40 J. Schofield: The archaeology of sound and music, World Archaeol. **46**(3), 289–291 (2014)

51.41 Europäische Vereinigung zur Förderung der Experimentellen Archäologie e.V. (Ed.): *Experimentelle Archäologie in Europa Bilanz 2010* (Isensee, Oldenburg 2010)

51.42 D.C.E. Millson: *Experimentation and Interpretation: The Use of Experimental Archaeology in the Study of the Past* (Oxbow Books, Oxford 2010)

51.43 E. Keefer (Ed.): Lebendige Vergangenheit. Vom archäologischen Experiment zur Zeitreise, Sonderheft 6 Archäologie in Deutschland (2006)

51.44 J.M. Coles: *Archaeology by Experiment* (Charles Scribner's Sons, New York 1974)

51.45 G. Lawson: Epistemology and imagination. Reconciling music-archaeological scholarship and ancient music performance today. In: *Musical Perceptions – Past and Present. On Ethnographic Analogy in Music Archaeology. Studien zur Musikarchäologie VII*, Orient-Archäologie, Vol. 25, ed. by R. Eichmann, E. Hickmann, L.-C. Koch (Marie Leidorf, Rahden/Westf. 2010) pp. 265–276

51.46 M.P. Georgiou: *Ancient Greek Musical Instruments. The Quest for the Resonance of the Universe* (En Tipis, Nicosia 2007), (Greek/English)

51.47 N.J. Conard, M. Malina, S.C. Münzel: New flutes document the earliest musical tradition in southwestern Germany, Nature **460**, 737–740 (2009)

51.48 S.C. Münzel, W. Hein, F. Potengowski, N.J. Conard: Flötenklang aus fernen Zeiten. Die ältesten Blasinstrumente von der Schwäbischen Alb. In: *Musikarchäologie. Klänge der Vergangenheit*, Archäologie in Deutschland Sonderheft 7, ed. by R. Eichmann, L.-C. Koch (Wissenschaftliche Buchgesellschaft, Darmstadt 2015) pp. 30–37

51.49 Y. Garfinkel: *Dancing at the Dawn of Agriculture* (Univ. of Texas Press, Austin 2003)

51.50 I. Morley: *The Prehistory of Music. Human Evolution, Archaeology & the Origins of Musicality* (Oxford Univ. Press, Oxford 2013)

51.51 F. Blocher: Gaukler im Alten Orient. In: *Außenseiter und Randgruppen: Beiträge zu einer Sozialgeschichte des Alten Orients*, Xenia, Vol. 32, ed. by V. Haas (Univ.-Verl., Konstanz 1992) pp. 79–112

51.52 L. Manniche: *Music and Musicians in Ancient Egypt* (British Museum, London 1991) pp. 108–119

51.53 A. von Lieven: Wein, Weib und Gesang – Rituale für die Gefährliche Göttin. In: *Rituale in der Vorgeschichte, Antike und Gegenwart*, ed. by C. Metzner-Nebelsick (Marie Leidorf, Rahden/Westf. 2003) pp. 47–55

51.54 O. Brison: Nudity and music in anatolian mythological seduction scenes and iconographic imagery. In: *Music in Antiquity. The Near East and the Mediterranean*, ed. by J. Goodnick Westenholz, Y. Maurey, E. Seroussi (De Gruyter Oldenbourg, Berlin, Boston, Jerusalem 2014) pp. 185–200

51.55 B. Lawergren: How Qin-zithers changed between 500 B.C.E. and 500 C.E. In: *Music Archaeological Sources: Finds, Oral Transmission, Written Evidence. Studien zur Musikarchäologie IV*, Orient-Archäologie, Vol. 15, ed. by E. Hickmann, R. Eichmann (Marie Leidorf, Rahden/Westf. 2004) pp. 295–310

51.56 J. Schween: Luren und Irische Hörner der Bronzezeit. Nordeuropäische Meisterwerke der Klangerzeugung. In: *Musikarchäologie. Klänge der Vergangenheit*, Archäologie in Deutschland

Sonderheft 7, ed. by R. Eichmann, L.-C. Koch (Wissenschaftliche Buchgesellschaft, Darmstadt 2015) pp. 38–50
51.57 D. Shehata: *Musiker und ihr vokales Repertoire. Untersuchungen zu Inhalt und Liedgattung in altbabylonischer Zeit*, Göttinger Beiträge zum Alten Orient, Vol. 3 (Universitätsverlag, Göttingen 2009)
51.58 P. Amory: *People and Identity in Ostrogothic Italy 489–554* (Cambridge Univ. Press, Cambridge 1997) p. 133
51.59 K. Volk: Musikalische Praxis und Theorie im Alten Orient. In: *Vom Mythos zur Fachdisziplin: Antike und Byzanz*, Geschichte der Musiktheorie, Vol. 2, ed. by K. Volk, F. Zaminer, C. Floros, R. Harmon, L. Richter, M. Haas (Wissenschaftliche Buchgesellschaft, Darmstadt 2006) pp. 1–51
51.60 R. Pruzsinszky, D. Shehata (Eds.): *Musiker und Tradierung. Studien zur Rolle von Musikern bei der Verschriftlichung und Tradierung von literarischen Werken*, Wiener Offene Orientalistik, Vol. 8 (Lit, Wien, Berlin 2010)
51.61 N. Ziegler: Florilegium marianum 9. Les musiciens et la musique d'après les archives de Mari. In: *Mémoires de N.A.B.U. 10* (G.N. Impressions, Bouloc 2007)
51.62 K. Volk: Improvisationsmusik im alten Mesopotamien? In: *Improvisation II*, ed. by W. Fähndrich (Amadeus, Winterthur 1994) pp. 160–202
51.63 A. von Lieven: The social standing of musicians in ancient Egypt. In: *Music Archaeology in Contexts. Studien zur Musikarchäologie V*, Orient-Archäologie, Vol. 20, ed. by E. Hickmann, A.A. Both, R. Eichmann (Marie Leidorf, Rahden/Westf. 2006) pp. 355–360
51.64 M. Schuol: *Hethitische Kultmusik. Eine Untersuchung der Instrumental- und Vokalmusik anhand der hethitischen Ritualtexte und der archäologischen Zeugnisse*, Orient-Archäologie, Vol. 14 (Marie Leidorf, Rahden/Westf. 2004)
51.65 M. Howell: An Organology of the Americas as painted by John White and other artists. In: *Flower World. Music Archaeology of the Americas/Mundo Florido*, Arqueomusicología de las Américas, Vol. 1, ed. by M. Stöckli, A.A. Both (2012) pp. 155–168
51.66 R. Eichmann: Extant lutes from the New Kingdom and the Coptic Period of Ancient Egypt. In: *ICONEA 2011* (2014) pp. 23–35
51.67 L. Crickmore: New light on the Babylonian tonal system. In: *Proceedings of the International Conference of Near Eastern Archaeomusicology (ICONEA)*, ed. by R. Dumbrill, I. Finkel (2008) pp. 11–22
51.68 L. Crickmore: The tonal system of Mesopotamia and ancient Greece: Some similarities and differences, Archaeomusicol. Rev. Ancient Near East **1**, 1–6 (2009), http://sas.academia.edu/RichardDumbrill/Books/182868/ARANE_2009
51.69 R. Dumbrill: Evidence and inference in texts of theory in the ancient Near East. In: *Proceedings of the International Conference of Near Eastern Archaeomusicology (ICONEA)*, ed. by R. Dumbrill, I. Finkel (2008) pp. 105–115
51.70 J. Goodnick Westenholz, Y. Maurey, E. Seroussi (Eds.): *Music in Antiquity. The Near East and the Mediterranean* (De Gruyter Oldenbourg, Berlin, Boston, Jerusalem 2014)
51.71 S. Hagel: Is *nid qabli* Dorian? Tuning and modality in Greek and Hurrian music, Baghdader Mitt. **36**, 287–348 (2005)
51.72 D. Halperin: Musical reconstruction of the Hurrian material by statistical analysis. In: *Proc. Int. Conf. Near East. Archaeomusicol. (ICONEA)*, ed. by R. Dumbrill, I. Finkel (2008) pp. 29–32
51.73 C. Ziegler: *Les Instruments de Musique Égyptiens au Musée du Louvre* (Éditions de la Réunion des Musées Nationaux, Paris 1979)
51.74 R.D. Anderson: *Catalogue of Egyptian Antiquities in the British Museum III: Musical Instruments* (British Museum Publications, London 1976)
51.75 F. Calament, R. Eichmann, C. Vendries: *Le luth dans l'Égypte byzantine. La tombe de la «Prophetesse d'Antinoë» au musée de Grenoble*, Orient-Archäologie, Vol. 26 (Marie Leidorf, Rahden/Westf. 2011)
51.76 R. Eichmann: Neuaufnahme einer Schalen-Spießlaute von deir el-Medina (Grab 1389)/Ägypten. In: *Music-Archaeological Sources: Finds, Oral Transmission, Written Evidence. Studien zur Musikarchäologie IV*, Orient-Archäologie, Vol. 15, ed. by E. Hickmann, R. Eichmann (Marie Leidorf, Rahden/Westf. 2004) pp. 551–568
51.77 R. Eichmann: Structure des échelles musicales dans l'antiquité égyptienne tardive, une perspective d'après l'étude des "luths coptes", Archéologie et musique. Actes du colloque des 9 et 10 février 2001, Les cahiers du muse de la musique **2**, 55–61 (2002)
51.78 F. Jianjun: Scale, tuning, and spectrum. The early Han Dynasty Chime stones and bronze bells excavated in Luozhuang, China. In: *Musical Perceptions – Past and Present. On Ethnographic Analogy in Music Archaeology, Studien zur Musikarchäologie VII*, Orient-Archäologie, Vol. 25, ed. by R. Eichmann, E. Hickmann, L.-C. Koch (Marie Leidorf, Rahden/Westf. 2010) pp. 167–188
51.79 P.F. Healy, C.L. Dennett, M.H. Harris, A.A. Both: A musical nature. Pre-Columbian ceramic flutes of northeast Honduras. In: *Musical Perceptions – Past and Present. On Ethnographic Analogy in Music Archaeology. Studien zur Musikarchäologie VII*, Orient-Archäologie, Vol. 25, ed. by R. Eichmann, E. Hickmann, L.-C. Koch (Marie Leidorf, Rahden/Westf. 2010) pp. 189–212
51.80 S. Hagel: Understanding the Aulos. Berlin Egyptian Museum 12461/12462. In: *Musical Perceptions – Past and Present. On Ethnographic Analogy in Music Archaeology, Studien zur Musikarchäologie VII*, Orient-Archäologie, Vol. 25, ed. by R. Eichmann, E. Hickmann, L.-C. Koch (Marie Leidorf, Rahden/Westf. 2010) pp. 67–87
51.81 J. Zhang, X. Xiao, Y.K. Lee: The early development of music. Analysis of the Jiahu bone flutes, Antiquity **78**, 769–778 (2004)
51.82 A. Gruszczyńska-Ziółkowska: Sound and its numbers. Interpretation of acoustical data from the

Nasca Culture (Peru). In: *The Archaeology of Sound: Origin and Organisation. Studies in Music Archaeology III*, Orient-Archäologie, Vol. 10, ed. by E. Hickmann, A.D. Kilmer, R. Eichmann (Marie Leidorf, Rahden/Westf. 2002) pp. 269–278

51.83 A.D. Kilmer: Mesopotamian music theory since 1977. In: *Music in Antiquity. The Near East and the Mediterranean*, ed. by J. Goodnick Westenholz, Y. Maurey, E. Seroussi (De Gruyter Oldenbourg, Berlin, Boston, Jerusalem 2014) pp. 92–101

51.84 A.D. Kilmer, R.L. Crocker, R.R. Brown: *Sounds from Silence. Recent Discoveries in Ancient Near Eastern Music* (BĪt Enki, Berkeley 1976), record with booklet

51.85 A.D. Kilmer: The Cult Song with Music from Ancient Ugarit: Another Interpretation, Rev. Assyriol. **68**, 69–82 (1974)

51.86 E. Pöhlmann, M.L. West (Eds.): *Documents of Ancient Greek Music. The Extant Melodies and Fragments* (Clarendon, Oxford 2001)

51.87 S. Rühling: Ancient sounds for modern ears – Archaeomusicology in media and performance. Essay of an experience report. In: *Musical Perceptions – Past and Present. On Ethnographic Analogy in Music Archaeology. Studien zur Musikarchäologie VII*, Orient-Archäologie, Vol. 25, ed. by R. Eichmann, E. Hickmann, L.-C. Koch (Marie Leidorf, Rahden/Westf. 2010) pp. 293–299

51.88 S. Rühling: Ein offenes Ohr für die Vergangenheit. Rekonstruktion und experimentelles Spiel antiker und byzantinischer Orgeln, Archäologie in Deutschland, Sonderheft 7. Musikarchäologie (2014)

51.89 S. Wyatt: Sound production in Early Aerophones. Short report on a work in progress. In: *Studien zur Musikarchäologie 8. Klänge der Vergangenheit/Sounds from the Past*, Orient-Archäologie, Vol. 2, ed. by R. Eichmann, F. Jianjun, L.-C. Koch (Marie Leidorf, Rahden/Westf. 2012) pp. 393–398

51.90 J.-L. Ringot: Upper Palaeolithic aerophones – flute or pipe? An experimental approach. Summary report. In: *Klänge der Vergangenheit/Sounds from the Past Studien zur Musikarchäologie 8*, Orient-Archäologie, Vol. 27, ed. by R. Eichmann, F. Jianjun, L.-C. Koch (Marie Leidorf, Rahden/Westf. 2012) pp. 389–390

52. The Complex Dynamics of Improvisation

David Borgo

This essay provides some general observations about the field of improvisation studies and surveys important theoretical and empirical work on the subject. It makes a distinction between referent-based and referent-free musical improvisation, placing particular emphasis on the specific issues that surround the latter, and highlighting recent research that has arisen to address them. Whether referent-based or referent-free, improvisation appears to involve a continual tension between stabilization through communication and past experience and instability through fluctuations and surprise. While many issues persist about how to frame and explore musical improvisation, there is broad agreement that improvisation involves novel output (for the individual, but only optionally for society) created in nondeterministic, real-time situations by individuals and collectives involving certain affordances and constraints. The critical questions involve how we chose to frame these improvisatory dynamics: either as information processing struggling to keep pace with the cognitive demands of the moment, or as an ecologically sensitive engagement with one's sonic and social world.

52.1 The Study of Improvisation

The study of improvisation, paraphrasing *Howard Gardner*'s view of cognitive science, has a relatively short history but a very long past [52.1]. All instances of human music making, from the most ancient to the most avant-garde, arguably involve at least some degree of improvisation, if by this we mean making musical decisions in the course of performance. Until the advent of sound recording, however, musical improvisation was extremely difficult to document, and therefore to study systematically.

Oral teachings, and to a lesser extent written treatises on the subject, can be found in a variety of musical traditions: from Western classical music (at least until the time of Beethoven), to folk and art music traditions of the non-Western world (especially in south, west and southeast Asia, but also throughout much of Africa and Latin America), to more contemporary popular music styles, with jazz serving as an important locus of activity. The thrust of these treatments, however, tends to be either prescriptive or anecdotal. General comparisons between various improvising traditions were sometimes made, but relatively little scholarship attempted to offer a more systematic or synthetic view of improvisation, with [52.2] being an important exception.

Audio recording provided researchers with the ability to capture and rehear a performance, allowing for microanalysis of sonic details and for comparative studies, often either involving a single musician improvising on different occasions [52.3] or different individuals improvising on the same underlying musical structure [52.4]. This methodology allowed for more nuanced descriptions of musical improvisation, but the inherent complexities of the transcription and analysis process, whether done by a human or by a machine, are nontrivial. *Peter Winkler* in [52.5] highlights many of the insurmountable challenges inherent to any attempt to reduce music as sound to music as notation. He likens a transcription to a blueprint drawn after the building is built, and cautions us not to mistake the blueprint for the building.

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_52

Beginning around the 1960s, the growth of both jazz studies and ethnomusicology as academic fields precipitated an increase in research on musical improvisation, with some of the most influential texts also emerging from nonmusic-related fields [52.6–8]. At roughly the same time, a musical practice specifically described as *improvised music* or *free improvisation* emerged that borrows from a panoply of musical approaches and at times seems unencumbered by any overt idiomatic constraints. This practice was originally championed by an eclectic group of primarily American and European artists with backgrounds in modern jazz and contemporary classical music. It now involves musicians emanating from around the world and draws on an even wider spectrum of influences, including electronic, experimental, and intermedia arts. This essay will provide some general observations about the field of improvisation studies and survey important theoretical and empirical work on the subject. It makes a distinction between referent-based and referent-free musical improvisation, placing particular emphasis on the specific issues that surround the latter, and highlights recent research that has arisen to address them. The scope of the research being surveyed is primarily limited to work published in English.

52.2 The Field of Improvisation Studies

In [52.9], *Bruno Nettl* argues that improvisation is *an art neglected in scholarship*. In [52.10] *Derek Bailey* observes that improvisation is both the most widely practiced of all musical activities and the least acknowledged and understood. Despite this apparent neglect, the academic field of improvisation studies has grown considerably in recent decades, with the publication of numerous scholarly books and articles on the subject, and with the emergence of academic journals, conferences, and graduate programs with improvisation as a central focus. The appearance of the journal *Critical Studies in Improvisation/Études critiques en improvisation* in 2004, the formation of the International Society for Improvised Music in 2006, and the forthcoming publication of the two-volume *Oxford Handbook of Critical Improvisation Studies* are all watershed marks for the field's increasing prominence.

While improvisation has often been studied in discipline-specific ways (e.g., in music, theater, dance, or visual arts), only recently have researchers embarked on more multi- and interdisciplinary work, at times spurred on by developments in the cognitive and neurological sciences, or by an interest in understanding improvisation across experience. Improvisation studies as an academic field tends to draw on many of the same theories that influence other scholarly work in the arts and humanities, such as critical theory, cultural studies, and science and technology studies. Researchers in the field have become more politically and socially engaged on the whole. Work on improvisation now extends well beyond the arts into fields such as education, philosophy, sociology, anthropology, literature, law, postcolonial studies, gender studies, human-computer relations, sports, and medicine, to name only a few.

Improvisation has also become a hot topic in management studies and organizational design, along with other business-related fields. Two special journal issues dedicated to this topic are [52.11, 12]. The former offers primarily a favorable assessment of employing the *jazz metaphor* to understand and generate creativity in the business realm, while the latter provides a more critical view of the underlying ethics and profit-driven motivations of attempting to *aestheticize* neoliberal economics.

What exactly is meant by the term *improvisation* across this variety of work can be remarkably diverse, and at times frustratingly vague. Certainly improvisation involves numerous different types of creativity (another term that defies easy definition). In the arts alone, whether it is theater, dance, comedy, painting or music, improvisation undoubtedly draws on different sets of abilities and experiences, and it offers different demands and rewards. Therefore, the metaphorical comparisons frequently made to improvising music from within other fields and pursuits can illuminate, but also obfuscate.

52.3 Challenges in Defining Improvisation

Definitions of musical improvisation tend to be vague, overgeneralized, or beholden to conventional notions of musical practice. For example, improvisation is often described as composing music on the spur of the moment, or as performing music spontaneously without the aid of manuscript, sketches, or memory. These

types of definitions tend to downplay the extensive practice and experience that seasoned improvisers bring to performance, and the ways in which memory (both declarative and procedural) and often some form of notation (perhaps functioning as an *aide memoire*) are still involved when one learns how to improvise.

Improvisation, according to *Bruce Benson* [52.13], does not fit the binary opposition of composition and performance that underpins how we tend to think about music making. He insists that we view improvisation along a continuum of musical practices running through various gradations of interpretive performance towards a type of stream-of-consciousness playing. Improvisation, however, can also be understood as qualitatively different from conventional notions of composition and performance, involving very different cognitive processes and reflecting different aesthetics, even a different ontology. Might improvisation challenge traditional notions that music necessarily involves a linear process leading from *compositional* activity to *performance* activity? Might it even get us to rethink conventional ideas about cognition that subscribe to a linear progression from sensation to thinking to action?

Jeff Pressing, whose work on the cognitive modeling of musical improvisation was some of the earliest and most influential on the subject, defines improvisation as *the simultaneous design and execution of musical ideas* [52.14], a formulation that seems both to call on and exceed the conventions of composition and performance. Writing in philosophy and aesthetics about musical improvisation is limited, but often argues that improvisation and composition are driven by unique mechanisms and expressive goals and therefore should not be assessed uniformly. In general, these treatments highlight the cumulative conception and irreversible temporality of improvisation; one can make reference to what has already occurred, but any *editing* must happen retrospectively.

Ed Sarath, for instance, asserts in [52.15] that improvisers experience time in an inner-directed manner in which the present is heightened and the past and future are perceptually subordinated. These observations underscore his conviction that improvisation demands a distinct *aesthetics of spontaneity*. *Ted Gioia* in [52.16] formulates an *aesthetics of imperfection* that cautions against making judgments about musical improvisation using the same formalist criteria that we use to judge notated composition. *Philip Alperson* in [52.17] frames his arguments through an *aesthetics of action*, insisting that, while improvisation affords access to the *composer's mind at the moment of creation*, it should not be viewed as a performative token of a compositional *megatype* or model.

In general, these different aesthetics may be explainable through differences between online and offline cognition, especially when viewed from both the vantage point of the individual and from a social cognitive perspective. Online cognition is concerned with *immediate input* from our local environment and is often used to describe an interactor's point of view, whereas offline cognition involves more *careful considerations*, such as lengthy editorial decisions or future planning, and it frequently describes an observer's – rather than an interactor's – point of view. Improvisation has certainly been disparaged in settings that place a greater value on offline cognition and/or that view an activity's intrinsic value from a *universalizing* perspective of observation over interaction.

My personal favorite illustration of the oft-noted differences between composition and improvisation comes from a chance meeting between Frederic Rzewski and Steve Lacy. In this frequently recounted tale, Rzewski asked Lacy to describe in 15 s the difference between composition and improvisation. Lacy replied [52.10, p. 141]:

> *In fifteen seconds, the difference between composition and improvisation is that in composition you have all the time you want to decide what to say in fifteen seconds, while in improvisation you have fifteen seconds.*

Lacy's formulation of the answer, according to the story, lasted exactly 15 s.

The pithiness of this statement and the timeliness of its delivery belies its profundity. On the one hand, Lacy improvised his verbal response to a question about improvisation given the constraints of time and context. If he had been provided a full minute, or 10 min, or 2 h for his response, or if he was speaking to a different audience, or perhaps in French, he would, undoubtedly, have improvised a different answer. In this way, his particular response is both in-the-moment and context-dependent; it is spontaneous, yet also made to conform to the expectations and demands of the moment.

Moreover, Lacy had likely given some previous thought to the general issue, and perhaps he had even offered similar albeit differently phrased responses on the subject in the past. So, in what ways was his statement improvised, worked out, or some combination of these? Would his remark have been less effective if he had not delivered it within the provided 15 s window of time? Should we view his response as *spontaneous*, or is it better viewed as *fluent*, in the same way that fluent speakers do not necessarily intend the construction of their sentences ahead of time? What if his response had been musical rather than linguistic? How would we judge its spontaneity or fluency then?

A priori definitions of creativity are certainly ill advised, but a posteriori definitions are likely unavoidable. *P.N. Johnson-Laird* [52.18] provides a helpful acronym for exploring creative activity, NONCE, according to which the outcome of a creative process is novel (N) for the person producing the result but only optionally novel (O) for society at large, and it is the result of a nondeterministic process (N) that is guided by constraints (C) and is based on existing elements (E).

52.4 Some Contemporary Research Directions

On the whole, researchers in the fields of music psychology and music cognition have focused the majority of their attention on how listeners perceive and process music. These approaches tend to use recorded or computer-generated music examples for their consistency. When researchers have explored the complexities of music performance, they most often do so by looking at the performance of notated music, again for the experimental control it provides. Studying in vivo improvisation from the perspective of the performer or listener remains a challenging proposition, and most writing on the subject remains theoretical, with much of it emanating from practicing musicians or from academics who also improvise.

The most widely accepted approaches to the study of improvisation in cognitive psychology theorize that improvisers deal with the *cognitive constraints* of the moment by drawing on a *model* [52.19], a *blueprint* or *skeleton* [52.20], or a *referent* [52.14] stored in long-term memory. Depending on the specific musical tradition, the relationship between a given improvisation and its model or referent may be more or less fixed, with free improvisation perhaps providing something of a boundary case for this approach, since many of its practitioners disavow the idea that the music is based on a model.

While it is undoubtedly true that real-time creative processes such as improvisation place great demands on cognitive resources, the brain is only one node in a complex nonlinear feedback system. Notably fewer authors, however, have approached the topic from the vantage point of social or ecological psychology [52.21, 22]. In other words, the question of *how does one improvise?* tends to elicit cognitivist or computational models concerned with individual *information processing* and *signal generation* instead of ecological inquiries that explore how one's perception, action, and sonic and social worlds may be intertwined or entangled. I will have more to say about this topic at the end of the essay.

Work on musical improvisation from within the field of ethnomusicology has corrected some of this imbalance. For instance, *Paul Berliner*'s [52.23] and *Ingrid Monson*'s [52.24] influential work on jazz improvisation highlights its dialogical nature and the culturally contingent ways that it is learned and conceptualized. *Vijay Iyer*'s work [52.25], as well as my own [52.26], also demonstrates how contemporary views of cognition as inherently embodied, situated, and distributed can shed further light on how and why musicians improvise together.

Theoretical writing on improvisation still outpaces empirical work on the subject, although this balance is beginning to shift, as more researchers analyze data drawn from, for example, video and other performance capture methods [52.27] or galvanic skin response [52.28]. By using brain imaging techniques [52.29–31] we are also beginning to identify specific neural regions that may be involved in heightened moments of improvisational creativity – sometimes referred to as a *flow* state [52.32, 33] – but this research is still in its earliest stages and questions abound with regards to how to structure experiments and how to interpret the data meaningfully.

Other recent approaches have used a variety of interview tactics [52.34] or grouping tasks [52.35] to understand the kinds of choices that improvisers either made in performance or might likely make given certain stimuli. Here again, challenges remain regarding how to frame tasks for participants and to employ measuring and recording technologies in ecologically sensitive ways that do not unduly impede or prefigure improvisational activities, as well as how to assess and account for differences in improvisational skill level.

Some recent work has shifted focus from primarily investigating performer note choice within harmonic contexts, akin to how jazz is often taught using chord-scale relationships, towards studies of intensity and timing as key factors in how musicians improvise together, and of how listeners hear and interpret improvised performances [52.36, 37]. The importance of timbre and nonlinear organizing principles in contemporary improvised music remains understudied in the literature, although fractal dimensional analysis of recorded improvisations has produced compelling results [52.26, Chap. 5 with Rolf Bader]. Another avenue of contemporary research involves developing automatic improvising systems with the intention of creating competent or even expert performances as judged by knowledge-

able human listeners [52.38]. Some of these approaches have turned to robotics as well in order to explore the important role that physical gestures often play in facilitating group interaction and in engaging listeners [52.39].

While many issues persist about how to frame and explore musical improvisation, there is broad agreement that improvisation involves novel output (again, for the individual, but only optionally for society) created in nondeterministic, real-time situations by individuals and collectives involving certain affordances and constraints. The following two sections will attempt to clarify these ideas within the domain of referent-based and referent-free musical improvisation.

52.5 Referent-Based Improvisation

Jeff Pressing, an improvising pianist himself, was one of the first psychologists to put forward a convincing cognitive model of improvisation, and he continued to refine his work on the subject until his untimely death. In brief, his model in [52.40] characterizes improvisation as a heterarchical process involving motor, psychological, and cultural aspects (which differentiates it from other models that emphasize only rule-based procedures, such as *Johnson-Laird*'s [52.18]; see [52.41] for an analysis of these competing theories).

It is first important to make a distinction between the various timescales involved in improvisation. For the purposes of this discussion, I will divide them into *short-term*, *intermediate*, and *long-term*. On the short-term scale of a second or less, Pressing argued that embodied patterns and learned gestures of procedural memory constitute the bulk of the grist for the improviser's mill. Improvising at this timescale, according to him, involves a nonlinear cycle of motoric function and cognitive feedback and feedforward. In reviewing the physiological and neurological literature, Pressing concluded that improvisers have the biological capacity to react to unexpected changes, and hence to one's own or one another's new ideas, about twice a second. He argued that the performance of experienced improvisers on short-term timescales can be quite nuanced and flexible, but it is also largely automatized.

On an intermediate timescale, lasting from a few seconds to the length of a full performance, Pressing argued that most improvised music utilizes an underlying scheme or formal image to facilitate the generation and editing of improvised behavior. Pressing called this a referent. His most succinct description of a referent was [52.40, p. 52]:

> *a set of cognitive, perceptual, or emotional structures (constraints) that guide and aid in the production of musical material.*

In other words, having a referent allows improvisers the possibility to prepare before the performance and a means during the performance to anticipate future developments. It provides *material* for variation or development, and since a referent can be shared, it reduces the need for detailed attention to all of the component parts of a given performance.

In mainstream jazz, for instance, a *standard* song form can provide a shared template of melodic, rhythmic, and harmonic information, and a means by which performers can synchronize or orient their performances with respect to those of the other musicians. But Pressing's notion of a referent was quite open-ended. It could be abstract or sublime, such as [52.14, p. 346]:

> *a mood, a picture, an emotion, a physical process, a story ... virtually any coherent image which allows the improviser a sense of engagement and continuity.*

A referent, in other words, is essentially anything that can provide shared musical *seeds* from which an improvisation can grow.

At the longest timescale, Pressing theorized about one's *knowledge base*, which includes the *passive expertise* an individual has about a music style and music culture, as well as the more personal history of one's own compositional choices and predilections. Pressing acknowledged that some improvisation could be referent-free on the intermediate timescale, but he insisted that all improvisation involves certain constraints at the short-term and long-term timescales, since embodied patterns and enculturated knowledge are always present in improvised performance.

Pressing's model is representational and reductionist by nature, but it offers considerable complexity in proposing that improvisers are constantly representing musical structure in parallel, via motor, musical, acoustic, and other ways. His model draws distinctions between improvisational features (e.g., loudness), objects (e.g., a motif), and processes (e.g., sequencing a motif), but it is primarily centered on the level of the musical event. A musical event, for Pressing, might involve an isolated feature, object, or process, but more often it will include all of them.

According to Pressing's model, improvisers first produce a set of musical events (E1). Then each subsequent set of events (E2, E3, E4) – at whatever time grain – takes into account the referent (R) that an improviser has stored in long-term memory (LTM), as well as each previous E – with increased weight given to E instances immediately prior. In ensemble settings, improvisers are also constantly taking account of and representing what the other musicians may have just done. All of this information makes arriving at an optimum solution likely too time consuming and resource intensive, but improvisers, according to Pressing, need only find a good solution, not the best. Pressing views these compounding and constraining factors as providing the potential for unique outputs and novel interactions between musicians. Experienced improvisers, according to him, tend to have a richer knowledge of the referent, an improved motor control and flexibility with their instruments, and an improved cognitive representation with more detail at a smaller grain size about the musical events produced by themselves and others, all of which allows for a mentally nimbler and more responsive performance.

52.6 Referent-Free Improvisation

What, then, of performances that do not, by their own admission, involve a referent? A variety of descriptive terms have circulated at various times and in various locales to describe post-1960 musical improvisation, each with its own group of adherents and each with its own semantic shortcomings. The preferred terms tend to highlight the creative or progressive stance of the performers and the cutting-edge or inclusive nature of the music itself: for example, free or free-form, avant-garde, creative, experimental, contemporary or new, collective, spontaneous, and so on. Stylistic references (e.g., jazz, classical, rock, world, or electronic) are variously included or excluded, as are cultural or national identity markers (e.g., Great Black Music, British Free Improvisation, or Japan Noise). In certain cases in which these specifying linguistic markers are not included, terms may still be indelibly associated with specific individuals or locations: for instance, conduction (Butch Morris), sound painting (Walter Thompson), chance or aleatoric music (John Cage), harmolodics (Ornette Coleman), intuitive music (Stockhausen), reductionism (Berlin), and onkyokei (Tokyo), etc.

The scholarly literature on improvisation also varies considerably on the question of how to interpret the notion of *freedom* in this music. Some authors explain freedom in purely musical terms, as varying degrees of liberation from functional harmony, metered time, and traditionally accepted performance roles and playing techniques [52.42, 43]. For others, defining free improvisation in strictly musical terms misses its most remarkable characteristic – the ability to explore and negotiate disparate perspectives and worldviews [52.44].

Many authors have interpreted free improvisation, and free jazz in particular, as a sociopolitical response to the appropriation and exploitation of African-American music styles [52.45–47]. They focus considerable attention on the birth of the practice during the Civil Rights Movement in the United States and on the music's place within the context of an emerging postcolonial world. Other authors have allied themselves with Marxist or neo-Marxist critiques of hegemonic culture and have focused on free improvisation's implied critique of capitalism and its related market- and property-based economy. For instance, *Eddie Prévost*, a pioneering English improvising percussionist, argues in [52.48] that free improvisation is least susceptible to commodification and that it is, therefore, primarily the domain of those who have felt stifled or excluded. Still others view collective improvisation as a fruitful site for generating social bonds, often across cultural and class divides, or argue for its promising therapeutic uses [52.49].

Given this range of opinion and approaches, how is it that musicians attempting to leave all compositional aspects undecided until the very moment of performance can produce coherent music together in real time? Opinions about free improvisation vary widely, according to *Derek Bailey* in [52.10, p. 85]:

> *They range from the view that free playing is the simplest thing in the world requiring no explanation, to the view that it is complicated beyond discussion. There are those for whom it is an activity requiring no instrumental skill, no musical ability and no musical knowledge of any kind,* he asserts*, and others who believe it can only be reached by employing a highly sophisticated, personal technique of virtuosic proportions.*

If there is a shared bond between these diverse performers, it may be a fascination with both the surprising musical occurrences and interpersonal possibilities inherent to an improvisatory method in which the content

and form of the music at the intermediate timescale are radically underdetermined. In other words, improvised music of this kind explores matters from the microscopic level of expressive detail to the macroscopic level of collective coordination and emergent form, all without an explicit reliance on a shared model or referent. Self-imposed and shared stylistic constraints do often play a generative role in the development of the music, but in referent-free improvisation there is no explicit precommitment to what will happen once the performance begins.

Even with these ideals in mind, it is clear that artists bring to any given performance a lifetime of musical engagement, experimentation, and expectation that is evident at both short-term and long-term timescales. Free improvisation is not a *pure* form of improvisation, or a type of ex nihilo creation, as some of its adherents occasionally suggest. Embodied patterns and learned gestures provide the foundations of this approach on short-term timescales, and across longer timescales free improvisation involves the development of a personal experiential knowledge base (and, some argue, necessitates developing an identifiable *voice* or style) as much as any other practice. Calling the music *nonidiomatic*, as *Bailey* does in [52.10], runs the risk of denying culturally shared sensibilities and understandings, which certainly come into play when the musical backgrounds of the performers and listeners are similar.

It should be noted that a considerable amount of music that is grouped under the broad heading *free improvisation* is not referent-free in the way that I am using the term here, perhaps highlighting that this binary between referent-based and referent-free improvisation is ultimately untenable. For instance, in settings in which musicians have played together before or those in which they are simply aware ahead of time of the body of recorded work that the other musicians have produced (which is to say almost every professional improvised music encounter), these previous experiences often serve as a type of pseudoreferent for the current situation.

Mike Heffley [52.50] highlights three kinds of freedom in improvised music: freedom-from-form, freedom-to-form, and freedom-in-form. Freedom-from-form describes the reactive process of stretching, challenging, and breaking rules and conventions that were once embraced as laws. Freedom-to-form is a proactive step in which rules, patterns, and conventions from other musical traditions, and those of idiosyncratic origin, are embraced as temporary and mutable structures or designs. Freedom-in-form, for *Heffley* [52.50, pp. 279–80], signifies the consummate stage as well as the point at which the process has gone full circle:

> *One path is chosen from among all possible, and its route, uncharted from without, has nonetheless imprinted its own order on the improvising body as a law unto itself . . . that will come in its turn to be so challenged and changed.*

Somewhat analogously, many improvisers confront at various points in their musical careers the seeming paradox that one's instrument can be a means, a subject, and a barrier to expression. In other words, does one choose to use one's instrument to express something musical beyond the instrumental (e.g., singing a song with a saxophone)? Does one follow the more instrumental impulse to explore the sonic possibilities inherent to one's instrument (e.g., keyclicks, slap tonguing, and multiphonics on a saxophone)? Or does one's instrument potentially dictate too much of the content of what one improvises (e.g., can one play the saxophone without thinking saxophonically?).

The dynamics in ensemble situations are even more complex. Clément Canonne argues that collective free improvisation offers a paradigmatic case of the coordination problems which are present, to some degree, in every form of group improvisation. He devises a mathematical model in [52.51] using dynamical systems theory that extends, rather than contradicts, the important work of Pressing on referent-based improvisation. Briefly, on short timescales Canonne agrees with Pressing that clusters of musical events are primarily determined by previous training and embodied patterns. Cognitive limitations play a role (one can't decide too much at the same time), but so does a desire to remain flexible (one does not want to decide too much at the same time). At the intermediate timescale, which may range from many seconds to many minutes, referent-free improvisers must work to create sonic *identities* in the absence of material provided by a shared referent. Sonic identities include processes or features that are developed or explored for a short time, only to give way to new emergent *identities* via some process of gradual or abrupt transition [52.52]. How these transitions arrive and are effected by the ensemble provides some of the more fascinating moments in referent-free improvised performance (see [52.53] for a taxonomy of transition types and [52.26] for a detailed performance analysis using this taxonomy).

In [52.54], *Canonne* adopts the notion of a *focal point* from game theory to explore how potential transitional moments in the music exert various forms of salience on the performance and for the musicians. At these moments, musicians are collectively trying to single out one of many possible trajectories, or to arrive at a convergence of expectations (although, admittedly, wonderful music can also happen, from the listeners

perspective, when the expectations of the performers remain divergent). In empirical studies involving musicians with different musical backgrounds and levels of experience with referent-free improvisation, Canonne found that experienced improvisers most often take advantage of sonic disruptions to signal transitional moments or to singularize musical ideas, turning them into focal points that ensure the vivacity of the musicians' coordination. His data show that seasoned free improvisers were also more skilled than their counterparts in jazz and classical music not only at determining the salience of the sonic environment for themselves, but also in anticipating the salience that sonic events will have for others, often taking into account particular instrumental constraints and affordances, or, when relevant, their shared history of improvising together.

Team reasoning strategies and skills are likely of particular importance to referent-free improvisation settings, since without a shared referent there is little musical *scaffolding* with which to solve group coordination problems. Canonne's dynamical systems model also demonstrates that referent-free improvisation can be self organizing, particularly in situations with a small number of musicians (< 5) who can sufficiently take advantage of their virtuosity, their leadership qualities, and their team reasoning.

Whether improvising in a referent-based or referent-free setting, one's *knowledge base* appears to encompass the embodied (overlearned motoric practices), personal (conceptual, aesthetic, ideological), and culturally shared (tacit understandings) elements that allow for the coherent unfolding of a musical performance. *Canonne*'s most recent empirical work [52.35] highlights how experienced improvisers come to share an implicit mental model of the practice of referent-free improvisation; a type of higher-level knowledge or metastructure that is task-specific instead of piece-specific. He and coauthor Aucoutourier asked experienced improvisers to sort short audio examples of free improvisation into groupings based on their *pragmatic similarity*; in other words, how they would react to the sounds in a performance setting. The authors analyzed these responses and found that the degree of similarity with which participants grouped sounds predicted with better-than-random accuracy their degree of musical familiarity.

In other words, musicians who played together tended to think about improvised music in the same way. The authors conclude that shared mental metamodels such as these may play a key role in the success of referent-free improvisation by allowing more confident *mind reading* of the intentions of fellow improvisers, more frequent cognitive consensus in the course of a performance, and swifter repair of *communication errors* when there is cognitive divergence among the improvisers. They note, however, that the normalizing force of familiarity is not always desired and is therefore frequently countered when musicians opt to play in unfamiliar situations with unfamiliar collaborators (e.g., Derek Bailey's Music Improvisation Company was founded on the idea of having musicians improvise together who had never shared the stage before).

Theorizing about improvisational processes always runs the risk of reifying the result, or of remaining centered on the individual, whether the focus is on skill acquisition, mental representations, rule-based procedures, or on a more integrated view of cognitive schema involving the parallel processing of motor memory, auditory imagery, and audio-motor integration. Much of the early modeling of referent-based improvisation by Pressing and others remained firmly ensconced in an individualistic information-processing paradigm. Berliner's and Monson's influential work on jazz highlighted the complexity and centrality of interactional dynamics in jazz performance, but our understandings of the social dimensions of improvisational performance have often lagged. Interestingly, *Pressing*'s final writing on the subject [52.55] took a more ecological turn, highlighting how the mind, body and environment are all part of an interacting dynamical system.

52.7 Final Thoughts

A distinction has been made here between referent-based and referent-free improvisation primarily dependent on the presence or absence of shared mental models at the intermediate timescale of performance. Referent-based improvisation tends to have well-established (though not necessarily unified) traditions of aesthetic evaluation, and the cognitive schemas used tend to be overlearned so that they can be rapidly accessed and adapted to the needs of the moment. Referent-free improvisation, by contrast, may involve less-well established schemas and less shared conventions, but motor patterns still play a prominent role, and shared metamodels can evolve when musicians play together over time. In the absence of an agreed-upon referent, implicit strategies for group coordination emerge to compensate.

However, we may be best served by considering these as constructed-capabilities-in-action instead of as stored artifact, referent, or model. Improvising without a referent does not discount the importance of culture and experience, nor does it downplay one's embodied and skilled relationship with the available musical resources. Rather, it frames the process of improvising as enactive instead of representational. Knowledge does not emerge from passive perception, or from the analytical study of an agreed-upon referent; it emerges from the need to act in an environment. In this way, referent-free improvisation may offer an exciting window into exploring cognition from an ecological vantage point.

When viewed ecologically, cognition is best understood as a process coconstituted by the cognizing agent, the environment in which cognition occurs, and the activity in which the agent is participating: action, perception, and world are dynamically coupled. In this light, improvisation may be seen as a cyclical and dynamic process, with no nonarbitrary start, finish, or discrete steps (i. e., it is not a token of a compositional megatype). The improviser and the environment coevolve; they are nonlinearly coupled and together they constitute a nondecomposable system.

Improvisers engage with the world and with one another in ways that can not be fully captured by an individual- or brain-centric understanding of cognition. The brain of an improviser is, ecologically speaking, always-and-already in a body and in a niche of musical activity. Instead of having to represent a musical model or shared referent, improvisers *bring forth* a musical world from recurrent sensorimotor patterns and actions. The burden of improvised performance, in this way, shifts from storing and recalling information to detecting it, in the form of ecological invariants and affordances. In place of a computational model of mind – one that stresses the constraints of our cognitive abilities – an ecological one only requires that individuals follow the need to act in their environment, that they orient their actions so as to make the world appear – or sound – a certain way now. Cognitively speaking, this solution is efficient and cheap, and it produces reliable results under a wide variety of conditions.

Whether referent-based or referent-free, improvisation appears to involve a continual tension between stabilization through communication and past experience and instability through fluctuations and surprise [52.56]. This is likely a dynamic of all creativity in the arts and in life, but improvisation draws attention to itself *as* performance: to how it defamiliarizes the familiar. *Gary Peters* offers two related allusions in [52.57]: the improviser as a contestant in a scrapyard reality TV show in which she must fabricate something original out of the discarded materials readily at hand; and the improviser as the subject of Paul Klee's *Angelus Novus*, with her face always turned toward the wreckage of the past as she is propelled into the future by the storm of progress. In both cases, improvisation has more to do with a way of being in the world than with the content or *value* of a fixed work. The critical questions, in the end, involve how we chose to frame these improvisatory dynamics: either as information processing that struggles to keep pace with the cognitive demands of the moment, producing a type of *imperfect* art; or, as an ecologically sensitive engagement with one's sonic and social world.

References

52.1 H. Gardner: *The Mind's New Science* (Basic Books, New York 1987)

52.2 E.T. Ferand: *Improvisation in Nine Centuries of Music* (A. Volk, Cologne 1961)

52.3 B. Nettl, R. Riddle: Taqsim nahawand: A study of sixteen performances by Jihad Racy. In: *Yearbook of the International Folk Music Council*, Vol. 5, ed. by C. Haywood (International Council for Traditional Music, Kingston 1973) pp. 11–50

52.4 S. Block: "Bemsha Swing": The transformation of a bebop classic to free jazz, Music Theory Spectr. **19**(2), 206–231 (1997)

52.5 P. Winkler: Writing ghost notes: The poetics and politics of transcription. In: *Keeping Score: Music, Disciplinarity, Culture*, ed. by D. Schwarz, A. Kassabian, L. Siege (Univ. Virginia, Charlottesville 1997)

52.6 A.B. Lord: *The Singer of Tales* (Harvard Univ. Press, Cambridge 2000)

52.7 D. Sudnow: *Ways of the Hand: The Organization of Improvised Conduct* (Harvard Univ. Press, Cambridge 1978)

52.8 D. Sudnow: *Ways of the Hand: A Rewritten Account* (MIT Press, Cambridge 2001)

52.9 B. Nettl: Introduction: An art neglected in scholarship? In: *In the Course of Performance: Studies in the World of Musical Improvisation*, ed. by B. Nettl, M. Russell (Univ. of Chicago Press, Chicago 1998) pp. 1–26

52.10 D. Bailey: *Improvisation, Its Nature and Practice in Music* (The British Library National Sound Archive, London 1992)

52.11 A. Meyer, P. Frost, K. Weick (Eds.): Organization Science **9**(5) (1998)

52.12 M. Laver, A. Heble, T. Piper (Eds.): Critical Studies in Improvisation **9**(1) (2013)

52.13 B. Benson: *The Improvisation of Musical Dialog: A Phenomenology of Music* (Cambridge Univ. Press, New York 2003)
52.14 J. Pressing: Cognitive processes in improvisation. In: *Cognitive Processes in the Perception of Art*, ed. by R. Crozier, A. Chapman (North Holland, Amsterdam 1984)
52.15 E. Sarath: A new look at improvisation, J. Music Theory **40**(1), 1–38 (1996)
52.16 T. Gioia: *The Imperfect Art: Reflections of Jazz and Modern Culture* (Oxford Univ. Press, New York 1988)
52.17 P. Alperson: On musical improvisation, J. Aesthetics Art Crit. **43**(1), 17–29 (1984)
52.18 P.N. Johnson-Laird: How jazz musicians improvise, Music Percept. **19**, 415–442 (2002)
52.19 B. Nettl: Thoughts on improvisation: A comparative approach, Music. Q. **60**(1), 1–17 (1974)
52.20 J. Sloboda (Ed.): *Generative Processes in Music: The Psychology of Performance, Improvisation, and Composition* (Clarendon/Oxford Univ. Press, New York 1988)
52.21 D. Bastien, T. Hostager: Jazz as a process of organizational innovation, Commun. Res. **15**(5), 582–602 (1988)
52.22 K. Sawyer: Improvisational creativity: An analysis of jazz performance, Creativity Res. **5**(3), 253–263 (1992)
52.23 P. Berliner: *Thinking in Jazz: The Infinite Art of Improvisation* (Univ. of Chicago Press, Chicago 1994)
52.24 I. Monson: *Sayin' Something: Jazz Improvisation and Interaction* (University of Chicago Press, Chicago 1996)
52.25 V. Iyer: Embodied mind, situated cognition, and expressive microtiming in African-American music, Music Percept. **19**(3), 387–414 (2002)
52.26 D. Borgo: *Sync or Swarm: Improvising Music in a Complex Age* (Continuum, New York 2005)
52.27 C. Palmer, E. Koopmans, J. Loehr, C. Carter: Movement-related feedback and temporal accuracy in clarinet performance, Music Percept. **26**(5), 439–450 (2009)
52.28 R. Dean, F. Bailes: Cognitive processes in musical improvisation. In: *Oxford Handbook of Critical Improvisation Studies*, ed. by G. Lewis, B. Piekut (Oxford Univ. Press, New York 2015)
52.29 C. Limb, A.R. Braun: Neural substrates of spontaneous musical performance: An fMRI study of jazz improvisation, PLoS ONE **3**(2), e1679 (2008)
52.30 A.L. Berkowitz, D. Ansari: Generation of novel motor sequences: The neural correlates of musical improvisation, NeuroImage **41**, 535–543 (2008)
52.31 P. Vuusta, M. Wallentina, K. Mouridsena, L. Østergaarda, A. Roepstorffa: Tapping polyrhythms in music activates language areas, Neurosci. Lett. **494**, 211–216 (2011)
52.32 M. Csikzentmihalyi, G.J. Rich: Musical improvisation: A systems approach. In: *Creativity in Performance*, ed. by R.K. Sawyer (Ablex, New York 1996) pp. 43–66
52.33 G. Mazzola, M. Rissi, N. Kennedy, P.B. Cherlin: *Flow, Gesture, and Spaces in Free Jazz: Towards a Theory of Collaboration* (Springer, Berlin, Heidelberg 2008)
52.34 M. Norgaard: Descriptions of improvisational thinking by artist-level jazz musicians, J. Res. Music Educ. **59**, 109–127 (2011)
52.35 C. Canonne, J.-J. Aucouturier: Play together, think alike: Shared mental models in expert music improvisers, Psychol. Music **44**(3), 544–558 (2015)
52.36 T. Järvinen, P. Toiviainen: The effects of metre on the use of tones in jazz improvisation, Musicae Sci. **4**, 55–74 (2000)
52.37 P.E. Keller, A. Weber, A. Engel: Practice makes too perfect: Fluctuations in loudness indicate spontaneity in musical improvisation, Music Percept. **29**(1), 1070–1112 (2011)
52.38 F. Pachet: Musical virtuosity and creativity. In: *Computers and Creativity*, ed. by J. McCormack, M. D'Inverno (Springer, Berlin, Heidelberg 2012)
52.39 G. Weinberg: Gesture-based human-robot jazz improvisation. In: *Proc. Int. Conf. Machine Learn.* (2011)
52.40 J. Pressing: Psychological constraints on improvisational expertise and communication. In: *In the Course of Performance: Studies in the World of Musical Improvisation*, ed. by B. Nettl, M. Russel (Univ. of Chicago Press, Chicago 1998)
52.41 M. Norgaard: How jazz musicians improvise: The central role of auditory and motor patterns, Music Percept. **31**(3), 271–287 (2014)
52.42 E. Jost: *Free Jazz* (Da Capo, New York 1994)
52.43 R. Dean: *New Structures in Jazz and Improvised Music Since 1960* (Open Univ. Press, Buckingham 1992)
52.44 J. Stanyek: Articulating intercultural improvisation, Resonance **7**(2), 44–47 (1999)
52.45 L. Jones: *Blues People: Negro Music in White America* (Morrow, New York 1963)
52.46 F. Kofsky: *Black Nationalism and the Revolution in Music* (Pathfinder, New York 1970)
52.47 V. Wilmer: *As Serious as Your Life: The Story of the New Jazz* (Allison and Busby, London 1977)
52.48 E. Prévost: Improvisation: Music for an occasion, Br. J. Music Educ. **2**(2), 177–186 (1985)
52.49 C.L. Edgerton: The effect of improvisational music therapy on the communicative behaviors of autistic children, J. Music Ther. **31**(1), 31–62 (1994)
52.50 M. Heffley: *Northern Sun, Southern Moon: Europe's Reinvention of Jazz* (Yale Univ. Press, New Haven 2005)
52.51 C. Canonne, N. Garnier: A model for collective free improvisation. In: *Mathematics and Computation in Music*, ed. by C. Agon, E. Amiot, M. Andreatta, G. Assayag, J. Bresson, J. Manderau (Springer, Berlin, Heidelberg 2011)
52.52 C. Canonne, N. Garnier: Cognition and segmentation in collective free improvisation. In: *Proc. 12th Int. Conf. Music Percept. Cogn.* (2012) pp. 197–204
52.53 T. Nunn: Wisdom of the impulse: On the nature of musical free improvisation, http://www20.brinkster.com/improarchive/tn.htm (1998)
52.54 C. Canonne: Focal points in collective free improvisation, Perspect. New Music **51**(1), 1–16 (2013)
52.55 J. Pressing: The physicality and corporeality of improvisation. In: *Sounds Australia, The Jour-*

nal of the Australian Music Center (Ad Lib Edition), http://www.abc.net.au/arts/adlib/stories/s858418.htm (2012)
52.56 A.L. Berkowitz: *The Improvising Mind: Cognition and Creativity in the Musical Moment* (Oxford Univ. Press, New York 2010)
52.57 G. Peters: *The Philosophy of Improvisation* (Univ. of Chicago Press, Chicago 2009)

53. Music of Struggle and Protest in the 20th Century

Anthony Seeger

This is a description of a sound, a poetics, and a political stance in the United States of America during a turbulent century and a half written by someone who grew up in the social milieu described. It endeavors to trace some of the historical and literary roots of 20th-century protest music and discusses the political and musical impact of certain musician-activists on the styles of protest music popular in the second half of the 20th century. These included Charles Seeger, John and Alan Lomax, and Pete Seeger, among others. The tradition of using song to express political ideas flourished in the first four decades of the century, declined due to political repression in the fifth decade, flourished again during the 1960–1980s, and moved to spoken poetry and rap toward the end of the century. For a brief period of time the 20th century forms and performances of music of struggle and protest in the United States had a major impact on music and how music was used in other struggles around the world.

Ethnographies are typically detailed accounts written about a group of people with whom an ethnographer belonging to a different group has had personal experience. That is what the word means – writing about (other) groups. This is an ethnography of a subgroup, a sound, a poetics, and a political stance in a place – the United States – during a turbulent century and a half written by someone who grew up in the social milieu described. Now that the poetry of struggle has largely moved to rap, it is a good moment to reflect on the history of the singer-songwriter of protest music in the 20th-century USA. Research for this essay has involved consulting primary sources such as sheet music, (among the songbooks [53.1–3]); audio recordings (particularly important historical anthologies of recordings include [53.4–8]). A flood of excellent, thoughtful books since 1990 on the US folk music revivals and biographies of some of their major figures, such as [53.9] on Handcox, [53.10, 11] on Woody Guthrie, [53.12] on John Lomax, [53.13] on Alan Lomax; and the publication of original documents of Alan Lomax [53.14, 15] and Pete Seeger [53.16], added to earlier ones on Charles Seeger [53.17] and Ruth Crawford Seeger [53.18], have also made it easier to reflect on the topic today than ever before. Recent books specifically addressing music and musicians in social change were also very useful [53.19–23] as were some older publications [53.24–29]. The author's first-hand experiences as a young member of the Seeger family backstage and in the audiences at Carnegie Hall, in the artists' seating at the Newport Folk Festival, attending numerous concerts and family gatherings, and his experiences as a musician participating in (a few) marches to end nuclear armaments or wars, were an additional resource for this essay. This essay endeavors to trace the historical and literary roots as well as the political and musical impact of certain musician/activists for the particular styles of

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_53

protest music popular in the second half of the 20th century.

Communities and individuals around the world use music to express their struggles or to protest their situations in many different ways. Mass choirs, workers choruses, and popular music styles have been used in the Baltic states, Europe, and Korea. In other parts of the world protest music may not be a local tradition but one adopted from that of another country, as in the performance of heavy metal in the Middle East or Indonesia. Some communities do not use music to express struggle and protest at all, but demonstrate dissatisfaction by refusing to make music and remaining still while others dance. For example, the Suyá/Kĩsêdjê Indians of Brazil considered melodies and texts to be revealed by powerful entities, not made up by individuals. Songs were all parts of ceremonies and could neither be altered nor used for other purposes [53.30]. They wrote no protest songs and had no lullabies. When they were unhappy their protest was expressed through silence – ceasing to interact with others or remaining silent and motionless while others sang and danced. The tradition of using song to express political ideas has flourished, declined, and flourished again in the United States during the past one hundred years. The socially conscious songs of such songwriters and performers as Joe Hill, John Handcox, Aunt Molly Jackson, Woody Guthrie, Pete Seeger, Malvina Reynolds, Phil Ochs, Tom Paxton, Gil Turner, Jim Collier, Joan Baez, Bob Dylan, Holly Near, Bruce Springsteen, and many others have had a powerful impact on those who listened to them, as have the rhythmic speech, beats, and sampling of rappers today. Union leader John L. Lewis is quoted as saying in 1939: *A singing army is a winning army and a singing labor movement cannot be defeated* [53.5, p. 13]. If music were not perceived to be a powerful resource in social and political struggles it would not be so widely censored, controlled, and surrounded with restrictions by governments and religious organizations in different parts of the world. The United States in the 20th century was no exception – politicized music was met with government repression and commercial ostracism.

During much of the 20th century a singer with an instrument performing alone or leading a trio or quartet group was the most characteristic voice of protest music in the US. Their vocal style was often untrained and their instrumentation usually avoided the virtuosic. They often conveyed the hard times and struggles of people other than themselves, though they may have written or arranged the song texts about them. Most of these performers had large repertories and sang many kinds of songs; only a few might be classified as songs of protest. Different performance venues and different political conjunctures elicited different repertories. Many of their topical songs set new words to familiar melodies in a style sometimes referred to as *zipper songs* – you *zip the old song open*, replace the old words with new ones but keep the melody, and *zip it closed* for an instantly recognizable song. Why were these figures so representative of the music of struggle and protest? Where did this model come from? Why was their vocal style purposefully untrained? Why did they write their own songs or song lyrics? How did *folk music* become defined in the United States as someone playing the guitar and singing his or her own songs? Why were certain stylistic features and compositional techniques chosen over others? Those are among the questions this short essay will try to address.

Several kinds of analysis are required to answer such questions. Analyzing the musical sounds of specific pieces is not sufficient, nor is describing the biographies of individual performers. While historical sources can point to some of the literary and scholarly antecedents for the figure of the singer-songwriter aligned with working people on the one hand and groups of people singing together on the other, history alone is not sufficient. It is essential also to consider the specific political contexts of the 20th century within which the music was performed and to recognize the powerful impact of the activities of a few key individuals in the 20th century whose influence on the music was particularly influential.

53.1 Historical Antecedents of Music of Protest and Struggle in the United States

This essay addresses two somewhat different kinds of music, that of struggle and that of protest. Struggle, defined as a *continued effort to resist force or free oneself from constraint* [53.31, p. 3104] is different from protest, which means to make a public declaration against something [53.31, p. 2335]. Descriptions of how times are hard and how it is a struggle to survive are *songs of struggle*. I use the word protest song to mean songs that call the audience to take a specific kind of action. The boundary isn't absolute, as a song of struggle can be used to make a protest and the most important factor in defining a protest song is how the audience

perceives its performance and acts upon it. Instrumental music can also be used to express both struggle and protest through sonic parody, through its tone quality and performance style, by changing the place and time of its performance, and how it relates to a specific social context.

Music of protest and struggle had a long history in Europe and probably also in Africa before it was employed in the United States. Many songs were passed through the oral tradition and most of them were forgotten after the events passed. Broadside ballads – cheaply published poetry about current events printed on single sheets of paper and sung to standard melodies published in England beginning in the early 16th century – often described disasters and political affairs (within limits set by the censors or publishers), and many of them have survived the centuries since their initial publication. Benjamin Franklin wrote ballads about current events for his father's Massachusetts newspaper in the 1700s; an important principle in the freedom of the press was established through a legal case over published ballads in New York City in 1733 [53.32, pp. 373–374]; and songs have been written about every war in US history (from *Yankee Doodle* to *The Star Spangled Banner*, *Waist Deep in the Big Muddy*, *Born in the USA* and hundreds of others). Elections were another perennial subject of songs that would praise one candidate and often ridicule or excoriate his opponent. From 1800 to the 1940s songs were written to support every presidential candidate and in the 19th century political parties regularly published songbooks about their candidates [53.33, 34]. Songs and the social movements that performed them have advocated the elimination of slavery, abstinence from liquor, farmers' rights, the right of workers to form unions, women's right to vote, access to equal civil rights for ethnic groups and for gay, lesbian, bisexual, or transgendered individuals and other causes over the centuries. A long tradition of African-American song has addressed hard life, oppression, resistance, anger, and hope. This centuries-long history of using of music, and especially song, to express struggle and political protest nourished the nation's protest music of the 20th century and advanced the struggle of African Americans for civil rights. Almost every community in the United States has music describing their hard times, and some perceived by community members as protest. But in addition to the existence of protest songs we must explain their specific form, performance style, and the figure of the singer-songwriter.

53.2 The Poet Walt Whitman's Influence on the Image of the of the Protest Singer–Songwriter

In addition to building on an earlier tradition of songs of protest, the United States also developed a tradition of small groups (often trios or quartets) and single creators who identify with worker's struggles and write influential poetry. Singer/poets who identified with the *people* had their roles established in American literature in the mid-19th century by the very influential poet Walt Whitman (1819–1892). Bryan Garman traces what he terms *a race of singers* (specifically male singer-songwriters) back to Whitman (and some of Whitman's ideas can in turn be traced to Herder [53.24]). Whitman has sometimes been called the *father of free verse*, which eschews regular meter patterns and rhyme and tends to follow the rhythm of natural speech – an important feature a good deal of US protest music. Whitman was an admirer of the popular Hutchinson Family Singers, who toured parts of the United States and England singing antislavery songs in the mid-19th century. They filled large concert halls in New York and Boston for high fees and were probably the first commercially successful popular singers of political songs in US history [53.9, 35].

Whitman promoted the idea of a working-class poet-hero whose poetry could have a political impact. He believed that literature could change politics because art and politics were inextricably interrelated [53.36, p. 81]. Writing about music, Whitman suggested that a simple unadorned music should supplant the *stale second-hand foreign method* and replace it with an an aesthetic that subordinated style to substance in the form of *heart singing* [53.36, p. 81]. These sentiments were echoed by musicologist Charles Seeger and Alan Lomax many years later in their calls for an American music based on vernacular or folk music and can be seen in the instructions Ruth Crawford Seeger laid down for singers of folk songs.

If the influence of a 19th-century poet on 20th-century protest music seems unlikely, consider the following: Woody Guthrie read Whitman's influential book of poetry *Leaves of Grass*; Alan Lomax acknowledged the importance of Whitman on his father's and his collections of traditional music and their view of Woody Guthrie. (Alan Lomax wrote *Woody really fulfilled ideal for a poet who would walk the roads of the country and sing the American story in the language*

of the people [53.36, p. 89].) Whitman also certainly influenced the way their supporters perceived singer-songwriters like Guthrie, Seeger, and Springsteen: as poet-heroes of the people singing music that is at once political and artistic and whose verses naturally follow patterns of vernacular speech.

53.3 Ballad Collectors, Songs of Struggle, and Versions of the American Identity

In the 1930s Charles Seeger, John Lomax, and his son Alan Lomax called for a shift from Eurocentric music to an American music based on its folk music that resembled Whitman's preferences for a simpler sound. The search for *folk* or vernacular (local) music had various objectives in the United States, and books of collected songs had been published since the 1840s. The American scholar *Francis Child* published his landmark study of English ballads in the 1880s [53.37], and much of the early collecting of folk songs in the United States focused on narrative ballads performed in the southern Appalachian region in the Eastern United States. African-American music was rarely included in the ballad studies or in the educational programs proposed to teach children their cultural heritage [53.38, p. 27]:

> *The most significant effect of the myth of the white ballad singer was to help block African-American folk song from gaining a central place in the canon of America's musical heritage.*

Returning African-American music and local musical styles to their place in America's musical heritage was one of the goals of John A. Lomax and his son Alan Lomax, whose enthusiasm for recording the music of a wide diversity of Americans led to the establishment of a very different canon of 20th century music and musicians than that of the ballad scholars. Their methods were different from those of the earlier ballad collectors who wrote down the songs from dictation. They went to prisons and sharecroppers' houses and made audio recordings not for publication but for documentation – most of them were deposited in the Library of Congress. The field recordings they made, the songbooks they prepared from them, their symbolically prominent positions at the Library of Congress Archive of Folk Culture, and Alan Lomax's activities in radio, theater, and record productions all had a huge impact on the music of the 20th century. Alan wrote this about how going to people's homes to record them gave the documenter a very different perspective on the music than scholars who studied only the song texts in manuscripts [53.39, p. 365]:

> *The literary scholars, who seldom went into the field and collected their songs in the homes of the singers, have failed generally to perceive the undercurrent of tragedy and of protest that underlay the songs they footnoted and published.*

A number of important figures in the future of the folk music revival and songs of struggle and protest would meet in the 1930s and their intertwined lives would have a powerful effect on subsequent decades. These were Charles Seeger, his son Pete, and children of his second wife, Ruth Crawford; John Lomax, his son Alan, his daughter Bess, and other members of his family; Woody Guthrie; Lead Belly; and other artists. The following paragraph briefly summarizes their biographies because they appear repeatedly in this essay.

Charles Seeger (1886–1979) merits the longest description not only because he was – among other things – a systematic musicologist but also because his career took many intricate turns and three of his children became influential musicians. Born in Mexico City to an American businessman with a successful career in imports, C. Seeger studied music at Harvard (where he hated music history) and then studied in Cologne, Germany, and did some conducting at the Cologne Opera. This is not a place to summarize his career, but suffice it to say that he was a theorist-composer with great self-assurance, unusually extensive international experience by virtue of his early years in Mexico and his studies in Germany, and an organizer. His political radicalization in 1913 or so led to his dismissal from Berkeley and was followed by the end of his marriage to Constance. He was an active member and one of the theorists of the Communist Party-funded Pierre Degueyter club and taught composition in New York City for a few years. He married a young composer Ruth Crawford and moved to Washington DC to work on government programs for the arts in 1936 and later was head of music and arts programs at the Pan American Union (precursor to the Organization of American States). He was very active in national and international professional organizations. He was a founding member of the New York Musicological Society, the American Musicological Society, The Society for Comparative Musicology (founded as a way to give continuity to the work of the nearly defunct Gesellschaft für Ver-

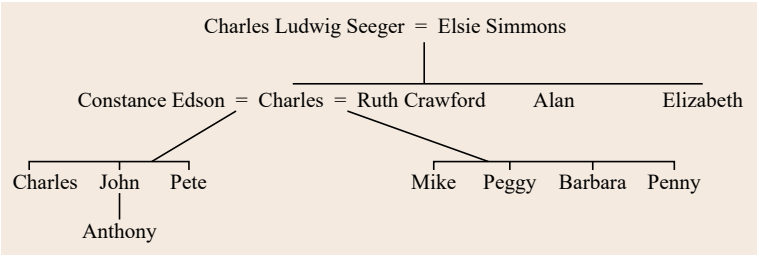


Fig. 53.1 Simplified diagram showing Charles Seeger, his parents, siblings, seven children, and the author

gleichende Musikwissenschaft in Germany because so many of its members left Germany), and the Society for Ethnomusicology. He was active in UNESCO NGO, including the International Music Council and the International Council for Traditional Music. He saw music as embedded in social life and this influenced his musicological work, his practical work in US government programs, and his theoretical writings about music and social transformation. He was also an advocate of what would later be called *applied musicology* [53.40, pp. 227–228]. Charles could look at disparate phenomena and theorize about them with an assurance that often convinced others, though much of his later writing was hampered by a very dense prose style that is difficult to read. Charles had three sons with his first wife, Constance, a concert violinist. The youngest was Pete. Their marriage eventually ended in divorce. Several years later Charles married one of his composition students, Ruth Crawford, and started another family that eventually included four children, two of which – Mike Seeger and Peggy Seeger – were active and influential professional musicians. Figure 53.1 is a very simplified family tree showing Charles and his children, and only one of his many grandchildren (the author).

John Avery Lomax (1867–1948) was raised on a central Texas farm, put himself through college and business courses, and worked for many years in college teaching and administration. Long interested in cowboy poems and songs, at Harvard he was encouraged to pursue the study of Texas traditions and published the influential book *Cowboy Songs and Other Frontier Ballads* [53.41] in 1910 with an introduction by President Theodore Roosevelt. He became a popular lecturer on folklore and often gave lecture-demonstrations in which he talked about and then performed the songs. Hard hit by the Great Depression and the death of his first wife, Lomax moved to Washington DC where he was appointed honorary consultant and curator of the Archive of Folk Song at the Library of Congress and, like Charles Seeger, worked for agencies of the US government. With his son Alan he collected thousands of recordings for the archive and published several songbooks. He had four children, Shirley, John A. Jr., Alan, and Bess. Like the Seeger family, several of his children and grandchildren were influential in American music. Alan was a brilliant recordist, a master of media production, and an intrepid traveler and researcher whose impact on American music was tremendous, and whose work was also influential in the popularization of music of struggle and protest.

The encounter of the Lomaxes in Washington DC with musicologist Charles Seeger, his son Pete Seeger, and Charles' second wife Ruth Crawford Seeger and their children was an important stimulant of the musical ferment and the performance style of singers of folk and protest music in the 1930s and 1940s and into the 21st century. Ruth Crawford's avant-garde sensibilities led her to recognize the musical subtleties of the musicians the Lomaxes recorded, which she carefully transcribed for some of their publications. Alan Lomax, Pete, Mike, and Peggy Seeger's lives were intricately intertwined in many ways over the decades.

53.4 The Vocal Style and Performance Practice of US Protest Music

Peggy Seeger has suggested that a key to understanding the stylistic continuity in American protest song is the use of vernacular speech patterns in the songs. The rhythms and melodies of the songs often follow those of spoken English (Peggy Seeger personal communication). In the late 1930s Peggy Seeger's mother, the composer Ruth Crawford Seeger, transcribed dozens of field recordings made by the Lomaxes of American vernacular music for the songbook *Our Singing Country* [53.39]. Although vastly shortened by the Lomaxes for their book, her introductory essay includes an important discussion of folk performance style as discerned from the recordings. The whole introduction, without edits, was only published in 2001 [53.42]). The sixteen instructions she gives to readers of the book who might want to sing the songs from her transcriptions are

a good characterization of the vocal style and performance practice of a lot of music of struggle and protest during the 20th century (adapted from [53.43, p. xxxi–xxxiii]):

1. Do not hesitate to sing because you think your voice is not good. The songs are better sung in a natural voice than a bel canto voice.
2. Do not sing *with expression* – maintain a level of more or less the same degree of loudness or softness from the beginning to the end of the song.
3. Do not slow down at ends of phrases, stanzas, or songs.
4. Do not hesitate to keep time with your foot [and] unless otherwise indicated sing with a fairly strong accent.
5. Do not *punch* or typewrite out each tone, when two or more tones are to be sung to one syllable of text.
6. Do not make too much difference between major and minor degrees in songs containing both.
7. Do not feel, in group songs, that these songs require harmonizing.
8. Do not hesitate to sing without accompaniment.
9. When doing so do not make noticeable pauses between. stanzas. When accompaniment is required a guitar or banjo is to be preferred.
10. Remember that most songs begin with the chorus and end with the chorus.
11. *Do not sing down* to these songs. Theirs are old traditions, dignified by hundreds of thousands of singers over long periods of time.
12. Her last injunction was *listen to phonograph recordings of these songs and others like them.*

Although recordings were available during the 20th century, radio and television rarely featured political and topical songs for most of the first half of the 20th century. Other kinds of mass media were important for the transmission of music of protest and struggle. The labor songs of the 19th century were transmitted through newspapers and in songbooks. Quite a few political movements generated their own songs in the late 19th century, but the Industrial Workers of the World (IWW) used songs quite extensively for entertaining members and for drowning out competing musical messages from other groups on public street corners. One of the IWW's best remembered singer-organizers was a man known as Joe Hill ([53.44], also known as Joseph Hillström). Hill wrote *zipper* songs of struggle set to popular melodies of the time, which made them easy to learn and to sing. Hill wrote from jail in 1914 that (cited in [53.45, p. 1]):

> *A pamphlet, no matter how good, is never read more than once, but a song is learned by heart and repeated over and over. There is one thing that is necessary in order to hold the old members and to get the would-be members interested in the class struggle and that is entertainment.*

Hill was probably framed for murder and executed in 1915. (Joe Hill is best remembered by many through the song written after his death, *I Dreamed I Saw Joe Hill*, written by Alfred Hayes in 1930 and memorably performed by Paul Robeson and later by Joan Baez.) The Industrial Workers of the World published a song book entitled *Industrial Workers of the World Songs to Fan the Flames of Discontent*, which became known as *The Little Red Songbook* [53.1, first edition 1909]. Although IWW membership diminished in the 1920s, the music and the organization continued and the idea that a good song is sung a thousand times and that workers need entertainment was influential to generations of songwriters and performers. Song books and songsheets continued to be published throughout the 20th century in order to introduce dispersed readers to new songs. Among them was a quarterly bulletin of people's songs from 1946–1950 [53.21, pp. 179–220], *Sing Out!* magazine from 1950–present, *Broadside* magazine from 1963–1988 [53.6] and *Fast Folk Magazine* (1982–1996) [53.7]. *Fast Folk Magazine* did not focus on songs of struggle and protest and used compilation LPs and CDs to supplement the printed page, but like the earlier magazines the work was collective, the song was thought to be more important than the singers, and the objective was to make the songs available for others to sing [53.46].

Audio recordings, both commercial and *field* recordings, reveal that before Charles Seeger and the Lomaxes called for a vernacular American sound a lot of Americans were already singing that way at home and in public. There is no space here to review the history of recorded sound in the United States, but both early commercial 78 rpm records and field recordings reveal a common singing style in the southern Appalachians that was also employed by local performers, textile workers and coal miners, and by Woody Guthrie, often by Pete Seeger and the Almanac Singers, and from them to other performers. Commercial audio recordings and private field recordings provided important models for generations of singer-songwriters in the 20th century. Eloquent songs of struggle are found among the large number of commercial recordings made of local southern musicians and released by major labels to sell back to members of the communities in which they were recorded in the 1920s through the mid-1930s. The recordings of rural musicians included songs of hard times, of tragic events, armed conflict, and of protests. They were often sung by solo

singers or family groups and the accompaniment was often on instruments of a string band (banjo, guitar, and or fiddle). Both the recording process and many of the artists on them ceased recording during the economic depression of the 1930s and after World War II most large record labels dropped the strategy of recording local artists for regional markets in favor of producing national hits. Preserved by collectors, the original 78 rpm records provided an inspiration to many urban performers in the 1950s and 1960s. Woody Guthrie, for example, is said to have listened to and imitated the guitar style of the Carter family (as well as using some of their melodies for his songs). In 1952 Harry Smith, an artist and record collector, assembled a long play record anthology from out-of-print 78 rpm records published between 1926 and 1934. His *Anthology of American Folk Music* was released on Folkways Records and became an important reference for the folk music revival. Among the songs of industrial transformation and rural or domestic struggle on that anthology that were subsequently played by folk revival artists were *Peg and Awl* by the Tar Heel Boys, *Down on Penny's Farm* by the Bently Boys, *Single Girl Married Girl* by the Carter family and *How Can a Poor Man Stand Such Times and Live?* by Blind Alfred Reed.

There were strong differences of opinion even within the Seeger family itself about the right way to perform these vernacular songs from an earlier era. Mike Seeger tended to take a curatorial approach, playing and singing in a way similar to the originals. Pete Seeger took a much freer approach, and often altered the melody, changed the accompaniment, and sometimes shortened or adapted the lyrics. Peggy Seeger seems to have done some of each. Both Pete and Peggy composed new songs in a variety of styles. Arguments about authenticity and representation surfaced throughout the folk revival of the 1950–1970s. Having the actual sounds from an earlier era available for consultation was thus both a boon and a point of contention. But the early commercial and field certainly influenced the endurance of a characteristic performance style.

53.5 20th Century Politics and Protest Music

The style and content of protest music does not simply derive from 19th-century poets, individuals, songbooks, and audio recordings. The policies of the Communist Party USA (CPUSA) played an important role in shaping a national culture of protest song in the US, as is ably described by *Reuss* [53.21, p. 1971] and *Dennisoff* [53.47]. Reuss traces the move to politicize music and all artistic expression to a speech by Bukharin at the 1928 World Congress of the Communist International [53.21, p. 40]:

> *The third period ideology that Bukharin promulgated had stern implications for American Communists' attitudes toward culture. All artistic expression was to be politicized.*

In 1931 the Workers Music League was formed and in 1932 the New York City section of the league founded the Pierre Degueyter Club (Pierre Degueyter had composed the *International* and had died in 1932). One of its divisions was the Composers Collective, which *developed the most important theoretical statements on music in the American Communist movement* [53.21, p. 44]. Among its active members were Charles Seeger, Mark Blitzstein, Henry Cowell, George Antheil and Elie Siegmeister. Writing under the pseudonym Carl Sands, Charles Seeger wrote in the *Daily Worker* 1934, (cited in [53.21, p. 45]):

> *Music is propaganda – always propaganda – and of the most powerful sort and: The special task of the Workers Music League is the development of music as a weapon in the class struggle.*

By the late 1930s the CPUSA was more enthusiastic about using local folk traditions to further the struggle. In the 1950s, declared an illegal organization and its members persecuted, the CPUSA lost many of its members when Stalin's atrocities were verified and the tanks rolled into Poland in 1956. But the impact of its involvement in the arts was felt by many protest musicians in the 1950s during the Cold War. After the Composer's Collective invited southern union organizer and singer Aunt Molly Jackson to attend their meetings, Charles Seeger realized that her vernacular style would be more successful in communicating radical ideas to the masses than the avant-garde compositions of his fellow members of the Composers Collective. He reportedly said to her *Molly, they didn't understand you. But I know some young people who will want to learn your songs*. One of them was his son, Pete Seeger [53.21, p. 53] whose political positions, organizational skills, writing, musicality, songwriting and commitment to getting audiences to sing with him had a very powerful impact on performances of music of protest and struggle in the second half of the 20th century.

Like his father, Pete Seeger founded and led a number of different organizations, was at the center of topical song activities and was a member of two influential musical groups. He was a founding member of a songwriting and performing collective that called itself the Almanac Singers. (Most histories of American topical song and the folk music revival of the 1960s highlight the Almanac Singers, two of whose members became important in the popular group The Weavers after World War II [53.21, pp. 147–148]. Although short-lived, the group and its songs had a strong impact on later developments.) The group was based in a shared apartment in New York City, wrote songs collectively, and recorded several albums of protest music and two albums of folk music standards in the early 1940s. The group recorded an album of antiwar songs shortly before Germany invaded Russia, after which it began to record anti-Hitler songs and virtually disbanded as its members entered the war effort. Pete Seeger volunteered in the Army and served in the Pacific as a musician; Woody Guthrie joined the Merchant Marine; Bess Lomax joined the defense effort as well. After returning to civilian life Pete Seeger became an organizing figure behind a postwar organization called People's Songs (1945–1950) and its associated booking agency, People's Artists. When those failed he was instrumental in starting *Sing Out!* magazine (1952–the present), where he wrote a regular column for decades. He also financially supported the establishment of *Broadside* magazine (1962–1988), which published new topical songs [53.6]. *Broadside* was published by Agnes 'Sis' Cunningham and Gorden Friesan, whose careers had been ruined by the anti-Communist persecution of suspected Communist Party members. There were many contributors to *Broadside* whose songs were first published in this magazine, among them Bob Dylan, Sis Cunningham, Phil Ochs, Tom Paxton, Peter LaFarge, The Fugs, Nina Simone, The Freedom Singers, Janis Ian, Luis Valdez, Jim Collier, and many others.

53.6 African-American Musical Traditions and Social Protest

African-American composers, musicians, and scholars have contributed in myriad ways to musical life in the United States and it is impossible to do them justice here. While their contributions to ragtime, jazz, and rap were original, well documented, and powerful, so too was their use of music to describe their struggles and hopes in blues forms, in church music, and in music of struggle and protest. There were many African-American union organizers in the 1930s and some of them composed powerful songs. One of these was John Handcox [53.9] who authored *Roll the Union On* and *Mean Things Happening in this Land* among other enduring songs. Another was Lucille Simmons, who arranged a composed religious song *If my Jesus Wills* by Louise Shropshire into what, after some further arrangement, became known as *We Shall Overcome*. Simmons was renowned for her very slow renditions of well-known songs on the picket line [53.3, pp. 32–35]. African-American artists' use of community singing in the political movements for unionization in the 1930s and for equal rights in the mid-20th century proved to be a powerful organizing tool and the use of music in demonstrations became a model for using music in other political struggles, especially the anti-Apartheid protests in South Africa.

The Civil Rights movement of the 1950s and early 1960s was one of the most musical social movements of the 20th century. Some of its music came out of earlier African-American community organizing and some of it was invented on the streets and in jail by young demonstrators who took popular songs of the moment and *zipped* into them new words for the occasion. Just as a great deal of the organizational strength of the movement came from African-American churches, some of the best-known movement songs come from well-known religious songs. An influential a capella group called The Freedom Singers performed at many demonstrations and other public events. Bernice Johnson Reagon, one of its members, said that singing together kept them united and also kept them from feeling afraid. Protesters sang on the streets, in the jails, and in concerts in the North to gain support outside the South. This musical movement drew some of its strength from group singing – community singing like that found in many African-American churches. This was quite different from the tradition of the singer-songwriter, and its use and successes were very different. The Civil Rights movement was marked by tragedy and only partially successful in obtaining equal rights for African Americans in the United States, but through it some important gains were made. The assassination of Dr. Martin Luther King, the continued institutionalized racism, and the shooting and imprisonment of young African-American men made the tone of *We Shall Overcome* too optimistic and has led to a variety of other musical expressions of protest including rock and rhymed poetry.

53.7 The Conservative Reaction

If the changing policies of the Communist Party in the first half of the 20th century influenced the shape and development of protest music, the subsequent government investigations, trial by innuendo, and repression of political singer-songwriters also shaped its virtual demise in the 1950s. The repression was also shaped by larger institutions, collusion between government agencies and civic organizations, and a resulting climate of fear and accusation that tore apart long-standing allies, shattered friendships, and ruined many careers.

The United States does not have an official government office of censorship. Freedom of speech has protection in the US Constitution. Social movements, especially union organizing and establishing equal rights for African Americans, however, were violently repressed throughout the early 20th century by local police, private protection agencies hired by companies, State-controlled national guards, discriminatory laws, and businesses. These were sometimes unofficially aided by the Federal Bureau of Investigation. Musicians and other performers were rarely imprisoned or executed because of their music. But they were investigated and their careers ended because of who they performed for. There was a period from the late 1940s into the 1950s when many songwriters and performers of music of protest and struggle were accused of being Communists. Many people not only lost their jobs, they were *blacklisted*, or marked as people who should not be hired by any other organization. Anyone hiring them or defending them would also be suspect. The US Congress established the House Un-American Activities Committee (HUAC), whose objective was to investigate subversive activities on the part of civilians, government employees, and organizations with suspected Communist associations. In 1950 a pamphlet called *Red Channels* appeared with a long list of alleged Communists, and many of those whose names were on the list soon found themselves unemployable in their fields, avoided by their former friends, and the subjects of harassment and government investigation. Those were terrible times. If a person under investigation confessed to his or her errors and agreed to denounce other people who were supposedly active in Communist activities, that person was sometimes taken off the blacklist (but even those who collaborated by *naming names* often found they were still unemployable). When I was about seven I was playing a 78 rpm record of the Almanac Singers singing union songs on a hot spring evening in our apartment in New York City. My father, John Seeger (Pete Seeger's older brother) stormed angrily into the room, shut the window with a crash, and said *don't you ever play those records again with the window open*. He feared that neighbors might denounce us for the music we played. This childhood moment, which I remember vividly, was probably the origin of my certainty that music is deeply intertwined with society and culture and the beginning of my becoming an anthropologist and ethnomusicologist. This was the insidious and terrible thing about the blacklist – it created fear and suspicion and caused musicians who had once been best of friends to turn upon one another. My parents, teaching at a private progressive elementary school, did not suffer much from blacklisting but most of the rest my extended family did. Charles Seeger had his passport confiscated and resigned from his job at the Pan-American Union; his oldest son, radio astronomer Charles Seeger Jr., found that no one would hire him in the US and emigrated to the Netherlands for a position at Leiden University; Pete Seeger was investigated by HUAC and found in contempt of Congress for refusing to answer their questions on the basis of the constitutional rights to free assembly and speech. His career was shattered, he was forced to mount an extended legal defense, and he did not appear on television for 15 years. Even those family members who were not overtly political themselves suffered from the other family members' blacklisting. Mike Seeger was paid through a friend because a concert organizer did not want to list his name on the roster of paid performers. Peggy Seeger moved to England, as did Alan Lomax. Even I may have been a target. At age 16 a friend and I entered and won a small-town talent contest in 1960 whose first prize was $ 25 and an appearance on local television. The television appearance never occurred, and I never found out why – but other Seegers were banned from television and my last name probably led to dropping us from the show. The 1950s is stereotyped as a bland moment in history between World War II and the turbulent 1960s – but this was partly the result of the active silencing of many voices whose songs (and films, and journalism, and other arts) could not reach a wider audience.

53.8 The Folk Music Revival and The Commercialization of Folk Music

The beginning of the American folk revival is given different dates by different authors. Some trace it back to the folk dance movement of the 1930s and 1940s; others start it with the popularity of the Kingston Trio singing *Tom Dooley* in 1958 [53.48]; yet others divide it into different parts. If *revival* means *the emergence of the commercial successful performers of traditional songs and new compositions in a similar style*, which I shall use here, I would consider the debut of the group The Weavers to be the key event. Their performance of Lead Belly's *Goodnight Irene* was on the top of the music charts for 13 weeks in 1950 and their song *Tzena, Tzena, Tzena* was a hit the following year. Comprised of two former members of the Almanac Singers, Pete Seeger and Lee Hays, and two new additions, Ronnie Gilbert and Fred Hellerman, their first manager suggested they downplay the political songs of their past. Their first album featured rich orchestral arrangements by bandleader/arranger Gordon Jenkins. Yet they did perform songs by Woody Guthrie, Lead Belly, and from the Lomax books, as well as an array of international songs, including an Indonesian lullaby, an Indian song favored by Gandhi, and the South African *Wemoweh* (Mbube). For a short time they were in high demand and sold many recordings, but Pete Seeger and Lee Hays were identified by an informer (who later recanted) as members of the Communist Party. They were called to testify before congress and The Weavers were blacklisted – their record contract canceled, denied radio play and television appearances, and most of their bookings canceled as well. The investigation of The Weavers and a number of other musicians drove the musicians and their songs *underground*. But The Weavers had established the possibility of folk music having a large popular audience and many future musicians were inspired by The Weavers and by Pete Seeger's solo concerts in colleges.

Folk music going *underground* in the 1950s meant little media exposure and playing live at small venues such as coffee houses, college auditoriums, children's schools and summer camps, which were not as influenced by the blacklist. Pete Seeger once reflected that a good result of the blacklist was that he could influence a new generation of musicians who were at the colleges and schools where he was forced to perform to make a living. Many future folk singers were influenced by his concerts. Dave Guard of the Kingston Trio heard Pete Seeger at a concert where Joan Baez was also in the audience [53.48, p. 5]. Pete's college concerts had a powerful impact on a generation coming of age in the late 1950s and 1960s. (Several recordings of live concerts given by Pete Seeger during that period are available. One is a complete 1960s concert at Bowdoin College [53.49] and his liner notes to P. Seeger 1993 discuss his approach to college concerts in some detail.) Students learned to play guitars, banjos, and other instruments and used the increasing number of LP records and music magazines and books to study technique and to learn words. This new enthusiasm for traditional music had several distinct venues.

Washington Square Park, in Manhattan, New York City was a meeting place for budding musicians in the 1960s. *The Village*, a neighborhood in New York City surrounding Washington Square Park, had long been associated with creative bohemians. It was a neighborhood of small coffee houses and nightclubs. It had stores featuring folk song books and instruments such as Izzy Young's Folklore Center. Washington Square itself became known as a gathering place for musicians who would jam together for hours on end on Sundays and go on to the clubs in the evening.

Music festivals were another important venue. Short gatherings of musicians performing for audiences have a long history in the United States [53.50], but they became an especially important venue for the folk revival in the 1960s. The most important of the folk festivals in the 1960s was the Newport Folk Festival. Conceived by George Wein and a board including Theodore Bikel, Oscar Brand, Pete Seeger and Albert Grossman, the festival drew large crowds by inviting popular folk revival musicians and then enriching the crowd's experience by also bringing more *traditional* or local artists and introducing their music to new audiences. The Festival also featured workshops and opportunities to talk with and learn from the participants. One of its most important policies was to treat all the artists alike: they all received the same performance fees for their participation, whether they were selling thousands of recordings or had sold a few hundred thirty years earlier and were coming out of retirement to showcase their skills at the Festival. The festivals also gave artists a chance to get to know one another and to play together – an important feature of most festivals. Representatives of record labels were there as well. Newport was famous for its role in the careers of Joan Baez and Bob Dylan, who went on to great commercial success and also for Cajun musician Dewey Balfa who found that his popularity at Newport changed the image of Cajun music in Louisiana and could build on that opportunity for cultural projects of his own.

Summer camps are often overlooked, but have been an important venue for introducing children to group singing. On the east coast of the United States there has been a long tradition of sending children away from the

cities to spend one or more weeks in the countryside playing sports, hiking, and (almost everywhere) singing songs. In 2014 there were nearly twelve thousand summer camps of all kinds that were attended by nearly eleven million children and adults [53.51]. The music at these camps varies, but almost everywhere children learn a distinctive camp-specific repertoire they remember most of their lives. In the 1940s some camps were run by unions, and their songs include socialist anthems. In the 1950s and 1960s, college students who were learning folk music took jobs teaching music at many summer camps where they introduced younger children to their favorite musicians and songs [53.52]. Summer camps have had a large impact on the music that children sing, and also on the training of song leaders and musicians [53.40].

In spite of these alternative, fairly noncommercial venues for performers, the blacklist ruined the careers of many fine people and created a toxic environment for the performance of protest songs. This slowly changed in the 1960s, partly because a younger generation of songwriters who had never been affiliated in any way with the Communist Party began to write their own songs – among them Phil Ochs, Tom Paxton, Bob Dylan, Jim Collier – and other groups began to sing protest songs again, like The Kingston trio (*Where Have All the Flowers Gone*), Peter, Paul, and Mary (*If I had a Hammer*), Joan Baez, and others. Although vocal quality changed, most of these musicians sang in fairly natural voices (not bel canto and with little vibrato) and played guitar or banjo either solo or in a group. Many of them continued to perform for decades, but other musicians innovated new styles to express their protest and opposition to the status quo – among them electrified bands playing rock, punk, and other genres. (The term *folk music* in the United States today usually refers to acoustic music performed by a singer-songwriter accompanied by a stringed instrument. The phrase has less to do with an affiliation with vernacular styles than with identifying it as distinct from rock, jazz, heavy metal, and other genres.)

Two of the most famous, and most popular, musicians to emerge partly through their performances at Newport Folk Festival were Joan Baez and Bob Dylan. Joan Baez, who grew up in a pacifist Quaker household of a university professor, sang a wide repertoire of old ballads, songs composed by others, and songs of protest and struggle in a beautiful, clear voice. Although she did write some songs of her own, her most riveting performances were interpretations of songs by others. Her performances at the 1959 Newport Folk Festival and in local clubs led to a contract with Vanguard Records that year. Baez's Quaker background and political sympathies led her to oppose war and injustice throughout her career. She made a lifelong commitment to using music as part of a protest against wars and injustice. In 1964 she cofounded the Institute for the Study of Nonviolence and was arrested for protesting the induction of soldiers in California. Joan Baez has sung for many social causes – antiwar, civil rights, opposition to the death penalty, gay and lesbian rights, and poverty – and participated in many public events during her entire career, including a performance at Occupy Wall Street and another in the White House for President Obama.

Bob Dylan (born Robert Alan Zimmerman) arrived in New York City in 1963 from Minnesota, well versed in the songs from Harry Smith's *Anthology of American Folk Music* as well as contemporary performers, as is evident from the eloquent first volume of his *Chronicles* [53.53]. He admired Woody Guthrie, visited him in the hospital, and became a *Woody Guthrie Jukebox* and, like Guthrie, became a brilliant and prolific songwriter. Dylan's first album was mostly covers of earlier songs but included two original compositions. After this Dylan mostly performed his own songs, driven by his poetic gifts and dedication to personal discovery. Dylan wrote some of the most iconic and popular songs of the movements of protest and struggle, among them *Blowing in the Wind*, *Hard Rain's Gonna Fall* and *The Times They Are a-Changing*. His restless creativity led him to move beyond folk music in the mid-1960s and to explore the musical potentials of rock, blues, country, and other genres. There is a very large literature about Bob Dylan, and this is not the place to enter the many debates that characterize it. His early protest songs endured through performances by other singers long after he stopped performing them.

The critically important thing about the folk music revival of the 1960s was that musicians could make much more money through their recordings and live performances than had previously been possible. Many musicians had some protest music in their repertoire, but few musicians of any generation sang only music of protest. Even the Almanac Singers recorded folk songs as well as Union and antiwar songs, and this was true of most later performers. Like love songs, songs that describe how hard life is can be sung almost any time; songs of protest tend to appear a specific times and aimed at specific things, such as abolishing nuclear weapons (1950s) or ending the Vietnam War (1960s). Many heated debates about *authenticity* and *selling out* to commercial interests appeared in the pages of *Broadside* and *Sing Out!* Magazine during that period and were extensively debated by fans and musicians like. Issues of copyright and attribution arose, and some of them affected the recording process. This was a very different period than

the 1940s when the Almanac Singers composed collectively and actively sought to perform for union meetings.

As the musical potential of electric instruments became apparent, some musicians used the new timbres and sensibilities for new kinds of protest music. Many rock musicians have also written and performed music of social commentary and been critical of political trends in the United States. One of the most famous and consistent of these is Bruce Springsteen, whose working-class roots and expressive song-writing at times expresses a deep empathy with and outrage about the struggles of workers, veterans, and the terrible results of economic depression. Like Guthrie's youth in dust-bowl Oklahoma, Springsteen's youth growing up in New Jersey marks some of his most powerful music. Garman writes *perhaps more than any contemporary popular artist, Springsteen has kept his eyes fixed firmly on the dirty ways of the world* [53.36, p. 251]. Springsteen has performed for quite a few different causes, and acknowledges his debt to Woody Guthrie and songwriters who came before him. Rock music is a huge field, of course, and many bands espouse racist right-wing opinions and use recordings to recruit supporters and spread their ideas own as a means of recruitment and affirmation [53.54]. (I could have written an entirely different essay that focused on the music of right-wing politics in the United States, but I know much less about it and as far as I know it has been far less researched. It was shaped by some of the same historical traditions and social processes as left-wing music, but musicians frequently used different musical genres for its expression.)

Woody Guthrie, Pete Seeger, and Bruce Springsteen all came together at a memorable moment in an outdoor concert in Washington DC on January 19th, 2009 as part of the inauguration events of the first African-American President of the United States, Barack Hussein Obama. On a cold afternoon they, along with Pete Seeger's grandson Tao Rodriguez Seeger and a school chorus, led a huge crowd in singing Guthrie's best known song, *This Land is Your Land*. They not only sang the well-known celebratory verses but also two verses Guthrie wrote that are rarely included in school books – verses about poverty and the excesses of private property. The truly political nature of Guthries' composition was highlighted by two of the best known singer-songwriters and *people's poets* who were strongly influenced by Guthrie and the way he wrote and used songs. Thousands of mobile phones documented their performance and sent it to people all over the world. In a country with a long tradition of enslaving African peoples and a recent history of discrimination, the election of President Obama was a moment when music could celebrate, commemorate, and affirm its active participation in the social processes of the United States. In the words of the last verse: *Nobody living can ever stop me as I go walking that freedom highway. Nobody living can make me turn back; this land was made for you and me*.

53.9 Conclusion

For most of the 20th century, music of protest and struggle in the United States had a particular musical and narrative style as well as a few characteristic performance practices. This can be traced to the Freedom Singers in the 1950s and the Almanac Singers in the 1940s and then further back to African-American musicians, labor organizers, and songwriters from the mines, cotton mills, and farms of the Southern US. In addition to those influences, a long history of protest music in the United States contributed to the 20th-century style. The music and the performers were shaped by Walt Whitman, by international and national political movements like the Communist Party, and by the vernacular singing styles of the southeastern US. Some of the performers were certainly exemplars of Walt Whitman's people's poet. Their singing style, their conviction that music could influence politics, their compositions, and their courage were an important part of the music of 20th century and continue to influence some strands of American music today. Twentieth-century political music was also shaped by the passion and talents of specific people at particular moments. Among these were the Lomaxes and the Seegers, who began their collaboration in the 1930s. Through their passion for vernacular music and their many different activities they had a profound influence on the music of the second half of the 20th century. But there were many other influential figures as well. The music they performed, encouraged, recorded, or studied was enriched by their work, their teaching, and by those who learned from them and contributed to music of struggle, protest, criticism, and hope that continues today in other forms.

Protest marches during the past fifteen years have been more often animated by group chants than by song. The winning unions (there aren't many – most unions are losing members) are not employing music as much as other organizing tools. Poets of protest are far more apt to use rap and rhythmic speech than melodic

verses – though there are of course many fine exceptions. Songs of protest and struggle are sung in many different languages as well, as they probably always have been. Many members of the general public move to their own soundtracks, hosted on personal electronic devices, rather than to collective music. But there is a lot to protest about in the United States and the centuries-long traditions of using music and speech to protest and to try to effect change will continue to stimulate activists in the future.

References

53.1 IWW (Industrial Workers of the World): *The Industrial Workers of the World Songs to Fan the Flames of the Little Red Songbook* (W. Oliver, London 1916), http://www.musicanet.org/robokopp/iww.html

53.2 S.H. Friedman: *The Rebel Songbook* (Rand School, New York 1935)

53.3 P. Seeger: *Where Have All the Flowers Gone? A Singer's Stories, Songs, Seeds, Robberies* (Sing Out Publications, Bethlehem 1993), https://en.wikipedia.org/wiki/We_Shall_Overcome

53.4 H. Smith: *Anthology of American Folk Music* (Smithsonian Folkways, Washington DC 1997), originally published on LP records in 1952, 6-CD boxed set with book

53.5 R.D. Cohen, D.H. Samuelson: *Songs for Political Action: Folkmusic, Topical Songs, and the American Left* (Bear Family BCD, Hambergen 1996)

53.6 J. Place, R. Cohen: *The Best of Broadside, Anthems of the American Underground 1962–1988* (Smithsonian Folkways Recordings, Washington DC 2000), 5 compact discs and book

53.7 R. Meyer: "Fast folk, for the art of it". In: *Fast Folk, A Community of Singers and Songwriters*, ed. by R. Meyer, J. Hardy, J. Place (Smithsonian Folkways Recordings, Washington DC 2002) pp. 6–8, (compilers and annotators) 2 CDs and booklet

53.8 P. Seeger: *American Industrial Ballads* (Smithsonian Folkways, Washington DC 1992), notes by Irwin Silber, originally published by Folkways Recordings in 1956, 1 CD, Recordings 40058

53.9 M.K. Honey: *Sharecropper's Troubadour: John L. Handcox, the Southern Tenant Farmers' Union, and the African American Song*, Studies in Oral History (Palgrave, New York 2013)

53.10 W. Kaufman: *Woody Guthrie, American Radical* (Univ. Illinois Press, Urbana 2011)

53.11 M.A. Jackson: *Prophet Singer, The Voice and Vision of Woody Guthrie* (Univ. Press Mississippi, Jackson 2007)

53.12 N. Porterfield: *Last Cavalier, The Life and Times of John A. Lomax 1867–1948* (Univ. Illinois Press, Urbana 1996)

53.13 J. Zwed: *Alan Lomax the Man Who Recorded the World, A Biography* (Viking, New York 2010)

53.14 R.D. Cohen: *Alan Lomax, Selected Writings* (Routledge, New York 2003)

53.15 R.D. Cohen: *Alan Lomax, Assistant in Charge: The Library of Congress Letters* (Univ. Press Mississippi, Jackson 2010)

53.16 B. Rosenthal, S. Rosenthal (Eds.): *Pete Seeger in His Own Words* (Paradigm, Boulder 2012)

53.17 A. Pescatello: *Charles Seeger, A Life in American Music* (Univ. Pittsburgh Press, Pittsburgh 1992)

53.18 J. Tick: *Ruth Crawford Seeger, A Composer's Search for American Music* (Oxford Univ. Press, Oxford 1997)

53.19 D.K. Dunaway, M. Beer: *Singing Out, An Oral History of America's Folk Music Revivals* (Oxford Univ. Press, Oxford 2010)

53.20 J.C. Friedman: *The Routledge History of Social Protest in Popular Music* (Routledge, New York 2013)

53.21 R.A. Reuss, J.C. Reuss: *American Folk Music and Left-Wing Politics 1927–1957* (The Scarecrow Press, Lanham 2000)

53.22 B. Rosenthal, R. Flacks: *Playing for Change, Music and Musicians in the Service of Social Movements* (Paradigm, Boulder 2013)

53.23 W. Roy: *Reds, Whites, and Blues: Social Movements, Folk Music, and Race in the United States* (Princeton Univ. Press, Princeton 2013)

53.24 G. Bluestein: *The Voice of the Folk: Folklore and American Literary Theory* (Univ. Massachusetts, Boston 1972)

53.25 A. Green: *Only a Miner, Studies in Recorded Coal Mining Songs* (Univ. Illinois Press, Urbana 1972)

53.26 A. Green: Labor song: An ambiguous legacy, J. Folklore Res. **28**(2/3), 93–102 (1991)

53.27 C. Sandberg: *The American Songbag* (Harper Brace, New York 1927)

53.28 A. Seeger: Musics of struggle. In: *Program Book, 1990 Festival of American Folklife*, ed. by S. Peter (Smithsonian Institution, Washington DC 1990) pp. 56–69

53.29 P. Seeger: *The Incomplete Folksinger* (Simon Schuster, New York 1972), ed. by J.M. Schwartz

53.30 A. Seeger: *Why Suyá Sing: A Musical Anthropology of an Amazonian People* (Univ. Illinois Press, Urbana 2004)

53.31 OED: *Compact Edition of the Oxford English Dictionary* (Clarendon, Oxford 1971)

53.32 R. Sanjek: *American Popular Music and its Business, the First Four Hundred Years*, Vol. 1 (Oxford Univ. Press, Oxford 1988)

53.33 I. Silber: *Songs America Voted by* (Stackpole Books, Harrisburg 1971)

53.34 O. Brand: *Presidential Campaign Songs 1798–1996* (Smithsonian Folkways Recordings, Washington DC 1999), CD 45051

53.35 S. Gac: *Singing for Freedom: The Hutchinson Family Singers and the 19th Century Culture of Reform* (Yale Univ. Press, New Haven 2007)

53.36 B.K. Garman: *A Race of Singers, Whitman's Working-Class Hero from Guthrie to Springsteen*

(Univ. North Carolina Press, Chapel Hill 2000)
53.37 F.J. Child: *The English and Scottish Popular Ballads*, Vol. 8 (Mifflin, Houghton 1898)
53.38 B. Filene: *Romancing the Folk: Public Memory and American Roots Music* (Univ. North Carolina Press, Chapel Hill 2000)
53.39 A. Lomax, W. Guthrie, P. Seeger: *Hard Hitting Songs for Hard-Hit People* (Univ. Nebraska Press, Lincoln 1999)
53.40 A. Seeger: Lost Lineages and Neglected Peers: Ethnomusicologists Outside Academia, Ethnomusicology **50**(2), 314–235 (2006)
53.41 J.A. Lomax: *Cowboy Songs and Other Frontier Ballads* (Sturgis and Walton, New York 1910)
53.42 R.C. Seeger: Notes on the songs and the manner of singing. In: *The Music of American Folk Song and Selected Other Writings on American Folk Music*, ed. by L. Polansky, J. Tick (Univ. Rochester Press, Rochester 2001)
53.43 R.C. Seeger: "Musical preface". In: *Our Singing Country Folk Songs and Ballads*, ed. by J.A. Lomax (Dover, Mineola 2000)
53.44 F. Rosemont: *Joe Hill: The IWW & The Making of a Revolutionary Working Class Counterculture* (Charles H. Kerr, Chicago 2003)
53.45 L. Taylor: *"Introduction" to the CD: Don't Mourn – Organize! Songs of labor songwriter Joe Hill*, CD with liner notes (Smithsonian Folkways, Washington DC 1990), recordings SF 40026
53.46 J. Place: *"introduction" to Fast Folk: A Community of Singers and Songwriters* (Smithsonian Folkways Recordings, Washington DC 2002), 2 CD compilation with notes, SFW 40135
53.47 S. Dennisoff: *Great Day Coming! Folk Music and the American Left* (Univ. Illinois, PressUrbana 1971)
53.48 R. Cantwell: *When We Were Good: The Folk Revival* (Harvard Univ. Press, Cambridge 1996)
53.49 P. Seeger: *Pete Seeger: The Complete Bowdoin College Concert, 1960* (Smithsonian Folkways, Washington DC 2012), CD with liner notes, Recordings 40184
53.50 R.D. Cohen: *A History of Folk Music Festivals in the United States: Feasts of Musical Celebration* (Scarecrow, Lanham 2008)
53.51 ACA (American Camp Association): http://www.acacamps.org/media/aca-facts-trends consulted (2014)
53.52 E. Badeaux: "Singing for the summer – and getting paid for it", Sing Out! **8**(4), 23–25 (1959)
53.53 B. Dylan: *Folkways, The Original Vision, Songs of Woody Guthrie and Lead Belly* (Smithsonian Folkways, Washington DC 2005), Recordings SFW
53.54 K. Hankes: *"Music and Money and Hate"*, *Intelligence Report*, Vol. 156 (Southern Poverty Law Center, Montgomery 2014) pp. 34–36

About the Authors

Jakob Abeßer

Fraunhofer IDMT
Ilmenau, Germany
jakob.abesser@idmt.fraunhofer.de

Chapter F.41

Jakob Abeßer holds a degree in Computer Engineering (Dipl.-Ing.) from Ilmenau University of Technology. He is a postdoctoral researcher in the Semantic Music Technologies Group at Fraunhofer IDMT and obtained a PhD degree (Dr.-Ing.) in Media Technology from Ilmenau University of Technology in 2014. His research fields include automatic music transcription, symbolic music analysis, machine learning, and music instrument recognition.

Judit Angster

Fraunhofer Institute of Building Physics (IBP)
Acoustics
Stuttgart, Germany
judit.angster@ibp.fraunhofer.de

Chapter A.8

Judit Angster graduated in Physics at Eötvös University, receiving her PhD in 1991. Since 1992, she has held the position of scientist at Fraunhofer IBP in Stuttgart, where she is Head of the Musical Acoustics/Photoacoustics Research Group. She is coordinator of several European pipe organ acoustics research projects and President of the Technical Committee of Musical Acoustics (Germany, since 2008).

Simha Arom

Paris, France
simha.arom@gmail.com

Chapter G.49

Simha Arom is Emeritus Research Director of the French National Scientific Research Center (CNRS). His work deals with the systematic features of music in the polyphony of Central Africa and Georgia, as well as the temporal organization, modeling, and cognitive aspects of oral tradition music. He holds the *CNRS Silver Medal* and many International awards.

Rolf Bader Chapter A.10 For biographical profile, please see the section "About the Editor".

Ana M. Barbancho

Universidad de Málaga
ATIC Research Group, Dep. Ingeniería de Comunicaciones, ETSI Telecomunicación
Malaga, Spain
abp@ic.uma.es

Chapter F.42

Ana M. Barbancho holds a Telecommunications Engineering and PhD degree from the University of Málaga (UMA; 2000 and 2006, respectively), and a degree in Solfeo Teaching from Málaga Conservatoire of Music (2001). Since 2000, she has been with the Department of Communications Engineering, UMA. Her research interests include musical acoustics, digital signal processing, new educational methods, and mobile communications.

Isabel Barbancho

Universidad de Málaga
ATIC Research Group, Dep. Ingeniería de Comunicaciones, ETSI Telecomunicación
Malaga, Spain
ibp@ic.uma.es

Chapters F.42, F.45

Isabel Barbancho received her Telecommunications Engineering and PhD degrees from the University of Malaga (UMA; 1993 and 1998, respectively), and a degree in Piano Teaching from the Malaga Conservatoire of Music (1994). She is currently a Professor at the Department of Communications Engineering, UMA. Her research interests include musical acoustics, signal processing, multimedia applications, audio content analysis, and serious games.

Juan Pablo Bello

New York University
New York, USA
jpbello@nyu.edu

Chapter F.39

Juan Pablo Bello is Associate Professor of Music Technology and Electrical & Computer Engineering at New York University. He studied Electronic Engineering in Venezuela and the UK, specializing in signal processing, machine listening, and music information retrieval, topics in which he teaches and publishes regularly. He is recipient of an NSF CAREER award and a Fulbright fellowship.

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5

Stefan Bilbao

University of Edinburgh
Acoustics and Audio Group
Edinburgh, UK
s.bilbao@ed.ac.uk

Chapter B.19

Stefan Bilbao (BA Physics, Harvard, 1992; MSc, PhD Electrical Engineering, Stanford, 1996 and 2001, respectively) is currently a reader in the Acoustics and Audio Group at the University of Edinburgh, and was previously a lecturer at the Sonic Arts Research Centre, at Queen's University Belfast, and a research associate at the Stanford Space Telecommunications and Radioscience Laboratories.

David Borgo

University of California San Diego
La Jolla, USA
dborgo@ucsd.edu

Chapter G.52

David Borgo is Professor of Music and Integrative Studies at the University of California San Diego. He authored a book that won the Alan P. Merriam Prize in 2006 from the Society for Ethnomusicology. As a saxophonist, David has toured internationally and has released eight CDs and one DVD of original music.

Jonas Braasch

Rensselaer Polytechnic Institute
Troy, USA
braasj@rpi.edu

Chapter B.14

Jonas Braasch received a Master's Degree in Physics from the Technical University of Dortmund (1998), and two doctoral degrees from the University of Bochum in Electrical Engineering and Information Technology (2001) and Musicology (2004). He joined Rensselaer Polytechnic Institute in 2006 and is Associate Professor in the School of Architecture and Director of the Center for Cognition, Communication, and Culture.

Elvira Brattico

Aarhus University
Department of Clinical Medicine
Aarhus, Denmark
elvira.brattico@clin.au.dk

Chapter C.22

Elvira Brattico (PhD in Psychology, University of Helsinki, 2007) is Full Professor of Neuroscience, Music, and Aesthetics at the Center for Music in the Brain (MIB), Department of Clinical Medicine, Aarhus University and Royal Academy of Music, Aarhus/Aalborg, Denmark. Her interests cover auditory neuroscience, neuroaesthetics, and brain plasticity. She has published more than 100 papers and 10 book chapters.

Estefanía Cano

Fraunhofer IDMT
Ilmenau, Germany
cano@idmt.fraunhofer.de

Chapter F.41

Estefanía Cano holds degrees in Electronic Engineering (BSc), Music-Saxophone Performance (BA), Music Engineering (MSc), and Media Technology (Dr.-Ing.). In 2009, Estefanía joined the Fraunhofer Institute for Digital Media Technology IDMT, where she currently works as a postdoctoral researcher on various music information retrieval topics such as sound source separation, audio matching, automatic music transcription, and music education.

Amin Chaachoo

Tetouan-Asmir Center
Tetuán, Morocco
chaachooamin@gmail.com

Chapter F.43

Amin Chaachoo is a musician and musicologist, founder and director of the Tetouan-Asmir Center for Musicological Research and Preservation of the Musical Heritage and the first violin soloist of the Orchestra of the Tetouan Conservatory. He has authored two books that are considered as inflection points in the development of Arab-Andalusian music.

Marshall Chasin

Musicians' Clinics of Canada
Toronto, Canada
marshall.chasin@rogers.com

Chapter F.40

Dr. Marshall Chasin is an audiologist and the Director of Auditory Research at the Musicians' Clinics of Canada, Adjunct Professor at the University of Toronto (in Linguistics), and Associate Professor in the School of Communication Disorders and Sciences at the Western University. He is the author of over 200 articles and 7 books.

Jesús Corral García

Universidad de Málaga
Departamento de Ingeniería de Comunicaciones, ETSI Telecomunicación
Malaga, Spain
jcorral@ic.uma.es

Chapter F.45

Jesús Corral García received his MSc degree in Telecommunication Engineering at Universidad de Málaga (UMA), Spain in 2014. He has been a researcher of the Application of Information and Communications Technologies (ATIC) Research Group at the Department of Ingeniería de Comunicaciones (IC) of UMA since 2015. His research is focused on interactive audio file formats and speech recognition and its applications.

Lola L. Cuddy

Queen's University
Dept. of Psychology
Kingston, Canada
lola.cuddy@queensu.ca

Chapter C.23

Lola L. Cuddy is Professor Emerita, Department of Psychology, Queen's University in Kingston, Ontario, Director of the Music Cognition Laboratory, and general editor of *Music Perception.* Research interests include representation of tonal hierarchies, implicit learning of musical structures, and musical memory and dementia. In 2015, she received the life time achievement award from the Society of Music Perception and Cognition.

Phillippe Depalle

McGill University
Schulich School of Music
Montreal, Canada
philippe.depalle@mcgill.ca

Chapter B.17

Philippe Depalle is Associate Professor of Music Technology in the Schulich School of Music, McGill University. He received MS and PhD degrees in Applied Acoustics from the Université du Mans, Le Mans, France, in 1984 and 1991, respectively. His research interests include sound synthesis, sound processing, sound analysis, simulation of musical instruments, control of sound synthesis, and spectral analysis.

José-Miguel Díaz-Báñez

Escuela Superior de Ingenieros,
Universidad de Sevilla
Departamento de Matemática Aplicada II
Sevilla, Spain
dbanez@us.es

Chapter F.43

José-Miguel Díaz-Báñez received his PhD in Mathematics in 1998 and serves as Professor of Applied Mathematics at the University of Seville. His primary research area is computational geometry and optimization, mainly applied to problems in aerial robotics and music technology. He has authored more than 200 publications, including 53 journal papers. He is the coordinator of the COFLA Project (www.cofla-project.com).

Christian Dittmar

International Audio Laboratories Erlangen
Erlangen, Germany
christian.dittmar@audiolabs-erlangen.de

Chapter F.41

Christian Dittmar studied Electrical Engineering at Jena University of Applied Sciences and after graduation (2003) joined the Fraunhofer Institute for Digital Media Technology (Ilmenau). He is CEO and co-founder of Songquito UG (since 2012), a start-up company responsible for marketing the music education game Songs2See. He has been pursuing his PhD thesis at the International Audio Laboratories Erlangen since 2014.

Peter Driessen

University of Victoria
Dept. of Electrical and Computer Engineering
Victoria, Canada
peter@ece.uvic.ca

Chapter F.46

Peter Driessen is a Professor in Engineering with cross-appointments in Music and Computer Science. His teaching and research interests include music technology, sound recording, audio signal processing, and multimedia. He has over 100 publications and 14 patents, and holds research funding from NSERC, SSHRC, and Canada Council for the Arts.

Zhiyao Duan

University of Rochester
Dept. of Electrical and Computer Engineering
Rochester, USA
zhiyao.duan@rochester.edu

Chapter B.15

Zhiyao Duan is an Assistant Professor in the Department of Electrical and Computer Engineering at the University of Rochester. He received his PhD in Computer Science from Northwestern University in 2013. His research interest is in the area of computer audition, i.e., designing computational systems that are capable of analyzing and processing sounds, including music, speech, and environmental sounds.

Tuomas Eerola

Durham University
Dept. of Music
Durham, UK
tuomas.eerola@durham.ac.uk

Chapter C.29

Tuomas Eerola received his MA and PhD degrees from the University of Jyväskylä, Finland, in 1997 and 2003. He is a Professor of Music Cognition at Durham University, UK. His research interest lies within the field of music cognition and music psychology, including musical similarity, melodic expectations, perception of rhythm and timbre, and induction and perception of emotions.

Ricardo Eichmann

Deutsches Archäologisches Institut
Orient-Abteilung
Berlin, Germany
ricardo.eichmann@dainst.de

Chapter G.51

Ricardo Eichmann, Director of the Orient Department of the German Archaeological Institute, is a Near Eastern archaeologist. He received his Dr. phil. from Ruprecht Karls University Heidelberg in 1984 and was a Professor of Near Eastern archaeology at Eberhard Karls University Tübingen in 1995/1996. Besides archaeological field work, he has been co-organizing international conferences on music archaeology since 2000.

Benoit Fabre

Sorbonne Universités, UPMC Univ Paris 06
LAM – Institut d'Alembert
Paris, France
benoit.fabre@upmc.fr

Chapter A.7

Benoît Fabre is currently a Professor at University Pierre & Marie Curie, in Paris, France. He was trained both in Music (Hochschule für Musik in Freiburg, Germany) and Acoustics (PhD at Université du Maine, France). Since 1993, he has carried out research on the physics, the making, and the playing of musical instruments at LAM (Laboratory for Musical Acoustics).

Ichiro Fujinaga

McGill University
Schulich School of Music
Montreal, Canada
ichiro.fujinaga@mcgill.ca

Chapter B.16

Ichiro Fujinaga is an Associate Professor and the Associate Dean of Research and Administration at the Schulich School of Music at McGill University. He has degrees in Mathematics (BSc, Alberta), Music/Percussion (BMus, Alberta), Music Theory (MA, McGill), and Music Technology (PhD, McGill). He is the Principal Investigator of the Single Interface for Music Score Searching and Analysis (SIMSSA) project.

Aaron Gibbings

The University of Western Ontario, Natural Sciences Centre
The Brain and Mind Institute, Department of Psychology
London, Canada
jgrahn@uwo.ca

Chapter C.27

Aaron Gibbings is a PhD student in Dr. Jessica Grahn's Music and Neuroscience Lab at the Brain and Mind Institute. Aaron completed his BA in Psychology at King's University College, and his MSc at the University of Western Ontario. His doctoral research investigates how neural oscillations and entrainment affect beat and rhythm perception.

Joël Gilbert

Université du Maine – CNRS
Laboratoire d'Acoustique
Le Mans, France
joel.gilbert@univ-lemans.fr

Chapter A.7

Joël Gilbert is directeur de recherche CNRS, Laboratoire d'Acoustique de l'Université du Maine. He holds the Agrégation of Physics (ENS Fontenay aux Roses, 1985) and a PhD in Acoustics from Université du Maine (1991). Since 1991, he has carried out research on linear and nonlinear acoustics of internal flows, and physics of dynamical systems applied to reed and brass instruments.

Nicholas Giordano

Auburn University
Dept. of Physics
Auburn, USA
njg0003@auburn.edu

Chapter A.6

Nicholas Giordano joined Auburn University in 2013, where he is a Professor in the Department of Physics and Dean of the College of Sciences and Mathematics. Prior to Auburn, he was a member of the Faculty at Purdue University where he served as Head of the Department of Physics from 2007–2013. He has been an Alfred P. Sloan Foundation Fellow.

Rolf Inge Godøy

University of Olso
Dept. of Musicology
Oslo, Norway
r.i.godoy@imv.uio.no

Chapter E.35

Rolf Inge Godøy is Professor at the Department of Musicology, University of Oslo. His main interest is in phenomenological approaches to music theory, meaning taking our subjective experiences of music as the point of departure for music theory. This work has been expanded to include research on music-related body motion in performance and listening, using various conceptual and technological tools.

Emilia Gómez

Universitat Pempeu Fabra
Music Technology Group
Barcelona, Spain
emilia.gomez@upf.edu

Chapter F.43

Emilia Gómez is an Associate Professor (Serra-Húnter Fellow) at Universitat Pompeu Fabra in Barcelona. She holds a BEng (Universidad de Sevilla), MSc in Music Technology (IRCAM), and PhD in Computer Science (UPF). Her research deals with music information retrieval (MIR): pitch, melody, tonality, similarity, and semantics. She has published a large number of publications on these topics.

Francisco Gómez

Technical University of Madrid
Dept. of Applied Mathematics
Madrid, Spain
fmartin@etsisi.upm.es

Chapter F.43

Dr. Gómez obtained an MSc in Mathematics from Universidad Autónoma de Madrid. Later he obtained a PhD in Computer Science from the Technical University of Madrid. He does research on music information retrieval and computational music theory. His main areas of research are: measures of similarity, mathematical measures of rhythm complexity and syncopation, and automatic analysis of music traditions.

Jessica Grahn

The University of Western Ontario
The Brain and Mind Institute, Dept. of Psychology
London, Canada

Chapter C.27

Jessica Grahn is an Associate Professor at the Brain and Mind Institute and Department of Psychology at the University of Western Ontario (Western). She has degrees in Neuroscience and Piano Performance from Northwestern University and a PhD from Cambridge, England, in the Neuroscience of Music. She specializes in rhythm, movement, and cognition.

Sascha Grollmisch

Fraunhofer IDMT
Ilmenau, Germany
goh@idmt.fraunhofer.de

Chapter F.41

Sascha Grollmisch holds a degree in Media Technology Engineering (Dipl.-Ing.) from Ilmenau University of Technology. He is a software developer and researcher at the Semantic Music Technologies Group at Fraunhofer IDMT and CTO and co-founder of Songquito UG. His main research field is automatic music transcription and gamification of music education.

Simon Grondin

Université Laval
École de psychologie
Québec, Canada
simon.grondin@psy.ulaval.ca

Chapter C.21

Simon Grondin is a Professor of Psychology at Université Laval (Québec). His research interests are mainly focused on timing and time perception, rhythm, psychological time, psychophysics, cognitive neurosciences, and the relative age effect in sports. He is a former editor of the Canadian Journal of Experimental Psychology (2006–2009) and a former associate editor of Attention, Perception and Psychophysics (2006–2015).

Peter Grosche

Huawei Technologies Duesseldorf GmbH
München, Germany
peter.grosche@huawei.com

Chapter F.39

Peter Grosche studied Electrical Engineering and Information Technology at the Technical University of Munich (TUM) and obtained a PhD degree in Computer Science from Saarland University for his work in the field of music signal processing and music information retrieval. Currently, he is working on various research topics in audio signal processing at Huawei European Research Center in Munich.

Brian Hamilton

University of Edinburgh
Acoustics and Audio Group
Edinburgh, UK
b.hamilton-2@sms.ed.ac.uk

Chapter B.19

Brian Hamilton received BEng (Hons) and MEng degrees in Electrical Engineering from McGill University in Montreal, QC, Canada, in 2009 and 2012, respectively. He is currently a postdoctoral research associate in the Acoustics and Audio Group at the University of Edinburgh. His focus of research is finite difference and finite volume methods for 3D room acoustics simulations.

Andrew Hankinson

University of Oxford
Bodleian Libraries
Oxford, UK
andrew.hankinson@bodleian.ox.ac.uk

Chapter B.16

Andrew Hankinson received his PhD from McGill University in 2014. He is currently a Senior Software Engineer at the Bodleian Libraries, University of Oxford. His research interests focus on tools, technologies, and best-practices for building large-scale document image recognition systems for libraries, with a particular focus on optical music recognition.

Reginald Harrison

University of Edinburgh
Acoustics and Audio Group
Edinburgh, UK
s0916351@sms.ed.ac.uk

Chapter B.19

Reginald L. Harrison (BSc Physics and Music, University of Edinburgh, 2013) is currently a PhD student in the Acoustics and Audio Group at the University of Edinburgh. His research interests include time domain modelling of brass instruments for sound synthesis.

Emi Hasuo

Tokyo Denki University
School of Information Environment
Chiba, Japan

Chapter C.21

Emi Hasuo received her Bachelor's degree from Kunitachi College of Music in 2006, and her Master's and PhD degrees from Kyushu University in 2008 and 2011, respectively. She worked at Laval University as a postdoctoral fellow and is currently a research fellow of JSPS. Her research focuses mainly on understanding the basic mechanisms of time and rhythm perception.

Avraham Hirschberg

Veldhoven, The Netherlands

Chapter A.7

Avraham Hirschberg is Emeritus Professor at Technische Universiteit Eindhoven (TU/e) and Universiteit Twente. He holds an MSc from TUDelft and a PhD from TU/e on Plasma Physics (1981). He researched the thermodynamics of asphaltenes from 1981–1985 (Shell Research). From 1985 on, research on aeroacoustics of internal flows including pulse tubes, microperforated plates, human speech production, and wind instruments among others.

Neil S. Hockley

Bernafon AG
Bern, Switzerland
neho@bernafon.com

Chapter F.40

Neil S. Hockley is a Senior Audiologist in research and development at Bernafon AG in Bern, Switzerland. Prior to joining Bernafon in 2001, he worked concurrently in clinical and academic settings in the Niagara region of Ontario (Canada) (1994–2001). One of his professional interests is to improve hearing aid technology for amateur and professional musicians.

Alexander Refsum Jensenius

University of Oslo
Dept. of Musicology
Oslo, Norway
a.r.jensenius@imv.uio.no

Chapter E.38

Alexander Refsum Jensenius (BA, MA, MSc, PhD) is a music researcher and research musician working in the fields of embodied music cognition and new interfaces for musical expression. He studied at Oslo, Chalmers, UC Berkeley and McGill, and is working as an Associate Professor of Music Technology at the Department of Musicology, University of Oslo.

Wilfried Kausel

University of Music and Performing Arts, Vienna
Dept. of Musical Acoustics
Vienna, Austria
kausel@mdw.ac.at

Chapters A.2, A.3

Wilfried Kausel has been the Head of the Department of Music Acoustics at the University of Music and Performing Arts Vienna since 2010, becoming a full Professor in 2015. He received his tenure as an Associate Professor in 2003 and his PhD in 1999. He is a Fellow of the ASA and a member of technical committees in musical acoustics.

Christian Kehling

Neways Technologies
Erfurt, Germany
christian@k-ling.de

Chapter F.41

Christian Kehling is a Media Technology Engineer and passionate musician. He graduated in 2014 at the Ilmenau University of Technology. His diploma work dealt with instrument-specific coding with a strong focus on automated music transcription, semantic audio analysis, and instrument synthesis. Currently he is working as a research and development engineer and software developer for spatial audio systems.

Peter E. Keller

Western Sydney University
MARCS Institute for Brain, Behaviour and Development
Penrith, Australia
p.keller@uws.edu.au

Chapter C.28

Professor Peter Keller holds degrees in Music and Psychology from the University of New South Wales in Australia. Peter is leader of the 'Music Cognition and Action' research program in the MARCS Institute for Brain, Behaviour and Development at Western Sydney University. He conducts research aimed at understanding the behavioral and brain bases of human interaction in musical contexts.

Stefan Koelsch

University of Bergen
Bergen, Norway
koelsch@cbs.mpg.de

Chapter C.24

Stefan Koelsch is Professor for Biological Psychology and Music Psychology at the University of Bergen (Norway). He has Masters degrees in Music, Psychology and Sociology, and did his PhD and Habilitation at the Max Planck Institute for Cognitive Neuroscience, where he also led the Independent Junior Research Group "Neurocognition of Music" after being a postdoctoral fellow at Harvard Medical School.

Nadine Kroher

Escuela Superior de Ingenieros,
Universidad de Sevilla
Departamento de Matemática Aplicada II
Sevilla, Spain
nkroher@us.es

Chapter F.43

Nadine Kroher received her Audio Engineering degree from the University of Technology and the University of Music and Dramatic Arts Graz, Austria. She obtained an MSc in Sound and Music Computing at Universitat Pompeu Fabra. Currently, she is a PhD candidate at the University of Seville, where she works on automatic transcription and computational analysis of Flamenco music.

Panos Kudumakis

Queen Mary University of London
School of Electronic Engineering and Computer Science
London, UK
panos.kudumakis@eecs.qmul.ac.uk

Chapter F.45

Panos Kudumakis holds a PhD in DSP from King's College London. His research interests include music and media standardization. He was head of DRM at EMI-Central Research Labs before appointment in 2007 as Research Manager at qMedia, Queen Mary University of London. He currently serves as Head of the UK delegation of ISO/IEC MPEG committee, which he joined in 1998.

Tsuyoshi Kuroda

Shizuoka University
Faculty of Informatics
Hamamatsu, Japan
tkuroda@inf.shizuoka.ac.jp

Chapter C.21

Tsuyoshi Kuroda received his PhD in Design from Kyushu University, Fukuoka, Japan in 2010, and is currently working as a postdoctoral fellow in the Cognitive and Brain Science Laboratory, Shizuoka University, Hamamatsu, Japan. His research interests include auditory organization and temporal processing in perception and motor action. He has authored publications in scientific journals such as JEP:HPP and AP&P.

Marc Leman

Ghent University
IPEM – Musicology, Department of Art,
Music and Theatre Sciences
Ghent, Belgium
marc.leman@ugent.be

Chapters E.34, E.37

Marc Leman is Methusalem Research Professor in Systematic Musicology and Director of IPEM at Ghent University. He works on the epistemological and methodological foundations of embodied music interactions. He is Laureate of the FWO Excellence Award "Ernest-John Solvay".

Alexander Lerch

Georgia Institute of Technology
Center for Music Technology
Atlanta, USA
alexander.lerch@gatech.edu

Chapter F.44

Alexander Lerch studied Electrical Engineering at the TU Berlin and Tonmeister at the UdK Berlin. He received his PhD from the TU Berlin. In 2001, he co-founded the company zplane.development – a research-driven technology provider for the music industry. In 2013, he joined the Georgia Tech Center for Music Technology, where he works on music information retrieval and intelligent audio software.

Micheline Lesaffre

Ghent University
IPEM – Musicology, Department of Art,
Music and Theatre Sciences
Ghent, Belgium
micheline.lesaffre@ugent.be

Chapter E.36

Dr. Micheline Lesaffre is a researcher in systematic musicology at IPEM, the music research institute of Ghent University. Her research has a focus on user-oriented analysis, the usability of music tools, user experiences, and cross-disciplinary approaches. Recently, this work has been expanded to the domain of health and wellbeing, with a focus on person-centered approaches to embodied music-based interaction.

Jukka Louhivuori

University of Jyväskylä
Department of Music, Art and Culture
Studies
Jyväskylä, Finland
jukka.louhivuori@jyu.fi

Chapter G.47

Prof. Louhivuori has held the post of President of the Finnish Society for Music Education (FiSME), the Finnish Society for Musicology, and the European Society for the Cognitive Sciences of Music (ESCOM). He is specialized in cross-cultural music cognition, ethnomusicology, and music and well-being. Recently, he has developed new musical interfaces for music pedagogical purposes.

Håkan Lundström

Lund University
Inter Arts Center
Lund, Sweden
hakan.lundstrom@kanslik.lu.se

Chapter G.50

Håkan Lundström is a Senior Professor at the Inter Arts Center, Malmö Faculty of Fine and Performing Arts, Lund University, Sweden. He is an ethnomusicologist who has worked on Kammu (Khmu) music, particularly singing traditions. He has conducted research in Japanese and Alaskan Native American music and has led a research program concerning documentation and revitalization of ethnic minority music.

Pieter-Jan Maes

Ghent University
IPEM – Musicology, Department of Art,
Music and Theatre Sciences
Ghent, Belgium
pieterjan.maes@ugent.be

Chapters E.34, E.37

Dr. Pieter-Jan Maes is working as postdoctoral researcher at IPEM, Ghent University. His research focuses on the role of body movement and auditory feedback in musical timing and entrainment. On the basis of empirical findings, he is developing the theory of embodied music cognition, as well as practical music applications.

András Miklós

Steinbeis Transfer Center Applied
Acoustics
Stuttgart, Germany
akustikoptik@t-online.de

Chapter A.8

András Miklós graduated in Physics at Eötvös University Budapest (Hungary), and received his PhD in 1990. From 1992–2005, he was a Senior Scientist at University Heidelberg. Since 2003, he has been a Director of the Steinbeis Transfer Center Applied Acoustics, and a scientific adviser at Fraunhofer IBP in Stuttgart. His main research fields are photoacoustics, photothermal phenomena, and musical acoustics.

Emilio Molina

Universidad de Málaga
Departamento de Ingeniería de Comunicaciones, ETSI Telecomunicación
Malaga, Spain
emm@ic.uma.es

Chapter F.42

Emilio Molina received his degree in Telecommunications Engineering from the University of Málaga in 2011, and his MSc in Sound and Music Computing from the Universitat Pompeu Fabra (Barcelona) in 2012. He then worked as a researcher at the ATIC Group (University of Málaga) in the field of singing voice analysis. Currently, he works as a research engineer for BMAT.

Thomas Moore

Rollins College
Dept. of Physics
Winter Park, USA
tmoore@rollins.edu

Chapter A.5

Thomas Moore is the Archibald Granville Bush Professor of Science and a Professor of Physics at Rollins College. His research interests lie in the physics of musical instruments and the application of optical methods to the study of vibrations. He is a Fellow of the Acoustical Society of America.

Joaquin Mora

Escuela Superior de Ingenieros,
Universidad de Sevilla
Departamento de Matemática Aplicada II
Sevilla, Spain
mora@us.es

Chapter F.43

Dr. Mora is a Professor at the University of Seville and in the Flamenco doctoral program. He has directed various doctoral theses, authored numerous articles, and given various talks on Flamenco and Spanish folk music. He is a member of the COFLA research group, where he works on computational analysis of Flamenco music.

Robert Mores

University of Applied Sciences
Faculty of Design, Media & Information
Hamburg, Germany
robert.mores@haw-hamburg.de

Chapter B.12

Robert Mores hold a Diploma in Electrical Engineering (1988) and a PhD in Computer Science (1994). From 1994 to 2001, he was with Philips Semiconductors. He holds some ten patents on in-vehicle communication systems. He is a Professor at the University of Applied Sciences Hamburg (since 2001), and teaches telecommunications and digital signal processing. His research focus is musical acoustics.

Andrew C. Morrison

Joliet Junior College
Dept. of Natural Sciences
Joliet, USA
amorriso@jjc.edu

Chapter A.9

Andrew Morrison is an Associate Professor of Physics at Joliet Junior College. He earned a BS in Physics from the University of Northern Iowa and a PhD in Physics from Northern Illinois University. He has taught at Illinois Wesleyan University, Northwestern University, and DePaul University. His research interests focus primarily on percussion acoustics, especially the Caribbean steelpan and related instruments.

Meinard Müller

International Audio Laboratories Erlangen
Erlangen, Germany
meinard.mueller@audiolabs-erlangen.de

Chapter F.39

Meinard Müller studied Mathematics and Computer Science at Bonn University, Germany. Since September 2012, he has held a Professorship for Semantic Audio Processing at the International Audio Laboratories Erlangen, Germany. His research interests include music processing, audio signal processing, and music information retrieval, which are also reflected in his recent textbook (Springer, www.music-processing.de).

Yoshitaka Nakajima

Kyushu University
Dept. of Human Science
Fukuoka, Japan
nakajima@design.kyushu-u.ac.jp

Chapter C.21

Yoshitaka Nakajima is a Distinguished Professor in the Faculty of Design, Kyushu University. He graduated in Psychology from the University of Tokyo in 1978, and received a PhD in Design at the Kyushu Institute of Design in 1999. His research fields include auditory perception, time perception, and language universals.

Tram Nguyen

The University of Western Ontario
The Brain and Mind Institute, Dept. of Psychology
London, Canada
tnguye95@uwo.ca

Chapter C.27

Tram Nguyen is currently a PhD student in the Music and Neuroscience Lab, directed by Dr. Jessica Grahn, at the University of Western Ontario's Brain and Mind Institute. Tram has completed BS and MSc degrees in Psychology at Western. Her doctoral research examines individual differences in beat perception and production, particularly between musicians and dancers.

Luc Nijs

Ghent University
IPEM – Musicology, Department of Art, Music and Theatre Sciences
Ghent, Belgium
luc.nijs@ugent.be

Chapters E.34, E.37

Luc Nijs is a postdoc at IPEM (Ghent University, BE), guest lecturer at the Royal Conservatory The Hague (NL), and clarinet teacher at "de Kunstbrug" (Ghent, BE). He holds a PhD in Arts Sciences and MA's in Music Performance (clarinet) and Philosophy. His research focuses on the musician–instrument relationship and on the role of movement in instrumental learning.

Giacomo Novembre

University College London
London, UK
giacomonovembre@gmail.com

Chapter C.28

Giacomo Novembre holds degrees in Cognitive Neuroscience (MSc, PhD) and Philosophy (BA). His research utilizes music as a model to explore the neurocognitive mechanisms that underpin action-perception coupling in the human brain. He has held lectureship and research associate positions at the MARCS Institute for Brain, Behaviour and Development (Western Sydney University, Australia) and at University College London.

Chiara Olcese

University of Ferrara
Dept. of Life Sciences and Biotechnology
Treviso, Italy
chiara.olcese@student.unife.it

Chapter C.22

Chiara Olcese holds a PhD in Biochemistry, Molecular Biology and Biotechnologies (2016). During her Master's studies she investigated sound waves' effects on human hormones and the mismatch negativity. Her postdoctoral research focuses on the molecular basis of Primary ciliary dyskinesia, using Next-generation sequencing and bioinformatics methods.

Bryan Pardo

Northwestern University
Ford Engineering Design Center
Evanston, USA
pardo@northwestern.edu

Chapter B.15

Bryan Pardo is an Associate Professor in the Northwestern University Department of Electrical Engineering and Computer Science. Prof. Pardo received an MMus in Jazz Studies in 2001 and a PhD in Computer Science in 2005, both from the University of Michigan. He has authored over 80 peer-reviewed publications and performs throughout the United States on saxophone and clarinet.

Marcus Pearce

Queen Mary University of London
School of Electronic Engineering and Computer Science
London, UK
marcus.pearce@qmul.ac.uk

Chapters C.25, C.26

Educated in Experimental Psychology at Oxford and Artificial Intelligence at Edinburgh, Marcus Pearce is Senior Lecturer in Sound and Music Processing at Queen Mary University of London. His research covers computational, psychological and neuroscientific aspects of music cognition, with a particular focus on dynamic, predictive processing of melodic, rhythmic and harmonic structure, and its impact on aesthetic experience of music.

Florian Pfeifle

University of Hamburg
Institute of Systematic Musicology
Hamburg, Germany
florian.pfeifle@uni-hamburg.de

Chapter B.20

Florian Pfeifle received his MA in Systematic Musicology and Electrical Engineering in 2010 and his PhD in 2014. His current research is concerned with real-time grand piano physical modeling on FPGAs. Additionally, he is interested in instrument acoustics, scientific computation, and jazz music research. Until recently, he worked as a musician/producer, being awarded gold records for two of his works.

Laurent Pugin

Swiss RISM Office
Bern, Switzerland
laurent.pugin@rism-ch.org

Chapter B.16

Laurent Pugin is trained both as a musicologist and a computer scientist. He obtained his PhD at the University of Geneva, Switzerland. He has been co-Director of the Swiss Office of the RISM (Répertoire International des Sources Musicales) since 2009. He is a member of the Board of the Music Encoding Initiative (MEI) and teaches Digital Humanities at the University of Bern.

Zafar Rafii

Gracenote
Emeryville, USA
zrafii@gracenote.com

Chapter B.15

Zafar Rafii received a PhD in Electrical Engineering and Computer Science from Northwestern University in 2014. He is currently a research engineer at Gracenote in the US. He also worked at Audionamix in France from 2007 to 2008. His research interests include source separation and audio identification for music signals.

Martin Rohrmeier

TU Dresden
Institute of Art and Music
Dresden, Germany
martin.rohrmeier@tu-dresden.de

Chapters C.25, C.26

Professor Martin Rohrmeier studied Philosophy, Musicology, and Mathematics at the University of Bonn. He continued with an MPhil and PhD in Empirical Musicology and Music Cognition at the University of Cambridge, under the supervision of Prof. Ian Cross. His main research interests are music perception and cognition, musical syntax, implicit learning, computational modelling, and cognitive links between music and language.

Carles Roig

Universidad de Málaga
ATIC Research Group, Dep. Ingeniería de Comunicaciones, ETSI Telecomunicación
Malaga, Spain
carles@ic.uma.es

Chapter F.42

Carles Roig received his degree in Telecommunication Engineering from Universidad Politécnica de Valencia in 2010, and his MSc in Sound and Music Computing from Universidad Pompeu Fabra (Barcelona) in 2011. Then, he worked as a researcher at ATIC Group (Universidad de Málaga) in the field of automatic composition of music. Currently, he is a PhD candidate at Universidad de Valladolid.

Thomas D. Rossing

Stanford University
Dept. of Music
Stanford, USA
rossing@ccrma.stanford.edu

Chapter A.9

Thomas Rossing received a BA from Luther College, and MS and PhD degrees in Physics from Iowa State University. He was Professor of Physics at St. Olaf College for 14 years. Since 1971, he has been a Professor of Physics at Northern Illinois University. He is presently a Visiting Professor of Music at Stanford University. His areas of research have included musical acoustics, psychoacoustics, speech and singing, and vibration analysis.

Mark Sandler

Queen Mary University of London
School of Electronic Engineering and Computer Science
London, UK
mark.sandler@qmul.ac.uk

Chapter F.45

Mark Sandler received his BSc and PhD degrees from the University of Essex, UK, in 1978 and 1984, respectively. He is a Professor of Signal Processing at Queen Mary University of London, London, UK, where he founded the Centre for Digital Music. He has published over 400 papers in journals and conferences and supervised over 30 PhD students.

Gary Scavone

McGill University
Music Research, Schulich School of Music
Montreal, Canada
gary@music.mcgill.ca

Chapter B.13

Dr. Gary Scavone is an Associate Professor of Music Technology at McGill University, where he directs the Computational Acoustic Modeling Laboratory (CAML). He received PhD and Master of Science degrees at the Center for Computer Research in Music and Acoustics (CCRMA) at Stanford University. His research includes acoustic modeling, analysis and synthesis of musical systems, and sound synthesis software development.

Albrecht Schneider

University of Hamburg
Institute of Systematic Musicology
Hamburg, Germany
aschneid@uni-hamburg.de

Chapters 1, D.30, D.31, D.32, D.33

Albrecht Schneider graduated from the University of Bonn. He worked as a Professor of Systematic Musicology in the Institute of Musicology, University of Hamburg (1981–2011). He also taught as Visiting Professor in the Department of Ethnomusicology and Systematic Musicology of UCLA. He is the author, co-author, editor, and co-editor of a number of books and many other works.

Katrin Schulze

Heidelberg University
Dept. of Clinical Psychology and Psychotherapy, Institute of Psychology
Heidelberg, Germany
katrin.schulze@psychologie.uni-heidelberg.de

Chapter C.24

Dr. Katrin Schulze is a research associate working at Heidelberg University, Germany. Her research interests include auditory memory and emotion regulation. She graduated in Psychology from Magdeburg University and finished her PhD at the Max Planck Institute for Human Cognitive and Brain Sciences/Leipzig University, Germany. Further stations include BIDMC/Harvard Medical School (US) and Claude Bernard University Lyon 1 (France).

Anthony Seeger

University of California Los Angeles (UCLA)
Dept. of Ethnomusicology
Annapolis, USA
aseeger@arts.ucla.edu

Chapter G.53

Anthony Seeger is an anthropologist, ethnomusicologist, audio-visual archivist, and musician. He is Distinguished Professor of Ethnomusicology, Emeritus, at UCLA and Director Emeritus of Smithsonian Folkways Recordings in Washington, DC. He is the author of three books, co-editor of three others, and has published over 120 other works. As a musician he was active during the US folk music revival.

Mohamed Sordo

University of Miami
Center for Computational Science
Coral Gables, USA
msordo@miami.edu

Chapter F.43

Mohamed Sordo is a postdoctoral researcher at the University of Miami, USA. He obtained his BEng, MSc, and PhD degrees at the Universitat Pompeu Fabra (Barcelona, Spain) in 2006, 2007, and 2012, respectively. Mohamed's research areas involve music text mining, music information retrieval, and machine learning. Mohamed has authored over 25 scientific articles in international journals and conferences.

Lorenzo J. Tardón

Universidad de Málaga
Departamento de Ingeniería de Comunicaciones, ETSI Telecomunicación
Malaga, Spain
lorenzo@ic.uma.es

Chapters F.42, F.45

Lorenzo J. Tardón holds a degree in Telecommunications Engineering (1995) and a PhD (1999). Since 1999, he has been with the Department of Communications Engineering, Universidad de Málaga. He is Head of the Application of Information and Communications Technologies (ATIC) Research Group. His research interests include serious games, audio signal processing, digital image processing, and pattern analysis and recognition.

Mari Tervaniemi

University of Helsinki
Cicero Learning and Cognitive Brain Research Unit
Helsinki, Finland
mari.tervaniemi@helsinki.fi

Chapter C.22

Research Director Mari Tervaniemi is a well-known expert in neurosciences of music. She has published over 150 empirical papers and reviews in peer-reviewed international journals. Her research topics cover auditory neurocognition, musical expertise, and music emotions. Moreover, she has been in a key position to initiate projects to determine the effectiveness of music therapy in stroke, depression, and dementia patients.

Leslie Tilley

Massachusetts Institute of Technology
Cambridge, USA
tilley@mit.edu

Chapter G.48

Leslie Tilley is an Ethnomusicologist who specializes in analytical approaches to world music, with particular focus on Bali, Indonesia. Her current research examines collaborative improvisation in the Balinese drum genre *kendang arja*, and seeks to discover the unspoken, unconscious constraints guiding improvisation. Leslie is an Assistant Professor of Music at Massachusetts Institute of Technology.

Alberto Torin

University of Edinburgh
Acoustics and Audio Group
Edinburgh, UK
s1164558@sms.ed.ac.uk

Chapter B.19

Alberto Torin (BA and MSc Physics, University of Padova, 2007 and 2011, respectively; PhD, Music, University of Edinburgh, 2016) is currently a postdoctoral research associate in the Acoustics and Audio Group at the University of Edinburgh. He is a member of the GMAAT project and his research topics include the numerical simulation of microphone components.

George Tzanetakis

University of Victoria
Dept. of Computer Science
Victoria, Canada
gtzan@cs.uvic.ca

Chapter F.46

George Tzanetakis is an Associate Professor in the Department of Computer Science with cross-listed appointments in ECE and Music at the University of Victoria. He is Canada Research Chair in Computer Analysis and Audio and Music. In 2011, he was Visiting Faculty at Google Research. His research focus is audio content analysis with a focus on music information retrieval.

Edith Van Dyck

Ghent University
IPEM – Musicology, Department of Art, Music and Theatre Sciences
Ghent, Belgium
edith.vandyck@ugent.be

Chapter E.34

Dr. Edith Van Dyck holds a PhD in Musicology from IPEM and currently works as a postdoctoral researcher at the same institute. Since 2009, she has been focusing on action-perception coupling in musical interaction and published several papers regarding the influence of music features and human emotions on music-induced movement (e.g., dance, running, and walking).

Doug Van Nort

York University
Computational Arts and Theatre & Performance Studies
Toronto, Canada
dvnt.sea@gmail.com

Chapter B.17

Doug Van Nort is Assistant Professor and Canada Research Chair in Digital Performance in the School of the Arts, Media, Performance & Design (AMPD) at York University. Van Nort holds degrees in Mathematics (MA), Electronic Arts (MFA), and Music Technology (PhD). His research integrates music, multimodal perception, improvisation, and computation. He is the founding director of DisPerSion Lab (York University).

Michael Vorländer

RWTH Aachen University
Institute of Technical Acoustics
Aachen, Germany
mvo@akustik.rwth-aachen.de

Chapter A.11

Michael Vorländer is a Professor at RWTH Aachen University, Germany. His research focus is auralization and acoustic virtual reality in its various applications in psychoacoustics, room and building acoustics, automotive and noise control. Michael Vorländer has published 1 book and 12 book chapters, 81 articles in peer-reviewed journals, more than 300 conference papers, and 17 plenary lectures at international conferences.

Chris Waltham

University of British Columbia
Dept. of Physics & Astronomy
Vancouver, Canada
cew@phas.ubc.ca

Chapter A.4

Before musical acoustics, Chris Waltham spent 20 years working on the Sudbury Neutrino Observatory project. In the last dozen years, he has worked on the acoustics of string instrument soundboxes, specializing in harps and qins. He is an amateur luthier, and has made several harps and violins, playing one of the latter in a Vancouver community orchestra.

Ron Weiss

Google Inc.
New York, USA
ronw@google.com

Chapter F.39

Ron Weiss is a software engineer at Google, where he has worked on content-based audio analysis, recommender systems for music, and noise robust speech recognition. Ron completed his PhD in Electrical Engineering at Columbia University in 2009. From 2009 to 2010, he was a postdoctoral researcher in the Music and Audio Research Laboratory at New York University.

Victoria Williamson

University of Sheffield
Sheffield, UK
v.williamson@sheffield.ac.uk

Chapter C.24

Dr. Victoria Williamson is the Vice Chancellor's Fellow for Music at the University of Sheffield, UK, and Director of the Music & Wellbeing research unit. This unit focuses on applied music psychology, the impact of music on everyday and extraordinary challenges to wellbeing. Dr. Williamson's special interest within this remit is the psychology of musical memory.

Shigeru Yoshikawa

Dazaifu, Japan
shig@lib.bbiq.jp

Chapter A.4

After graduating from the Physics Department of Nagoya University (1974), he began acoustical research on organ pipes. He worked for Technical R&D Institute of Defense Ministry in 1980 and investigated underwater organ pipes, while personally studying musical acoustics. He was a Professor of Musical Instrument Acoustics at the Graduate School of Design, Kyushu University from 1998 and retired in 2015.

Tim Ziemer

University of Hamburg
Institute of Systematic Musicology
Hamburg, Germany
tim.ziemer@uni-hamburg.de

Chapter B.18

Dr. Tim Ziemer is a musicologist working in applied psychoacoustics for wave field synthesis and music information retrieval at the University of Hamburg. His research interests include measurement, reconstruction, and perception of radiation characteristics of musical instruments. He is a freelance author for a computer magazine and has a background in music production, room and building acoustics, among other areas.

Detailed Contents

Part E Music Embodiment

Part F Music and Media

Subject Index

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z