

Michael Vorländer

Auralization

Fundamentals of Acoustics, Modelling,
Simulation, Algorithms and Acoustic
Virtual Reality

Second Edition



RWTHedition

Series Editor

RWTH Aachen University, Aachen, Germany

More information about this series at <http://www.springer.com/series/7858>

The ASA Press

ASA Press, which represents a collaboration between the Acoustical Society of America and Springer Nature, is dedicated to encouraging the publication of important new books as well as the distribution of classic titles in acoustics. These titles, published under a dual ASA Press/Springer imprint, are intended to reflect the full range of research in acoustics. ASA Press titles can include all types of books that Springer publishes, and may appear in any appropriate Springer book series.

Editorial Board

Mark F. Hamilton (Chair), University of Texas at Austin

James Cottingham, Coe College

Timothy F. Duda, Woods Hole Oceanographic Institution

Robin Glosemeyer Petrone, Threshold Acoustics

William M. Hartmann (Ex Officio), Michigan State University

Darlene R. Ketten, Boston University

James F. Lynch (Ex Officio), Woods Hole Oceanographic Institution

Philip L. Marston, Washington State University

Arthur N. Popper (Ex Officio), University of Maryland

Christine H. Shadle, Haskins Laboratories

G. Christopher Stecker, Boys Town National Research Hospital

Stephen C. Thompson, The Pennsylvania State University

Ning Xiang, Rensselaer Polytechnic Institute



Michael Vorländer

Auralization

Fundamentals of Acoustics, Modelling,
Simulation, Algorithms and Acoustic
Virtual Reality

Second Edition



Michael Vorländer
Institute of Technical Acoustics
RWTH Aachen University
Aachen, Germany

Jointly published with ASA Press

ISSN 1865-0899

ISSN 1865-0902 (electronic)

RWTHedition

ISBN 978-3-030-51201-9

ISBN 978-3-030-51202-6 (eBook)

<https://doi.org/10.1007/978-3-030-51202-6>

© Springer Nature Switzerland AG 2020

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

The Acoustical Society of America

On 27 December 1928 a group of scientists and engineers met at Bell Telephone Laboratories in New York City to discuss organizing a society dedicated to the field of acoustics. Plans developed rapidly, and the Acoustical Society of America (ASA) held its first meeting on 10–11 May 1929 with a charter membership of about 450. Today, ASA has a worldwide membership of about 7000.

The scope of this new society incorporated a broad range of technical areas that continues to be reflected in ASA's present-day endeavors. Today, ASA serves the interests of its members and the acoustics community in all branches of acoustics, both theoretical and applied. To achieve this goal, ASA has established Technical Committees charged with keeping abreast of the developments and needs of membership in specialized fields, as well as identifying new ones as they develop.

The Technical Committees include acoustical oceanography, animal bioacoustics, architectural acoustics, biomedical acoustics, engineering acoustics, musical acoustics, noise, physical acoustics, psychological and physiological acoustics, signal processing in acoustics, speech communication, structural acoustics and vibration, and underwater acoustics. This diversity is one of the Society's unique and strongest assets since it so strongly fosters and encourages cross-disciplinary learning, collaboration, and interactions.

ASA publications and meetings incorporate the diversity of these Technical Committees. In particular, publications play a major role in the Society. *The Journal of the Acoustical Society of America* (JASA) includes contributed papers and patent reviews. *JASA Express Letters* (JASA-EL) and *Proceedings of Meetings on Acoustics* (POMA) are online, open-access publications, offering rapid publication. *Acoustics Today*, published quarterly, is a popular open-access magazine. Other key features of ASA's publishing program include books, reprints of classic acoustics texts, and videos. ASA's biannual meetings offer opportunities for attendees to share information, with strong support throughout the career continuum, from students to retirees. Meetings incorporate many opportunities for professional and social interactions, and attendees find the personal contacts a rewarding experience. These

experiences result in building a robust network of fellow scientists and engineers, many of whom become lifelong friends and colleagues.

From the Society's inception, members recognized the importance of developing acoustical standards with a focus on terminology, measurement procedures, and criteria for determining the effects of noise and vibration. The ASA Standards Program serves as the Secretariat for four American National Standards Institute Committees and provides administrative support for several international standards committees.

Throughout its history to present day, ASA's strength resides in attracting the interest and commitment of scholars devoted to promoting the knowledge and practical applications of acoustics. The unselfish activity of these individuals in the development of the Society is largely responsible for ASA's growth and present stature.

Preface and Introduction

Auralization has been in a rapid development process of integrating up-to-date simulation algorithms, sound synthesis and signal processing tools. Merging efforts in various fields in science and technology created opportunities for expanding into a multitude of applications of auralization. Advances in two fields had particularly large impacts: Immersive Spatial Audio is one of the biggest drivers in audio engineering. Virtual Reality technology is a megatrend which expands into the consumer market, due to the availability of head-mounted displays at relatively reasonable cost. Both technology trends entailed large efforts in understanding better and implementing solutions for personalized audio systems. It seems that binaural technology receives a revival, although physical, technological and perceptual foundations have been known for decades.

Current research in hearing, psychoacoustics and otolaryngology increasingly involves complex acoustic simulation and 3D audio, in order to overcome the limitations of laboratory tests with insufficient ecological validity. The field of auditory research in itself has an enormous complexity in the interdisciplinary connection of acoustics with psychoacoustics, audiology, and hearing systems technology. Therefore, quality requirements are set that differ significantly from those for applications in home cinema or games. Not just 3D “effects” or auditory illusions have to be achieved but realistic sound field or sound signal stimuli, best embedded into a multimodal setting. If sound perception is the focus of the investigation, the simulation and auralization systems shall create authentic, or at least plausible, sounds.

In music production, spatial audio is still a very big issue. The music format standards will be further developed towards the application of spherical microphone arrays and object-oriented coding. Therefore, headphone technology experiences not just a revival but a drastic extension regarding new dimensions of individualized high-fidelity solutions.

Thinking about the creation of virtual sounds in acoustic simulation, new methods and algorithms were recently presented that include more complexity in the models and in the simulation of the often neglected wave effects. Furthermore,

acoustic simulation and auralization in particular are now applied to outdoor sound as well as to classical indoor sound in performance spaces. This sets new requirements on source characterization. Outdoor sound sources (vehicles, trains, etc.) are typically large, much larger than musical instruments or humans. They require consideration of near-field effects apart from the power and directivity.

The second edition of this book presents updated and new content in several chapters. In particular, interactive virtual acoustics environments with moving sources and spatial audio technology are discussed in more depth. Some errors and misprints were corrected. Thanks to all who submitted their observations!

The new edition of “Auralization”, like the last, emphasizes that acoustics – the science of sound – is a very broad discipline. It is intertwined with many other natural and human sciences and electrical, mechanical and civil engineering. The essence of acoustics is “sound”: waves in fluid or elastic media, their generation, transmission, reception, cognition and evaluation. Acoustics thus involves aspects of physics, engineering and psychology. Ultrasound applied in physical acoustics, medical diagnosis or material testing is an example where acoustics meets several other disciplines, such as solid-state physics and image processing for medical diagnosis or for material testing. Therefore, it is in the nature of acoustics to deal with various methodologies. The physical side, of course, is best understood by applying linear and nonlinear wave theory in a straightforward way and especially in academic case studies where analytic calculations are possible. In engineering, more complex geometries or structures must often be considered and simplifications and approximations come into play, the limits of which stem from another area: psycho-acoustics. Physical data is examined, evaluated and condensed to give meaningful information on the particular characteristics of acoustic impression perceived by humans. This description is indeed typical, although not generally applicable to all acoustical sciences.

Scientific methods used in acoustics are therefore based on mathematically well-grounded wave theory, on experimental methods and usage of high-standard instrumentation and analysis software. Today, theoretical analytical calculations are increasingly augmented by simulation tools. These tools are the result of tremendous progress in numerical mathematics. Either geometrical methods, similar to those used in computer graphics, or wave-based methods similar to those applied in radar, microwave propagation or other electromagnetic waves for mobile communication, for instance, are used. Numerical methods of field problems also have great impact on scientific progress and innovation in engineering in mechanical engineering, heat conduction, flow dynamics and climate models.

Prediction of the acoustic behaviour of components or systems is called “modelling”. Modelling acoustics is the everyday task of acoustic engineering. The areas of activity for acoustic engineers are extremely diverse. Accordingly, methods and tools of acoustic engineering are very different in complexity and accuracy. One reason for this diversity is the necessity for a pragmatic approach to reduce the acoustical problem to a simple, one-dimensional scale of sound and noise levels expressed in global decibels, which are linked to legal requirements, norms or other

Table 1 Examples of quantities for applied acoustics

Problem	Quantity	Unit	Goal
Noise emission	L_w , directivity	dB, dB(A)	Noise limits for the source
Noise immission	L_{den}	dB(A)	Protection against noise (urban, traffic) at receiver
Sound insulation	D, TL	dB	Protection against excessive noise
Sound insulation	R, D, TL	dB	Protection against noise from neighbours
Auditorium acoustics	T, EDT, G, C_{80}	s, dB	“Good acoustics” for musical performances
Auditorium acoustics PA systems Communication systems	STI, AICons	%	Speech Intelligibility
Product sound quality	Loudness, roughness, tonality	sone, asper	Acoustic comfort

regulations of sound and noise effects. Another aspect might be historical development of international standards¹ in acoustics, describing acoustics in relation to easily measurable quantities like dB(A) levels, sound level differences or any weighted level number of acoustic absorption, attenuation or insulation. Engineering models of acoustics are often described in standards, as listed in Table 1. They are related to real-life problems of everyday life by which people are affected, entertained, informed or disturbed by acoustical phenomena.

If the acoustic behaviour of a component or a system is to be predicted by rather limited effort, we talk about “modelling”. The result is a numerical quantity or a set of quantities. Numerical quantities, however, have a clearly limited descriptive meaning. Imagine you have to explain a painting to a person who does not see it. The verbal characterization will be based on descriptors like size, colour and brightness and maybe on information about objects or resolution of details. A painting reproduced according to this verbal or numerical description will never be exactly like the original. The reproduction will contain a lot of subjective interpretations, even if the descriptors are “objective parameters”. How much easier and unambiguous would it be if we just looked at the painting instead of discussing its parameters! When it comes to acoustic problems one might ask why we discuss numbers describing the character of sound instead of listening to the sound and evaluating its loudness, timbre, roughness, sharpness, character or quality directly.

If the behaviour of an acoustic object or system is described in a more complex way including the creation of acoustic signals in time or frequency domain, we talk about “simulation” and “auralization”.

¹International Standard Organization (ISO).

Auralization is the technique to create audible sound files from numerical (simulated, measured or synthesized) data.²

Any sound, noise, music or in general any signal generated, transmitted, radiated and perceived can more precisely be interpreted and compared by people if it is made audible instead of discussing “levels in frequency bands”, “single number quantities” or “dB(A)”.

Perception of sound signals has multiple dimensions, some of which are listed as follows: type of sound generation, direction of the event, movement of the source, listener movement and outdoor or indoor environment (kind, shape and size). The auralization must cover all relevant cognitive aspects of the specific case.

An appropriately authentic model of the sound and vibration field is required to allow simulations of the psychoacoustically relevant features. Sound radiation models must therefore represent at least the correct sound propagation constant (distance law) and directivity. In complex coupled systems of sound and vibration, a further problem is the identification of relevant signal paths or vibration patterns, the degrees of freedom of motion in structural paths and the definition of interface line or planes between distributed velocity/pressure interaction effects or impedance coupling. In practice, the complexity of the vibro-acoustic problem might be very challenging, maybe too challenging to be solved on the basis of tools and algorithms presented in this book. But even then the concept of auralization can be a valuable source of information and motivation for research and development in acoustic engineering. New numerical methods and new techniques of measurement and testing may be developed in the future new materials and new constructions invented. Information on sound signals and sound transmission data of new sources and new materials can still be expected to be available. Hence the methodology to construct digital filters on the basis of computer data will be of interest in any case. Chapter 7 “Signal processing for auralization”, Chap. 8 “Characterization of sources” and Chap. 9 “Convolution and binaural sound synthesis” thus form the core of this book.

Towards Virtual Reality

In the end, the goal is to achieve an auralization in real-time, a dynamic interaction with the user and the user’s immersion and presence in the virtual scene. This most challenging auralization method arises when acoustics is integrated into the technology of “virtual reality”. Virtual reality (VR) is an environment generated in the

²It should be clearly distinguished between auralization and what is called “sonification”. Sonification means creation of sounds for experiencing non-acoustic data by auditory perception (e.g. flow dynamics, neutrino events and gravitational waves).

computer, which the user can operate and interact with in real time. One characteristic of VR is a three-dimensional and multimodal interface between computer and human being. Besides vision and acoustics, more senses covering haptics (force feedback), tactiles and eventually others should be added. In several applications of science, engineering and entertainment, CAD tools are well established. Visualization in CAD environments and VR is mostly the leading technology. Acoustics in VR (auralization) is not present to same extent and is often added just as an effect and without plausible or authentic reference to the virtual scene.

The process of generating the cues for the respective senses (3D image, 3D audio, etc.) is called “rendering”. Apparently, simple scenes of interaction, for instance when a person is leaving a room and closes a door, require complex models of room acoustics and sound insulation. Otherwise, the colouration, loudness and timbre of sound in and between rooms are not represented sufficiently. Another example is the movement of a sound radiating object behind a barrier or inside an opening of a structure, so that the object is no longer visible but can still be touched and heard. Sound also propagates by diffraction, one of the most difficult phenomena in general problems. The task of representing a realistic acoustic perception, localization and identification is therefore a big challenge.

Another particular difficulty in acoustic rendering can be explained by considering the large bandwidth and range of wavelengths involved. In acoustics, we must deal with three decades of frequencies (20 Hz to 20 kHz and wavelength from about 20 m to 2 cm). Therefore, it is necessary to model and simulate physical wave phenomena in environments of about the same dimensions as these wavelengths: the built environment has dimensions of several metres up to several tens of metres, while the dimensions of the objects of daily use, furniture, tools or finally of humans are in the range of metres and fractions of metres. Wave physics is most difficult (and most interesting) when wavelengths are in the same order of magnitude as the objects in the sound path. Neither approximations of small wavelengths (like in optics) nor large wavelengths (like in radio broadcasting) can be made. This might be the reason for the delayed implementation of acoustic components in virtual environments. Personal computers have recently become capable of simulating acoustics in real time, even though numerous approximations must be made to reach this goal. In the end, however, the resulting sound need not be physically absolutely correct, but only perceptually correct. Knowledge about human sound perception is, therefore, a very important prerequisite to evaluate auralized sounds and to set goals.

While interacting with real scenes, several senses are stimulated. Acoustics, besides vision, haptics and tactile cognition, yield important information about the environment and the objects in the environment. The cognition of the environment itself, external events and, very importantly, a feedback of the user’s own actions are supported by the auditory event. If a high degree of immersion in virtual environments is to be obtained, all these sensory events must be matched in timing and magnitude. With respect to acoustics, various kinds of generations must be taken into account: ambient speech, user’s own speech, sounds from objects, sound from

collision of objects or simply loudspeakers are elementary examples. The total sound of these events is characterized by resonances in the generating and transmitting systems and by external signals. If the sound is bounded in a cavity, it will be reflected by walls and finally reaches the receiver's ears. Humans as recipients evaluate the diverse characteristics of the total sound, segregate it into the individual objects and, furthermore, evaluate the environment, its size and mean absorption (state of furniture or fitting). In the case of an acoustic scene in a room, which is probably the majority of VR applications, therefore, an adequate representation of all these subjective impressions must be simulated, auralized and reproduced.

Acknowledgements

This book revision would not have been possible without the help of many students and colleagues. The contributions of the past and current doctoral students Sönke Pelzer, Frank Wefers, Lukas Aspöck, Jonas Stienen and Michael Kohnen in our activities in auralization and virtual reality systems kept the focus of our research in virtual acoustics. Professor Torsten Kuhlen in the RWTH Aachen Virtual Reality Group has always been a very important discussion and project partner.

I am also grateful for the support by Armin Erradj, Philipp Eschbach and Nils Rummler who re-designed and created figures. I am also grateful for excellent feedback and support from all members of the Institute of Technical Acoustics, ITA, of RWTH Aachen University.

This new edition is still published as RWTH Aachen edition but now under the umbrella of the Acoustical Society of America. I'm very grateful for the quick and smooth integration into "ASA Books"!

Sara Kate Heukerott, David Packer and Clement Wilson Kamalesh from Springer Publishers were very patient and always helpful in clarifying formal and technical aspects in template style and layout. The cooperation was very smooth indeed.

Aachen, Germany
May 2020

Michael Vorländer

Preface of First Edition

Acoustic Virtual Reality is a rather new and growing discipline. It integrates methods of physics and acoustic engineering with psychoacoustics and electroacoustic sound reproduction. The keywords of Acoustic Virtual Reality are “simulation”, “auralization” and “spatial sound reproduction”. The expression “auralization” is analogous to the well-known technique of “visualization”. In visual illustration of scenes, data or any other meaningful information, in movie animation and in computer graphics, we describe the process of “making visible” as visualization. In acoustics, auralization is when acoustic effects, primary sound signals or means of sound reinforcement or sound transmission, are processed into an audible result. The word auralization is used today to describe the process of signal generation, processing and reproduction as well as its result: the perceivable sound as auralization of an acoustic problem, a room, a building, a car, or any other industrial product.

The concept of auralization was first introduced in relation to acoustic modelling of sound fields in rooms. Already in 1929 in Munich, Spandöck and colleagues tried to process signals from measurements in model scale in such a way that one could listen to the acoustics of a room, in spite of the fact that the “room” was just a scale model on a laboratory table. The idea was formulated quickly, but at that time it was extremely difficult to put it into action. Twenty years later, in 1949, magnetic tape recorders were available. Spandöck et al. finally presented their system based on ultrasonic signals, scale models and a tape recorder working at several speeds (Spandöck 1934). This kind of technique can be regarded as an analogue computer. The basic elements were already present in this approach: Sound field modelling, processing with an arbitrary sound signal, and sound reproduction. The results at that time were not comparable with the expectations of audio quality we have today. Background noise and low bandwidth limited the listening experience, but nevertheless these experiments must be considered the first successful auralization.

With the dramatic development of computers, the concept of simulation and auralization was re-invented by (Schroeder et al. 1962) in the beginning of the 1960s. In 1968, the first room acoustical simulation software developed by

(Krokstad et al. 1968) was applied in room acoustics. By the 1990s the processor speed, memory space and convolution machines were sufficiently powerful to allow room acoustical computer simulation and auralization on a standard personal computer or on relatively small workstations. In the early 1990s the word “auralization” was introduced (Kleiner et al. 1993). Since this time, several improvements in modelling algorithms, in binaural processing and in reproduction technique have been made. Today, commercial software for room acoustical simulation considered incomplete without an option for auralization through the sound card of the computer. Several other fields of acoustics also included the term auralization in their vocabulary, particularly product sound design and building acoustics.

The main theme of this book is the general technique of auralization and its manifold applications, both offline and in real time. It covers three main components: Sound generation, sound transmission and sound reproduction. The methodology of simulation and auralization is developed from discussion of methods for acoustic modelling. After starting with chapters on a general introduction into linear acoustics in fluids and in structures and on basic features of sound perception (psychoacoustics) in free field and in rooms, Chaps. 7, 8 and 9 an introduction is given into the aspects of signal processing and auralization methodology in a general sense. In Chap. 10 the reader will learn about simulation models and techniques of applying numerical approaches for various acoustic problems.

Virtual environments must usually be created in virtual rooms or similar enclosures. The technique of simulation is therefore explained in more detail by setting more emphasis on acoustics in enclosures in Chap. 11; including room acoustics auralization in Sect. 11.9 (It is presented as a first example where the technique of auralization is applied³). Chapters 12 and 14 deal with strategies of simulation and auralization for applications in architectural acoustics other than room acoustics: airborne and impact sound insulation, structure-borne sound sources in buildings. Another field with application of auralization is vehicle acoustics. The global model of binaural transfer synthesis, BTPS, used in automotive acoustics, is discussed in Chap. 15. The final Chaps. 16 and 17 are focused on aspects of real-time processing and 3D sound reproduction technology.

This book is organized as a comprehensive collection of acoustic fundamentals, of methodology and strategies of acoustic simulation and auralization. With the mathematical background of a third-year university student, the reader will easily be able to follow the main strategy of auralization and apply their own implementations of auralization in various fields of acoustic engineering and virtual reality. For readers interested in basic research the technique of auralization may be useful to create sound stimuli for specific investigations in civil engineering and in architecture. Virtual reality applications with acoustic stimuli are also crucial in linguistic, medical, neurological and psychological research and in the field of human-machine interaction.

³Also historically it was first established in room acoustics.

Acknowledgements

“Auralization” has become a big field of interest in the acoustics community. Having taught a course in “Acoustic Virtual Reality” over the past eight years, I have recognized the need to concentrate all work done in acoustic simulation, auralization and acoustic virtual reality systems into a textbook. This book would not have been possible without the contributions of many helping hands. The contributions of the doctoral students Oliver Schmitz, Tobias Lentz and Dirk Schröder in our activities in auralization and virtual reality systems created a strong focus of research in the institute. Also Ingo Assenmacher must be mentioned who is collaborating with us to implement acoustics in the RWTH Aachen Virtual Reality center. Inspired by requests from numerous graduate students and colleagues in Germany and abroad, the idea of extending the lecture notes into a book was born quickly. Finally, while teaching “Acoustic Simulation and Auralization” during a sabbatical in Florianópolis, Santa Catarina, Brazil, the book project was started.

I am particularly grateful for the support by Pascal Dietrich who created figures in a way I never could have done myself. The permission to use the world’s largest database of acoustic absorption was given by Dr. Ingolf Bork from PTB Braunschweig. Many more colleagues could be named here who supported the book project by contributing with information, discussion or figures. I am grateful also for excellent feedback and support from all members of the Institute of Technical Acoustics, ITA, of RWTH Aachen University. First and foremost, I thank my predecessor as head of the institute and my academic teacher, Prof. Heinrich Kuttruff, for teaching me not only the physical and technical basics but also the inspiration, motivation, beauty and fun of acoustics.

Jonathan Rathsam from the University of Nebraska checked the linguistic quality and corrected my “German English”. Due to his patience and kind proposals for modifications I managed to finish the book in hopefully acceptable English, at least from the viewpoint of non-native speakers.

Thomas Lehnert from Springer Publishers was very patient and always helpful in clarifying formal and technical aspects in template style and layout. The cooperation was very smooth.

Last but not least, my family must be mentioned. For several months I was not able to contribute to our daily life of our home, school and other fields of family activities in an acceptable amount. I can only promise to Angelika, Paul, Tobias and Benjamin that this will be better from now on.

Aachen, Germany
June 2007

Michael Vorländer

Contents

1	Fundamentals of Acoustics	1
2	Sound Sources	17
3	Sound Propagation	31
4	Sound Fields in Cavities and in Rooms	51
5	Structure-Borne Sound	67
6	Psychoacoustics	75
7	Signal Processing for Auralization	99
8	Characterization of Sources	119
9	Convolution and Binaural Sound Synthesis	135
10	Simulation Methods	145
11	Simulation of Sound in Rooms	171
12	Simulation and Auralization of Outdoor Sound Propagation	225
13	Simulation and Auralization of Airborne Sound Insulation	235
14	Simulation and Auralization of Structure-Borne Sound	255
15	Transfer Path Analysis and Synthesis	265
16	Filter Construction for Real-Time Processing	277

17	3D Sound Reproduction	297
18	Acoustic Virtual Reality Systems	323
Annex	333
References	347
Index	361

Chapter 1

Fundamentals of Acoustics



1.1 Introduction

Acoustic waves are fascinating phenomena in gases, liquids or solids with numerous physical and technical aspects. Air is the medium of interest when it comes to audible sound. The gas dynamics of air determine the equations of the sound pressure and other field quantities from which the wave equation can be derived and solved for basic wave types. Properties of the waves such as frequencies and sound levels are associated with the human hearing sensation. The chapter on fundamentals of acoustics deals with the introduction into sound field physics and terminology of acoustics.

“Waves” are well-known to everybody, even if they are not cognitively identified as waves. While reading this book, light waves as an example of electromagnetic waves are scattered from the white paper and absorbed by printed letters. Electromagnetic waves were predicted by James Clerk Maxwell’s theory from 1864 and experimentally discovered by Heinrich Hertz in 1888. Their spectrum from gamma radiation, X-rays, ultraviolet, the visible range, infrared towards spectra for technical communication systems offers a fascinating wide range of natural phenomena and technical applications. Waves are to be considered as local oscillations in a physical “field” with the inherent effect of energy and information transport and are found in numerous areas of physics. The common approach in these areas is that small perturbations of the equilibrium yield linear or approximately linear forces and oscillating states of permanently recycled potential and kinetic energy.

Although water waves (Fig. 1.1) are possibly one of the most enjoyable examples, we now focus on a mechanical system to illustrate the nature of waves. A chain of masses is connected by springs. In this example, a one-dimensional wave is excited. When one mass is moved (by forced excitation), it takes kinetic energy; transfers this energy to the attached spring, which is compressed (and stores potential

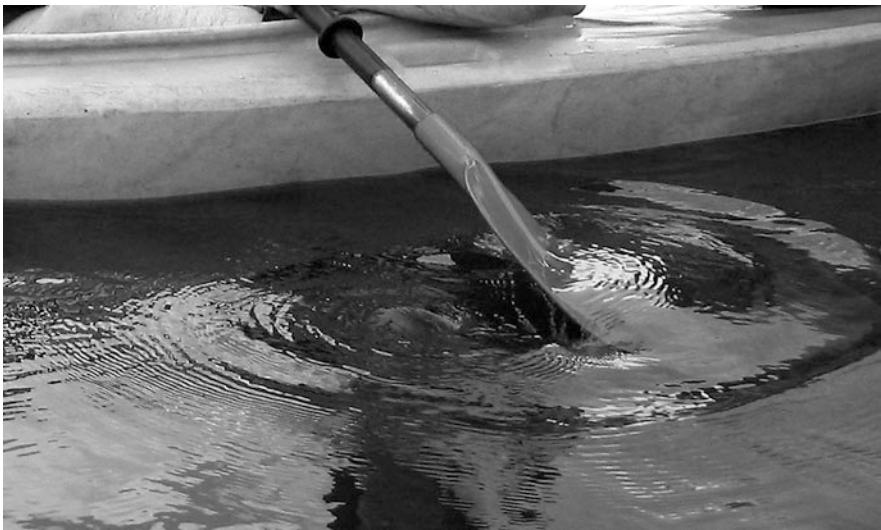


Fig. 1.1 Waves in water

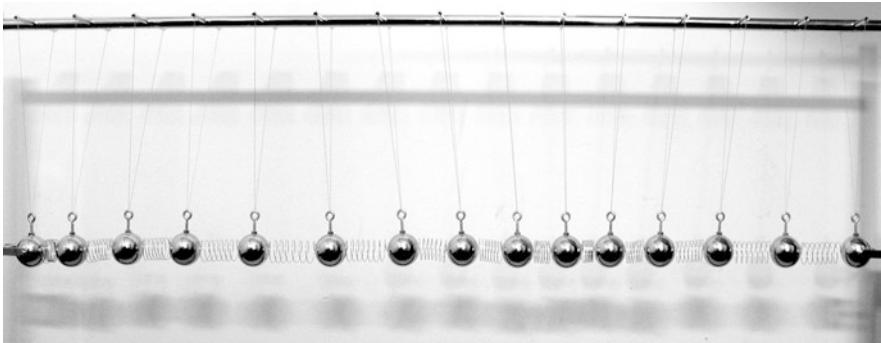


Fig. 1.2 Picture of metal spheres connected by springs

energy); recycles and transfers its energy to the next mass; and so on (Fig. 1.2). By intuition, we can easily imagine that heavy masses with large inertia and soft springs with high flexibility provide this transport effect at a slower speed than lightweight masses and stiff springs. We can also think of a row of children holding hands. When the first child pushes his neighbour, the neighbour will move to the side, pushing the child next to him or her and so on. When all children now stiffen their arm muscles, the wave movement in the row runs faster. By this analogy, the phenomena of energy transport in a wave and the microscopic nature of the wave speed are already understood.

1.2 Sound Field Equations and Wave Equation

Sound is a wave phenomenon in fluid or solid media. The main areas of acoustics are accordingly called airborne sound, underwater sound and structure-borne sound. The differential equations of vibration and waves in acoustics can be derived from dynamic physical laws of continuum mechanics. The physical foundations of linear acoustics are introduced in this chapter: the one-dimensional wave equation, its solutions and the three-dimensional generalization. We start with an image of sound. This picture was taken at a certain time and we observe that the medium's molecules or atoms are somehow displaced from their original position, of course, independent of the constant irregular thermal movement.

1.2.1 Sound Field Quantities

In a sound wave the particles (gas molecules, crystal lattice atoms, etc.) follow a space- and time-dependent displacement vector, $\vec{s} = (\xi, \eta, \zeta)$ (see Fig. 1.3). Accordingly the time derivative of this displacement is the particle velocity

$$\vec{v} = \frac{\partial \vec{s}}{\partial t} \quad (1.1)$$

with the components $v_x = \dot{\xi}(x, y, z, t)$ etc.

The displacements are neither homogeneous nor isotropic.¹ Therefore the medium will be compressed and decompressed. ρ_{tot} is the space- and time-dependent total density, ρ_0 the density of the medium in rest. The density fluctuations due to sound are thus

$$\rho = \rho_{\text{tot}} - \rho_0 \quad (1.2)$$

and the local sound-induced pressure fluctuations, closely related to density is given by

$$p = p_{\text{tot}} - p_0. \quad (1.3)$$

The latter quantity, p , is particularly important. We call it “sound pressure”. Note that the sound pressure is a scalar. In acoustics the sound pressure is usually the most important quantity of interest, mainly because the human ear is sensitive to sound pressure. Hence calculations or measurements of sound pressure yield directly the input quantity of the human hearing system.

¹Homogenous and isotropic means independent of translation or rotation, respectively.

Fig. 1.3 Microscopic view of a medium with sound



In fluid media like air, the elasticity of the medium is described by its compressibility. In this discussion, the thermodynamic state of the medium and its capability of storing energy (heat) are of crucial importance. Particle displacement and compression affect the pressure, but so do temperature and heat transfer. For acoustic waves in air, however, we might assume that the sound-induced oscillations are so fast that diffusion of heat between local areas of the medium is not possible. This is related to the simplified model of adiabatic processes for which we can use the adiabatic Poisson equation

$$\frac{p_{\text{tot}}}{p_0} = \left(\frac{\rho_{\text{tot}}}{\rho_0} \right)^{\kappa} \quad (1.4)$$

with κ denoting the adiabatic exponent, the ratio of heat capacities at constant pressure and volume, respectively. $\kappa = C_p/C_V$ ($= 1.4$ for air).

1.2.2 Derivation of the Wave Equation

Imagine a small volume element of thickness Δx in a one-dimensional fluid medium bounded by a tube with cross-section S . In the tube is an acoustic source pushing and pulling the volume element periodically with a strength of $qSdx$ (in units of $[m^3/s]$). It excites a small disturbance from the pressure equilibrium. The source can be assumed anywhere in the tube. It can, for instance, be represented by a small piston mounted flush in the tube wall. The pressure and particle velocity on the left-hand side of our test element, at x , might differ from the conditions at the right-hand side at $x + \Delta x$. A pressure difference will lead to a net force on the volume element (Fig. 1.4).

We obtain a movement of medium mass according to the Euler equation of all forces involved:

$$[(p_{\text{tot}})_x - (p_{\text{tot}})_{x+\Delta x}] S = \rho_{\text{tot}} S \Delta x \frac{dv}{dt}. \quad (1.5)$$

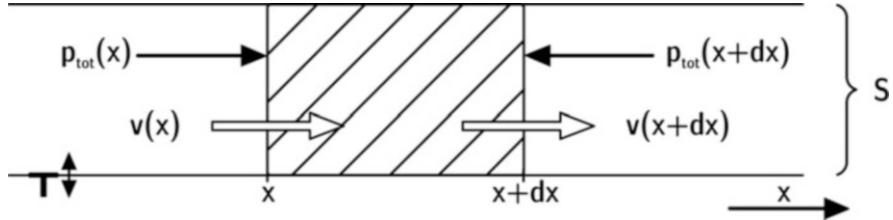


Fig. 1.4 Volume element in a one-dimensional fluid medium. Pressure variations are induced by a “pumping” source on the left-hand side

Since v is a function of $x(t)$ we must apply the chain rule

$$\frac{dv}{dt} = \frac{\partial v}{\partial t} + \frac{\partial x}{\partial t} \frac{\partial v}{\partial x} = \frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} \quad (1.6)$$

and, on the one hand, we find for infinitesimal Δx by setting $\Delta p/\Delta x \rightarrow \partial p/\partial x$

$$-\frac{\partial p_{\text{tot}}}{\partial x} = \rho_{\text{tot}} \left(\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} \right). \quad (1.7)$$

On the other hand, a movement of mass must comply with the conservation of mass. Mass reduction or increase within the volume element, thus, must correspond to a change in mass density. Mass changes can be induced by medium flow due to density differences or due to injections caused by the source.

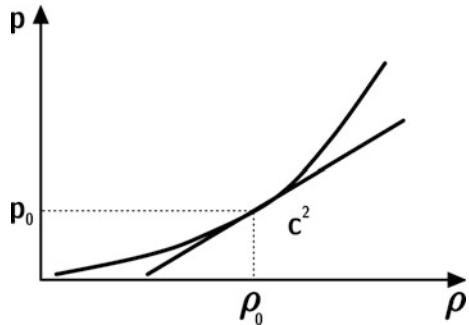
$$[(\rho_{\text{tot}} v)_x - (\rho_{\text{tot}} v)_{x+dx}] S = \left(\frac{\partial \rho_{\text{tot}}}{\partial t} - \rho_{\text{tot}} q \right) S \Delta x \quad (1.8)$$

or, again with $\Delta x \rightarrow \partial x$

$$-\frac{\partial(\rho_{\text{tot}} v)}{\partial x} = \frac{\partial \rho_{\text{tot}}}{\partial t} - \rho_{\text{tot}} q. \quad (1.9)$$

These two Eqs. (1.7) and (1.9), allow the derivation of an acoustic theory. They are, however, coupled in three variables, namely pressure, density and particle velocity, and they are non-linear. Two linear equations can be easily found if the effects of sound are assumed to be small. We will see later that this approximation is quite sufficient for almost all sound events of interest in this book. Now, if the sound pressure is small compared with the static pressure, $p \ll p_0$, and the densities follow the same prerequisite, $\rho \ll \rho_0$, we can decompose the Eqs. (1.7) and (1.9) into Taylor series and neglect small terms of higher order. A pure factor ρ_{tot} can also be replaced by ρ_0 .

Fig. 1.5 Linearization for small amplitudes of density as a function of pressure



Furthermore, the fact that the density and the pressure are linked in the adiabatic process lets us change the variable ρ to p :

$$p = \left(\frac{dp_{\text{tot}}}{d\rho_{\text{tot}}} \right)_{\text{ad}} \rho = c^2 \rho \quad (1.10)$$

with a constant c^2 as abbreviation of $dp/d\rho$. (Fig. 1.5)

The result of linearization and replacing the density is the set of two linear sound field equations

$$-\frac{\partial p}{\partial x} = \rho_0 \frac{\partial v}{\partial t} \quad (1.11)$$

$$-\rho_0 \frac{\partial v}{\partial x} = \frac{1}{c^2} \frac{\partial p}{\partial t} + \rho_0 q \quad (1.12)$$

From these two equations, one can easily eliminate one variable (the particle velocity v) to obtain one differential equation containing our variable of highest interest, the sound pressure p :

$$\frac{\partial^2 p}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = -\rho_0 \frac{\partial q}{\partial t} \quad (1.13)$$

or in short

$$\Delta p - \frac{1}{c^2} \ddot{p} = -\rho_0 \dot{q}. \quad (1.14)$$

The equation is well known in mathematics and physics as wave equation. The same formal notation can be used to derive this equation for particle velocity, density or temperature. In three dimensions, the spatial differential operator, the Laplace operator, in Cartesian coordinates (x, y, z) is added according to common notation used in mathematics and physics

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \quad (1.15)$$

in cylindrical coordinates (r, φ, z)

$$\Delta = \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2}{\partial \varphi^2} + \frac{\partial^2}{\partial z^2} \quad (1.16)$$

or in polar coordinates (r, ϑ, φ)

$$\Delta = \frac{1}{r^2} \left(\frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{\sin \vartheta} \frac{\partial}{\partial \vartheta} \left(\sin \vartheta \frac{\partial}{\partial \vartheta} \right) + \frac{1}{\sin^2 \vartheta} \frac{\partial^2}{\partial \varphi^2} \right). \quad (1.17)$$

1.3 Plane Waves in Fluid Media

The direct solution of the example of one-dimensional sound propagation in Sect. 1.1 is the plane wave. We will find terms describing wave acoustics such as speed of sound, sound intensity and energy density and can define the “sound pressure level”. Equation (1.13) holds for the case that the spatial displacements, particle velocities, their gradients or density or pressure have only a component in x direction.

The wave equation is solved by any function f with a variable in the form of $x-ct$, or g as function of $x+ct$ (d'Alembert solution), see, for instance, (Kuttruff 2007)

$$p(x, t) = f(x - ct) + g(x + ct). \quad (1.18)$$

The first term, f , describes the propagation of the local state of sound pressure $p(x, t)$ in space and time in positive x direction, the latter term, g , in negative x direction. This can be easily understood by considering the case of propagation to the right side assuming a function $f(x)$ with a maximum at $x = 0$ at time zero. After the time t has passed, the maximum will be found at the location $x = ct$.

The speed of propagation is c , the speed of sound. As described in 1.1.1 the constant c is calculated in a first approach based on the ideal gas theory from

$$c^2 = \left(\frac{dp_{\text{tot}}}{d\rho_{\text{tot}}} \right)_{\text{ad}} = \frac{\kappa p_{\text{tot}}}{\rho_{\text{tot}}}. \quad (1.19)$$

Taking into account more thermodynamic effects such as the humidity, altitude, etc., it can be estimated rather precisely (ISO 9613). For most cases of sound propagation in air, the approximation

$$c = 343.2 \sqrt{\frac{273.15 + \theta}{293.15}} \text{ m/s} \quad (1.20)$$

is a sufficient estimate (θ is the temperature in degrees Celsius).

Liquid media in hydrostatics are usually considered incompressible. Acoustic waves in liquids are in fact a result of small perturbation of the zero compressibility. Liquids are thus characterized by their adiabatic compressibility

$$\beta_{\text{ad}} = \frac{1}{\rho_0} \left(\frac{d\rho_{\text{tot}}}{dp_{\text{tot}}} \right)_{\text{ad}} \quad (1.21)$$

and the speed of sound is

$$c = \frac{1}{\sqrt{\rho_0 \beta_{\text{ad}}}}. \quad (1.22)$$

Now we come back to the elementary solution $p(x, t)$. Assuming $g = 0$ we now look at a wave in positive x direction. All locations (y, z) in planes parallel to the x direction have the same conditions. By using Eq. (1.11), we can calculate the particle velocity:

$$v = -\frac{1}{\rho_0} \int f' dt = \frac{1}{\rho_0 c'} f \quad (1.23)$$

and

$$\frac{p}{v} = \rho_0 c = Z_0. \quad (1.24)$$

This ratio, Z_0 , is called wave impedance or characteristic impedance for plane waves. It is an important reference. It can be interpreted as the characteristic resistance of the medium against pressure excitation in some kind of cause-and-effect interpretation: the amount of the driving pressure needed to set the medium's particles into motion (Table 1.1).

Table 1.1 Characteristic acoustic data for air and water

At normal conditions, 20 °C	Sound speed in m/s	Characteristic impedance in kg/m ² s
Air	344	414
Water	1484	1,48 · 10 ⁶

1.4 Plane Harmonic Waves

We get a harmonic wave of sound pressure in the positive x direction by choosing a harmonic function representing $f(x-ct)$ (without loss in generality g is set to zero). In complex form

$$\underline{p}(x, t) = \hat{p} e^{-jk(x-ct)} = \hat{p} e^{j(\omega t - kx)} \quad (1.25)$$

\hat{p} is called pressure amplitude, k wave number and ω angular frequency ($\omega = kc$). Note the symmetry between the terms kx and ωt in the harmonic function and that space and time domain are coupled by a factor of c . Now, it is well known that the period of harmonic functions is 2π . Hence the obvious role of the wave number and angular frequency is to re-scale the periods of the wave in space and time, respectively, to 2π . The period in space is called wavelength, λ . We obtain it by

$$k = \frac{2\pi}{\lambda}. \quad (1.26)$$

With the important relation

$$c = f \cdot \lambda = \frac{\lambda}{T} \quad (1.27)$$

we can introduce the temporal period, T , of the wave and with f denoting the frequency in the unit Hertz:

$$\omega = \frac{2\pi}{\lambda} = 2\pi f. \quad (1.28)$$

1.5 Wideband Waves and Signals

From the harmonic wave, we can directly construct any other wave function or, referring to one measurement point in space, a time function of sound pressure, called an acoustic “signal”. The procedure of superposition of harmonic signals into a more complex waveform is given by Fourier transformation.

For continuous periodic pressure-time functions with period T_0 , the transformation into the frequency domain yields the set of Fourier coefficients, \underline{S}_m , which are the complex amplitudes of the respective harmonic signal components. The set of complex amplitudes is called “spectrum”.

$$\underline{S}_m = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} p(t) e^{-jm2\pi f_0 t} dt. \quad (1.29)$$

Vice versa, the pressure-time signal, $p(t)$, can be reconstructed by Fourier synthesis, i.e. by adding all harmonic components:

$$p(t) = \sum_{m=-\infty}^{\infty} \underline{S}_m e^{jm2\pi f_0 t} \quad (1.30)$$

Periodic signals have a specific fundamental frequency, f_0 , and the spectrum is composed of discrete components at multiples of the fundamental frequency. It is called a line spectrum and has accordingly a frequency resolution of $\Delta f = f_0 = 1/T_0$. The same concept can be extended towards aperiodic signals ($T_0 \rightarrow \infty$), which then have a continuous spectrum (line spacing $\rightarrow 0$). More details will be discussed in [Chap. 7.2](#).

When the spectrum contains several frequencies, we talk about wideband or broadband sound. Typical spectra of acoustic signals are shown in [Fig. 1.6](#).

1.6 Energy and Level

Usually the sound pressure is not presented in linear form in its unit Pascal ($1 \text{ Pa} = 1 \text{ N/m}^2$). In daily life, the strength of sound is indicated by “decibels”. One reason is the enormous range of sound pressures in music, speech, and the urban and working environment, another is the somewhat better match with human hearing sensation (at least in first approximation, see also [Chap. 6](#)). The range of sound pressures should be also discussed in relation to our initial assumption of sound being a small displacement or pressure or density fluctuation compared with static conditions. If we just assume a static pressure of 100 kPa in the atmosphere at sea level, we have to deal with sound pressures of orders of magnitude lower, between 0.00001 Pa and 1000 Pa. The lower limit is near the “hearing threshold”, the limit of sensation of human hearing. The upper limit is called the threshold of pain which needs no further explanation. Thus typical sound pressures are indeed by orders of magnitude lower than the static pressure p_0 .

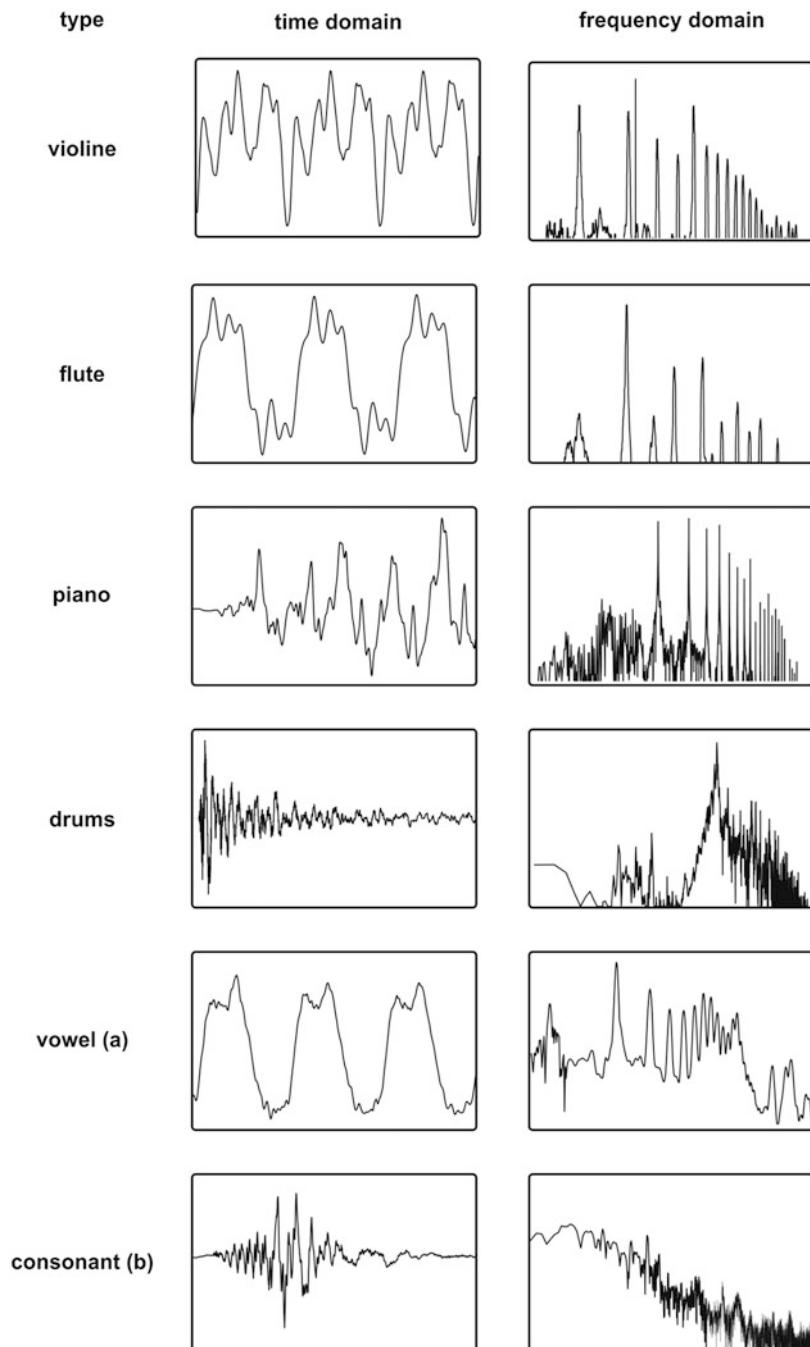


Fig. 1.6 Typical acoustic signals and their spectra

In practical acoustics, this enormous range is mapped to a logarithmic scale between about 0 dB and 130 dB, more or less similar to a scale counted in percent with a resolution practically in steps of 1. This resolution is appropriate for discussing the differences in audible sounds, and it corresponds to the just noticeable difference, jnd, of about 1 decibel.

The decibel scale surely has its merits, although we will later emphasize that sound evaluation purely based on decibels or related quantities will not be sufficient without a more thorough analysis or with auralization.

The decibel scale is based on sound energy. Similarly as in other wave and vibration phenomena in radio waves, in voltage and current measurements, etc. the level is defined by the energy of the wave. Due to the local harmonic medium particle movement, the total energy contained in a small volume element can be interpreted similar to the energy in little pendulums in terms of the kinetic and potential energy. The volume will be so small that all particles in it move in the same way. At the time of maximum particle velocity, the total energy is purely kinetic, whereas at zero velocity, the total energy is purely potential. This approach leads to two possible equations for the total energy density:

$$w = \frac{\rho_0 \hat{v}^2}{2} = \frac{\tilde{p}^2}{2\rho_0 c^2} \quad (1.31)$$

in which we used Eq. (1.24) for changing from velocity to pressure in a plane wave.

Having introduced the “root mean square” sound pressure (rms), the sound pressure level can finally be written as:

$$L = 10 \log \frac{\tilde{p}^2}{p_0^2} = 20 \log \frac{\tilde{p}}{p_0} \quad (1.32)$$

in dB (decibel), with

$$\tilde{p} = \sqrt{\frac{1}{T} \int_0^T p^2(t) dt} = p_{\text{rms}} \quad (1.33)$$

while setting the reference sound pressure $p_0 = 20 \mu\text{Pa}$ which is approximately the human hearing threshold at mid frequencies (Table 1.2).

Table 1.2 Sound pressure levels of typical sound events

Event	Level in dB
Hearing threshold at mid frequencies	0
Anechoic chamber	0–15
Bedroom	25–30
Living room	40–55
Conversation	60
Office	70
Typical noise limit for factories	85
Pneumatic hammer	100
Rock concert, disco or walkman maximum	110
Jet engine, 25 m away	120
Rocket at start	> 190

1.7 Sound Intensity

The microscopic energy in a sound wave is not just a static phenomenon. Energy is also transported. Wave propagation should not, however, be mixed up with particle flow. At zero mean flow (like random wind), the net particle displacement is zero, while energy is transported to the neighbouring volume element and so forth. This effect allows a very deep and detailed investigation of sound fields, particularly of sound fields that are more complex than simple plane waves. The basic quantity for describing the mean energy flow is the energy transported per second through a reference area of 1 m^2 . It is called sound intensity:

$$\vec{I} = \overline{\vec{p} \cdot \vec{v}} = \frac{1}{T} \int_0^T \vec{p} \vec{v} \, dt. \quad (1.34)$$

In a plane wave, sound pressure and particle velocity are in phase (Eq. (1.24)) and the sound intensity formula reduces to

$$|\vec{I}| = \frac{\overline{p^2}}{\rho_0 c} = \frac{\tilde{p}^2}{\rho_0 c} \quad (1.35)$$

with the direction of sound intensity in direction of propagation. Sound intensity can also be denoted with a level, the intensity level:

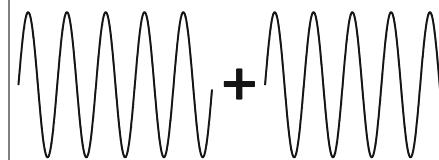
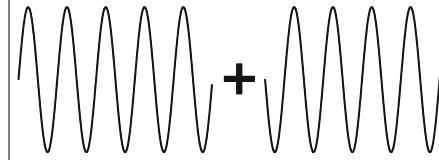
$$L_I = 10 \log \frac{|\bar{I}|}{I_0} \quad (1.36)$$

$I_0 = 10^{-12} \text{ W/m}^2$. This reference intensity is chosen to adjust it to the same numbers of levels of sound pressure and sound intensity in a plane wave.

1.8 Level Arithmetic

If several sound pressure signals are present at a time, the total pressure is the sum of the individual pressures. In case of coherent signals, i.e. with identical frequency and specific phase relation, the pressure-time functions must be added, and the rms value and the level are calculated in the end. In the case of incoherent waves of different frequencies or frequency compositions, the superposition reduces to adding the energies. The reason for the simplification can be interpreted as the effect of the binomial formula $(p_1 + p_2)^2 = p_1^2 + p_2^2 + 2p_1p_2$ with the latter term of the pressure signal multiplication cancelling with incoherent p_1 and p_2 in the averaging process (Table 1.3).

Table 1.3 Example: level addition of two signals of 50 dB

Phase relation		Total level in dB
Coherent, in phase		56
Coherent, anti-phase		$-\infty$
Incoherent		53

Thus for incoherent signals the quadratic pressures or the energy densities can be added directly:

$$w_{\text{total}} = \sum_{i=1}^N w_i = \frac{1}{\rho_0 c^2} \sum_{i=1}^N \tilde{p}_i^2 \quad (1.37)$$

or in level representation

$$L_{\text{total}} = 10 \log \sum_{i=1}^N 10^{L_i/10} = 10 \log \frac{\sum_{i=1}^N \tilde{p}_i^2}{p_0^2} \quad (1.38)$$

1.9 Frequency Bands

Standardized frequency bands are often used in acoustics, usually one-third octave bands or octave bands. The midband frequencies of one-third octave bands are defined on a logarithmic frequency scale as follows (here in the example of the base-2 logarithm)

$$\begin{aligned} f_u &= 2^{1/3} \cdot f_l \\ \Delta f &= f_u - f_l = f_l (2^{1/3} - 1) \\ f_m &= \sqrt{f_l \cdot f_u} \\ f_{m+1} &= 2^{1/3} f_m \end{aligned} \quad (1.39)$$

with f_l and f_u as lower and upper edge frequency and f_m , f_{m+1} as midband frequencies of the bands m and $m + 1$. Similarly, for octave bands holds:

$$\begin{aligned} f_u &= 2 \cdot f_l \\ \Delta f &= f_u - f_l = f_l \\ f_m &= \sqrt{f_l \cdot f_u} = \sqrt{2} f_l \\ f_{m+1} &= 2 \cdot f_m \end{aligned} \quad (1.40)$$

Fractional bands with bandwidth of 1/6 or 1/12 octave are also in use. The separation of broadband spectra into several frequency bands allows the discussion

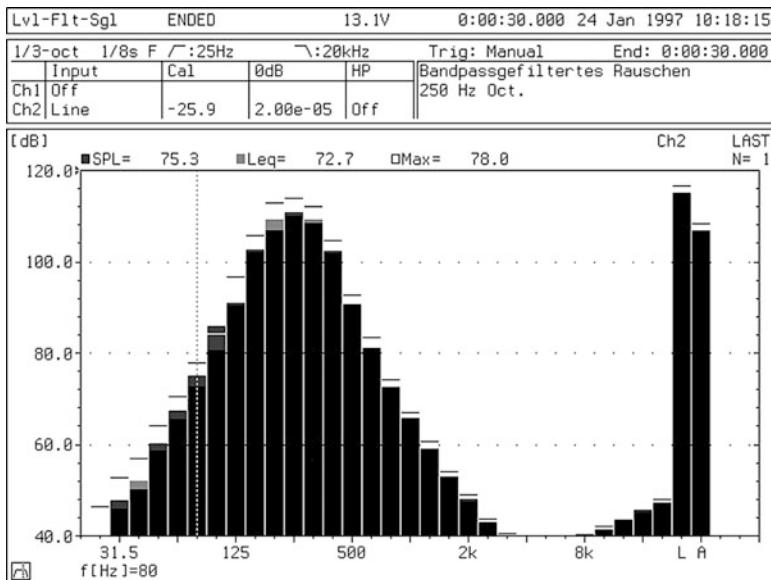


Fig. 1.7 Band-filtered spectrum of a broadband signal. (Courtesy of Norsonic A/S)

of transmission characteristics or noise frequency content. The band-filtered results, however, still express sound levels in decibels (Fig. 1.7). A rigorous time-frequency transformation such as a Fourier transformation provides the complete information based on amplitude and phase spectrum. This specific complex spectrum in its interpretation of Fourier transformation theory is discussed in more details below (see Sect. 7.4).

Chapter 2

Sound Sources



2.1 Introduction

In the chapter on fundamentals in acoustics and for the discussion of plane waves, we concentrated on the sound field at arbitrary observation points. At this point, sound sources were not yet taken into consideration. The characterization of sound sources, however, is one of the most important tasks in simulation and auralization. Solving the wave equation in polar coordinates is the first step to obtain a physical description of a source. The solution will lead to fundamental properties of spherical waves, to sound source directivities and to mathematical methods for acoustic signal processing of radiation problems.

2.2 Spherical Waves

At the origin of the coordinate system, at $r = 0$, we assume a small, point-like source (dimensions $\ll \lambda$), called a monopole, which emits a certain volume per second. The source is characterized by its capacity to set particles in the medium in motion. The corresponding strength, Q , is denoted as “volume flow” or “volume velocity”, see below.

Due to the spherical symmetry and to its omnidirectionality, the sound field does not depend on ϑ and φ . In this case, the Laplace operator reduces to

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} \quad (2.1)$$

and the wave equation reads

$$\frac{\partial^2 p}{\partial t^2} + \frac{2}{r} \frac{\partial p}{\partial r} - \frac{1}{c^2} \ddot{p} = -\rho_0 \dot{Q}. \quad (2.2)$$

By using the solution with the form

$$p(r, t) = \frac{f(r, t)}{r} \quad (2.3)$$

the differential equation is transformed into the simpler equation

$$\frac{\partial^2 f}{\partial t^2} - \frac{1}{c^2} \ddot{f} = -\rho_0 \dot{Q}. \quad (2.4)$$

It is solved by the d'Alembert solution $f = f(r - ct)$ in the same way as shown above. The general solution of Eq. (2.4) is thus

$$p(r, t) = \frac{1}{r} f(r - ct). \quad (2.5)$$

$p(r, t)$ represents a pressure wave propagating from $r = 0$ towards larger r . It is called a “spherical wave”.¹

We still need to clarify the relation between $f(r - ct)$ and the source term $-\rho_0 \dot{Q}$. The source strength Q will lead to a temporal volume motion of $Q(t)$ in [m^3/s]. Q is called the volume velocity or “sound flow”. Then it yields (according to (Kuttruff 2007))

$$p(r, t) = \frac{\rho_0}{4\pi r} \dot{Q} \left(t - \frac{r}{c} \right). \quad (2.6)$$

This is the general expression for a spherical wave excited by a point source of sound flow $Q(t)$. It is worth mentioning that a source signal $Q(t)$ is not radiated into sound pressure at the observation point in its original form as a volume velocity but as its time derivative. This, by the way, is the reason why small loudspeakers have a poor low-frequency output.

2.3 Harmonic Monopole Source and Sound Power

We assume a harmonic excitation

¹The second formal solution is given by $g = g(r + ct)$. It represents sound coming from an infinite sphere towards a point sink in the centre.

$$\underline{Q}(t) = \hat{Q} e^{j\omega t} \quad (2.7)$$

Then, with Eq. (2.6), we obtain

$$\underline{p}(r, t) = \frac{j\omega\rho_0\hat{Q}}{4\pi r} e^{j(\omega t - kr)} \quad (2.8)$$

and

$$\underline{v} = \frac{-1}{j\omega\rho_0} \frac{\partial p}{\partial r} = \frac{\hat{Q}}{4\pi} \left(jk + \frac{1}{r} \right) \frac{e^{j(\omega t - kr)}}{r}, \quad (2.9)$$

the latter by using Eq. (1.11).

The wave impedance is no longer real and frequency independent, but given by

$$\underline{\underline{v}} = \frac{j\omega\rho_0}{jk + \frac{1}{r}} = Z_0 \frac{1}{1 + \frac{1}{jkr}}. \quad (2.10)$$

Furthermore, the sound intensity is

$$I = \frac{\rho_0 \hat{Q}^2}{32\pi^2 c} \cdot \frac{\omega^2}{r^2} \approx \frac{\tilde{p}^2}{\rho_0 c} \text{ for } kr \gg 1. \quad (2.11)$$

The sound intensity increases with frequency and decreases with distance, both quadratically. The total radiated sound power can be calculated by integrating the sound intensity over a surrounding measurement surface. Here, the integration reduces to multiplication with an observation surface area of $4\pi r^2$. The result is

$$P = \iint I(r) r^2 dr \sin \theta d\theta d\varphi = \frac{\rho_0 \hat{Q}^2}{8\pi c} \cdot \omega^2. \quad (2.12)$$

Thus, we obtain the relation

$$I = \frac{P}{4\pi r^2}. \quad (2.13)$$

In practice, the sound power is also expressed in decibels with reference to $P_0 = 10^{-12}$ W:

$$L_W = 10 \log \frac{P}{P_0} \quad (2.14)$$

These findings can be applied for any kind of small volume-moving sound source.

Table 2.1 Sound power and sound power levels of typical sources

Sound source	Sound power P / W	Sound power level L_w / dB
Speech	10^{-5}	70
Human voice, maximal	10^{-2}	100
Grand piano	10^{-1}	110
Orchestra, fortissimo	100	140
Aircraft at takeoff	$10^3\text{--}10^4$	150–160

Sound power levels are part of the characterization of sound sources (Table 2.1). These levels of sound emission, however, should never be mixed up with the sound pressure at the receiver. If the distance between source and receiver is unknown, the sound pressure affecting the receiver cannot be determined.

Source characterization will be one of the central questions in auralization. Simple conditions such as those we assumed for our discussion of point sources are usually not present. First of all, a real source will have a limited power capacity. Accordingly the force applied to the medium and the amount of particle motion and sound radiation will be limited somehow. The key to describing the efficiency of a source is the resistance of the sound field to accept the force and velocity of the source vibration (often transmitted by a certain surface or membrane). Since the point source has no “surface”, we will therefore now discuss distributed source arrangements. A pulsating sphere will serve as starting point (Fig. 2.1).

2.4 Pulsating Sphere and Radiation Impedance

A spherical wave can also be generated by a spherical surface of any size. With harmonic movement of the surface we obtain the same result as for the point source. The sphere, however, must deal with the medium and its impedance load. At its radius, a , the counterforce of the medium is

$$4\pi a^2 \underline{p}(a, t) = \underline{Z}_r \underline{v}(a, t), \quad (2.15)$$

with $\underline{p}(a, t)$ denoting the sound pressure at the sphere surface and \underline{Z}_r the radiation impedance. $S = 4\pi a^2$ is the nominal surface area of the sphere. We can thus calculate the radiation impedance.

$$\underline{Z}_r = \frac{\rho_0 c S}{1 + \frac{1}{jka}} \quad (2.16)$$

According to basic power output relationships in electrical or mechanical systems, the radiated power of the source depends on the source surface velocity and on the real part of the radiation impedance, W_r

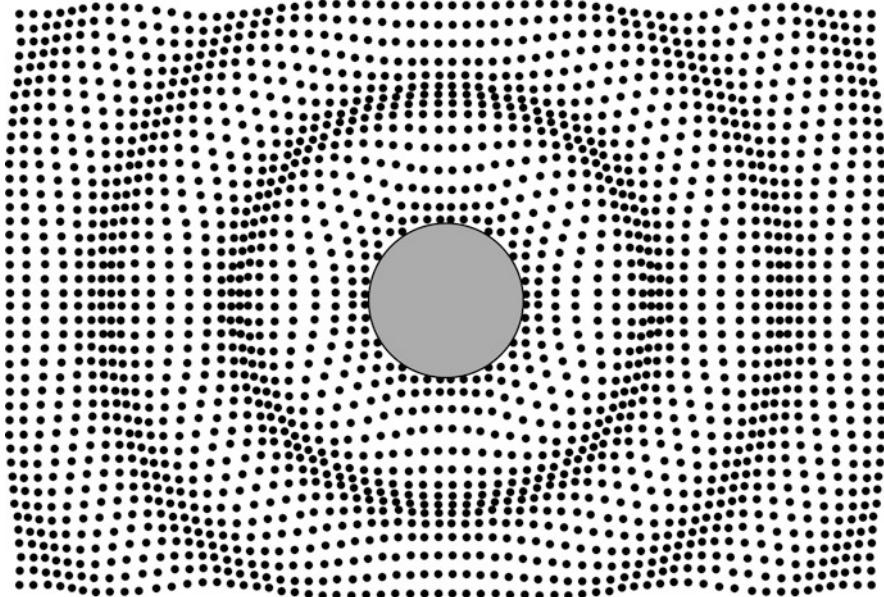


Fig. 2.1 Radiation from a spherical source (after (Möser 2004))

$$P = \frac{1}{2} |v(a, t)|^2 \operatorname{Re} \{Z_r\} = \frac{1}{2} |v(a, t)|^2 W_r \quad (2.17)$$

$$W_r = \frac{\rho_0 c S}{1 + \frac{1}{k^2 a^2}} \approx \begin{cases} S \rho_0 c k^2 a^2 & \text{for } ka \ll 1 \\ S \rho_0 c & \text{for } ka \gg 1 \end{cases} \quad (2.18)$$

As the equation for the radiation impedance has the same mathematical structure as the impedance of a parallel circuit, it can be re-written in the form of admittance

$$\underline{Y}_r = \frac{1}{S \rho_0 c} + \frac{1}{j \omega m_s}. \quad (2.19)$$

In this notation, apparently the source moves a certain amount of air, the mass of which is m_s .

$$m_s = 4\pi a^3 \rho_0. \quad (2.20)$$

At low frequencies, the admittance (mobility) is dominated by the latter term. More energy is then spent moving air loosely forth and back, but less energy is effectively transformed into sound radiation. The higher the frequency, the more the air is compressed and the greater the resulting sound pressure.

These relations are very basic, but they apply in principle to any kind of sound source (piston tube opening, plate, wall) with a radiating area. The radiation area may be a real plate or membrane or just a virtual radiating area in an open aperture. A real, distributed source has a specific near field as well. For more details see (Morse and Ingard 1968).

2.5 Multipoles and Extended Sources

Multipole sources or extended sources consisting of an arrangement of monopoles can be composed by summing or integrating over a distribution of monopoles or infinitesimally small surface elements, respectively. Due to interferences of the sound pressure contributions of all sources in this distributed arrangement, a complicated spatial radiation pattern is generated, which depends on the direction and the distance. Another approach towards a general description of direction-specific radiation, particularly for sources with spherical symmetry, is based on spherical harmonics. This technique will be discussed later in Sect. 2.6.

To understand the principle and effect of sound radiation of higher order, it is sufficient to discuss a dipole source created from two harmonic monopoles with opposite phase, $\hat{Q}_2 = -\hat{Q}_1 = \hat{Q}$ (Fig. 2.2).

The general equation for point source superposition

$$\underline{p} = \frac{j\omega\rho_0}{4\pi} \sum_n \hat{Q}_n \frac{e^{j(\omega t - kr_n)}}{r_n}, r_n = |\vec{r}_n| \quad (2.21)$$

reads in this case

Fig. 2.2 Dipole source

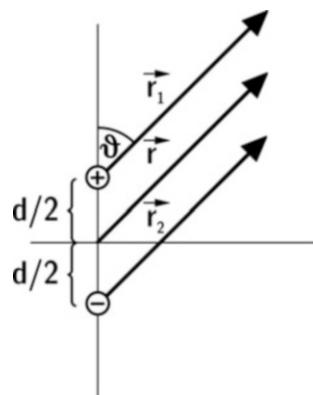
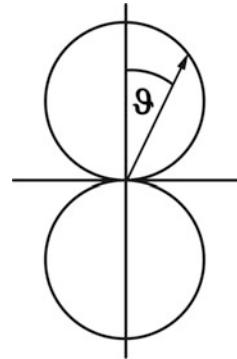


Fig. 2.3 Figure-of-eight directionality of a dipole source. Note the rotational symmetry in the vertical axis



$$\underline{p} = \frac{j\omega\rho_0\hat{\underline{Q}}}{4\pi r} e^{j(\omega t - kr)} \left(e^{jkd \cos \vartheta} - e^{-jkd \cos \vartheta} \right). \quad (2.22)$$

The distance between the two sources is assumed to be small ($kd \ll 1$). Then Eq. (2.22) reduces

$$\underline{p} \approx \frac{j\omega\rho_0\hat{\underline{Q}}}{4\pi r} e^{j(\omega t - kr)} \cdot jkd \cos \vartheta = -\rho_0 c \frac{k^2 d \hat{\underline{Q}}}{4\pi r} \cos \vartheta \cdot e^{j(\omega t - kr)}. \quad (2.23)$$

The result of the sound pressure is composed of a monopole term including the “differentiation effect” of the sound field ($j\omega$) and the $1/r$ distance law. Another factor is introduced which contains the “directivity factor”. In this case it has a figure-of-eight characteristic (Fig. 2.3).

Directional characteristics are relevant for any kind of source with a surface velocity distribution differing from spherical uniformity. In most cases it can be calculated by solving Eq. (2.21) or the corresponding integral form with the assumption of large wavelengths ($kd \ll 1$) and for the so-called “far field” ($d \ll r$), with d in this case being a characteristic source dimension. Based on these assumptions, the summation or integration result can be divided into (r) and (ϑ, φ), thus, into a distance-dependent monopole term and a directional term dependent on the polar and azimuth angles.

If all monopole sources have an identical phase, we can further define

$$\hat{\underline{Q}}_{\text{tot}} = \sum_n \hat{\underline{Q}}_n \quad (2.24)$$

and the directionality factor $\Gamma(\vartheta, \varphi)$, which is usually depicted in polar diagrams,

$$\underline{\Gamma}(\vartheta, \varphi) = \frac{1}{\hat{\underline{Q}}_{\text{tot}}} \sum_n \hat{\underline{Q}}_n e^{jkr'_n \cos [\bar{r}, \bar{r}_n]} \quad (2.25)$$

which in the end yields

$$\underline{p} = \frac{j\omega\rho_0\hat{Q}_{\text{tot}}}{4\pi r} e^{j(\omega t - kr)} \cdot \underline{\Gamma}(\vartheta, \varphi). \quad (2.26)$$

(Kuttruff 2007; Möser 2004), like others, also describe other cases of sources such as arrays or the piston source. For the purpose of understanding the basic principles of separation into the monopole term and the directivity, this overview is sufficient.

The integration of the radiated intensity over a spherical surface in the far field yields a sound power of

$$P = \frac{\rho_0\hat{Q}_{\text{tot}}^2}{32\pi^2 c} \omega^2 \iint_{\Omega} |\underline{\Gamma}(\vartheta, \varphi)|^2 d\Omega \quad (2.27)$$

For monopoles the result of the integral is 4π . Note that normalization with respect to the direction of maximum radiation is also used. The far-field normal intensity is then

$$I_n = \frac{P}{4\pi r^2} |\underline{\Gamma}(\vartheta, \varphi)|_{\max=1}^2, \quad (2.28)$$

and the sound pressure level calculated from the sound power level is

$$L_I = L_W - 20 \log r - 11 + L_D \quad (2.29)$$

with r in metres and L_D denoting the directivity factor

$$L_D = 10 \log \iint_{\Omega} |\underline{\Gamma}(\vartheta, \varphi)|^2 d\Omega. \quad (2.30)$$

2.6 Spherical Harmonics

The general solution of the Helmholtz equation in polar coordinates can as well be expressed by spherical harmonics. Starting with the Laplace operator (Eq. (1.17)), we obtain a separable solution (Williams 1999)

$$p = \Psi(r, \vartheta, \varphi) e^{j\omega t} = R(r) P(\vartheta) \Phi(\varphi) e^{j\omega t} \quad (2.31)$$

which leads to the three differential equations including three constants, m , C and k .

$$\begin{aligned} \left(\frac{\partial^2}{\partial \varphi^2} + m^2 \right) \Phi &= 0 \\ \frac{1}{\sin \vartheta} \frac{\partial}{\partial \vartheta} \left(\sin \vartheta \frac{\partial P}{\partial \vartheta} \right) + \left(C - \frac{m^2}{\sin^2 \vartheta} \right) P &= 0 \\ \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial R}{\partial r} \right) + \left(k^2 - \frac{C}{r^2} \right) R &= 0 \end{aligned} \quad (2.32)$$

The solution for the azimuth angle is straightforward.

$$\Phi = e^{im\varphi} \quad (2.33)$$

For the elevation angle ϑ , the orthonormal of Legendre polynomials of degree m and order n completes the result regarding the spatial angles. They are defined as follows ($x = \cos \vartheta$):

$$P_n^m(x) = (-1)^m (1-x^2)^{m/2} \frac{d^m}{dx^m} P_n(x), \quad 0 \leq m \leq n \quad (2.34)$$

$$P_n^{-m}(x) = (-1)^m \frac{(n-m)!}{(n+m)!} P_n^m(x), \quad (2.35)$$

with

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n. \quad (2.36)$$

The intermediate result for the angular variables is

$$Y_n^m(\vartheta, \varphi) = \sqrt{\frac{2n+1}{4\pi}} \frac{(n-m)!}{(n+m)!} P_n^m(\cos \vartheta) e^{im\varphi}. \quad (2.37)$$

These functions, here given in normalized form, are called “spherical modes” or “spherical harmonics”, in short “SH”. They are orthonormal with respect to integration over the surface of a unit sphere.

Spherical harmonics for negative and positive m can be computed by using the complex conjugates

$$Y_n^m(\vartheta, \varphi) = (-1)^m Y_n^{-m*}(\vartheta, \varphi). \quad (2.38)$$

Legendre polynomials are easily found in mathematical software and, most useful for the purpose of auralization, in software libraries for signal processing. The spatial radiation distribution in general is thus given by linear combinations of Eq. (2.37); see also Sect. 2.7.

If the near field, in the source-receiver constellation shall be considered, too, the third part of the differential equation is further analysed (with the constant C derived from the polar angle equation).

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial R}{\partial r} \right) + k^2 R - \frac{n(n+1)}{r^2} R = 0 \quad (2.39)$$

the solutions for incoming and outgoing waves of which are linear combinations of Bessel and Neumann functions, called spherical Hankel functions of 1st and 2nd kind and n th order

$$R(r) = h_n^{(1)}(kr) + h_n^{(2)}(kr). \quad (2.40)$$

The spherical modes in all three coordinates are finally given by

$$\Psi_n^m(k, r, \vartheta, \varphi) = \left(h_n^{(1)}(kr) + h_n^{(2)}(kr) \right) P_n^m(\cos \vartheta) e^{im\varphi} \quad (2.41)$$

of order n and degree m separated into the even and odd components. The modes are orthogonal:

$$Y_n^m Y_{n'}^{m'} dV = \begin{cases} 0 & \text{for } n \neq n' \vee m \neq m' \\ 1 & \text{for } n = n' \wedge m = m' \end{cases}, \quad (2.42)$$

with $dV = \sin \vartheta d\vartheta d\varphi$.

Directivity functions obtained by measurement or calculation can now be decomposed into series of spherical harmonics (Fig. 2.4). Accordingly the complexity of spatial sound radiation can, for instance, be studied, encoded (see Sect. 8.2) or modified by spatial filtering. This feature offers hence very interesting aspects for the description of a sound source. For the purpose of simulation, a directional source can be represented by its analytical components of the spherical harmonics, and thus the sound field expansion can be calculated according to the basic monopole, dipole, etc. response and later superposed to the response corresponding to the specific directional pattern.

2.7 Spherical Harmonics (SH) Transformation

In the previous section, it was shown that the set of spherical harmonics is an orthogonal basis, which solves the wave equation. Thus, any function $f(\vartheta, \varphi)$ of incoming or outgoing waves in spatial coordinates can be decomposed into its content expressed in weighted sums of spherical harmonics (Fig. 2.5).

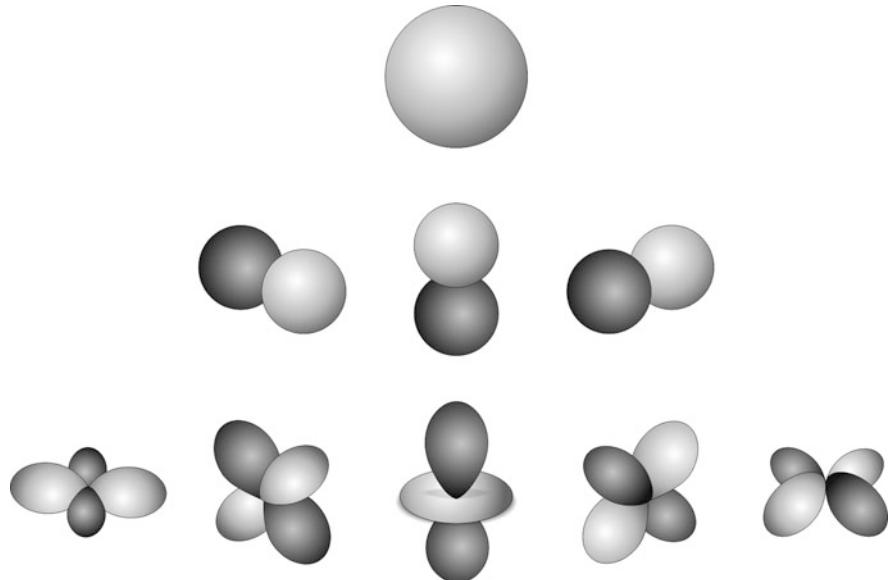


Fig. 2.4 Spherical harmonics

This interpretation is equivalent to the expression of temporal functions (signals) in amplitudes and phases of the harmonic functions (sine and cosine) in the Fourier transformation; see Sect. 7.4. The core of the transformation property is the orthonormality of the base functions:

$$\oint_{\text{unit sphere}} Y_n^m(\vartheta, \varphi) Y_{n'}^{m'*}(\vartheta, \varphi) d\Omega = \delta_{nn'} \delta_{mm'} \quad (2.43)$$

and the finite energy of the function to be transformed (Williams 1999).

$$\oint_{\text{unit sphere}} |f(\vartheta, \varphi)|^2 d\Omega < \infty \quad (2.44)$$

Accordingly, the specific SH coefficients, \$f_{nm}\$, related to the directional function \$f(\vartheta, \varphi)\$ can be found, which introduce weights to the spherical harmonics and finally sum up to the original function \$f(\vartheta, \varphi)\$.

$$f_{nm} = \oint_S f(\vartheta, \varphi) Y_n^m(\vartheta, \varphi) dS \quad (2.45)$$

with \$dS = \sin\vartheta d\vartheta d\varphi\$. This equation is called spherical harmonic transform, SHT.

The inverse transformation, ISHT, reads

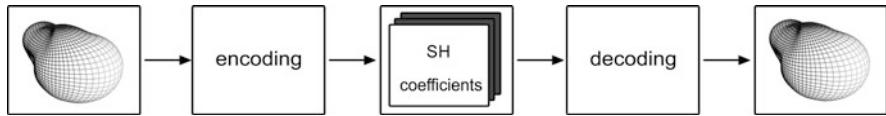


Fig. 2.5 Source directivity encoding in SH coefficients

$$f(\vartheta, \varphi) = \sum_{n=0}^{\infty} \sum_{m=-n}^n f_{nm} Y_n^m(\vartheta, \varphi). \quad (2.46)$$

The SH coefficients f_{nm} may represent normalized directivity functions of sources and receivers, $\Gamma(\vartheta, \varphi)$, they may also represent an outgoing sound pressure wave field $p(r, \vartheta, \varphi, t)$, if appropriate driving coefficients for the source strength are introduced, for example, radiated spherical “modes”:

$$p_{nm}(k, r) = f_{nm}(k) h_n^{(2)}(kr). \quad (2.47)$$

2.8 Plane Wave Representation with Spherical Harmonics

For a given receiver point, it can be assumed that a plane wave in free field arrives at the angles ϑ and φ . Expanded in a spherical harmonics decomposition reads (Williams 1999)

$$p(r, k, \vartheta, \varphi) = 4\pi \sum_{n=0}^{\infty} \sqrt{-1}^n j_n(kr) \sum_{m=-n}^n Y_n^m(\vartheta, \varphi) Y_n^{m*}(\vartheta, \varphi) \quad (2.48)$$

This equation is of fundamental importance for the mathematical description of surround sound systems. Surrounding arrays of loudspeakers can be used to reconstruct such a plane wave if the loudspeakers play signals which correspond in their sum to the spherical harmonics series. The series, however, can be approximated only in truncated form. A very important application for decomposing incoming wave fields into SH coefficients and reproduction of those wave fields by surround loudspeaker arrays is the “Ambisonics” technique; see Sect. 17.3.2.

2.9 Multipole Synthesis

The prerequisite of multipole synthesis is that sound fields can be developed into a series of spherical harmonics (Sect. 2.6) or of multipoles. The aim of this procedure is coding of measured or calculated directivity patterns. In simulations the sets of multipole coefficients are used for reconstruction of the original source directivity. This is achieved by linear superposition. The difference between a spherical harmonic superposition and a multipole synthesis is the geometric reference. Spherical harmonics strictly are related to polar coordinates and a source origin in one point, whereas multipoles can be distributed in space, so that more degrees of freedom in sound field approximation of arbitrary sources can be used as synthesis parameters.

Due to the fact that multipoles are special solutions of spherical harmonics, the principle will yield identical results for sources of spherical geometry. The decomposition of the sound radiation problem into spherical functions allows the a posteriori reconstruction of the directivity. The reconstructed results are best matching the original source directivity for sources of spherical geometry. Furthermore, it can be used as a general basis for other series of elementary radiators such as monopole and multipole sources and combinations of those.

The total pressure field is formed by superposition of spherical wave functions with proper relative amplitudes. For a complete set of field synthesis coefficients, the velocity is of interest, too. Thus not only the continuous pressure but also its gradient can be optimized regarding approximation and interpolation of arbitrary source data. We obtain the velocity addressed to spherical wave functions by using the pressure gradient. As mentioned above, the straightforward expansion of the sound field in spherical wave functions can produce severe problems of poor convergence, if the source is of non-spherical symmetry. In this case the required order of SH coefficients may be very high. Multipoles can be placed, however, at various positions. With an expansion of the sound pressure measured at some reference points into coefficients of a set of multipoles the field can be reconstructed to achieve an approximation of the reference sound pressures and, furthermore, interpolated sound pressures at other positions. The same strategy, by the way, can also be applied with regard to the particle velocity. Guidelines for the number and spatial discretization of the reference points are available (Ochmann 1990).

Chapter 3

Sound Propagation



3.1 Introduction

Sound waves propagate from a source into the environment where they interact with objects. In indoor environments, sound is reflected and scattered at walls, floors, ceilings, and furniture. In outdoor environments, sound is scattered at trees, reflected at building façades or at the ground, or it may be diffracted around a noise barrier. Furthermore, properties of the medium like inhomogeneity of sound velocity and viscous effects can cause influences such as refraction and attenuation, respectively. Modeling of these phenomena is often a challenging task but nevertheless interesting, because many of the theoretical concepts involve approximations related to object dimensions in comparison with the acoustic wavelength. For audible sound these span from metres to centimetres, so there is no general solution covering all frequencies in broadband sound. Any time a sound propagation model is discussed, it must be kept in mind its limited frequency range of validity. In this chapter the fundamental effects and their quantities of sound propagation are introduced.

Sound radiation from sources often occurs in free space. This applies to the sound that reaches the receiver directly over a free line-of-sight.¹ An important law of sound propagation describes the dependence of sound intensity and sound pressure level on the propagation distance. The following table summarizes several basic distance laws (Table 3.1).

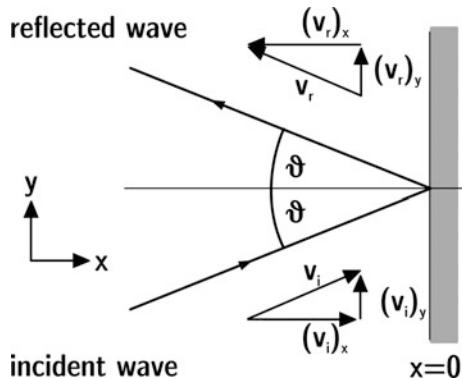
From here, we proceed by assuming a wave which has almost plane wavefronts, and in particular, the wave impedance is in very good approximation equal to the characteristic impedance of the medium, Z_0 . This condition is generally justified in the far field, i.e. in a distance corresponding to $kr \gg 1$.

¹Obviously this applies usually to outdoor sound propagation or to sound propagation in close distance between source and receiver, so that no obstacle is in the propagation path.

Table 3.1 Free-field propagation from elementary sources in direction of maximum sound radiation

Type of source	Distance law of sound pressure level	Level reduction per distance doubling, dB
Monopole	$L = L_w - 20 \log r - 11$	6
Multipole or other directional source, on axis	$L = L_w - 20 \log r - 11 + L_D$	6
Incoherent line above ground ($P' = P/\Delta x$), $L_w - 10 \log \Delta x/1m$	$L = L'_w - 10 \log r - 3$	3

Fig. 3.1 Reflection of plane waves at impedance plane



3.2 Reflection of Plane Waves at an Impedance Plane

A plane wave incident on a (infinitely) large smooth wall is reflected according to Snell's law. This can be described as a "specular" reflection (Fig. 3.1).

The amplitude might be reduced and the phase changed. If the wave is incident at the angle ϑ

$$\underline{p}_i(x, y, t) = \hat{p} e^{j(\omega t - kx \cos \vartheta - ky \sin \vartheta)} \quad (3.1)$$

the reflected wave is

$$\underline{p}_r(x, y, t) = \hat{p} R e^{j(\omega t + kx \cos \vartheta - ky \sin \vartheta)}, \quad (3.2)$$

with R denoting the reflection factor, $R = |R|e^{j\gamma}$. It is related to the wall impedance, \underline{Z} , by

$$\underline{R} = \frac{\underline{p}_r}{\underline{p}_i} = \frac{\underline{Z} \cos \vartheta - Z_0}{\underline{Z} \cos \vartheta + Z_0}. \quad (3.3)$$

$Z_0 = \rho_0 c$ is the characteristic impedance of air. The wall impedance, \underline{Z} , is defined as the ratio of sound pressure to the normal component of particle velocity, both determined at the wall. The impedance may be independent of the angle of incidence. This phenomenon is called a “local reaction”. Due to this local reaction adjacent sections of the same wall surface are independent from each other, so that no tangential waves are transmitted along the wall surface. This is a good approximation for heavy walls, for walls with low bending stiffness (see Sect. 5.3) and for porous absorbers with high flow resistivity.

The absorption coefficient, α , is particularly important

$$\alpha = \frac{\left| \underline{p}_i \right|^2 - \left| \underline{p}_r \right|^2}{\left| \underline{p}_i \right|^2} = 1 - |\underline{R}|^2 \quad (3.4)$$

and the specific impedance at the surface

$$\underline{\zeta} = \frac{1}{Z_0} \left(\frac{\underline{p}}{\underline{v}_n} \right)_w = \frac{1}{\cos \vartheta} \frac{1 + \underline{R}}{1 - \underline{R}} \quad (3.5)$$

The absorption coefficient for locally reacting surfaces is

$$\alpha = \frac{4 \operatorname{Re} \{ \underline{\zeta} \} \cos \vartheta}{1 + 2 \operatorname{Re} \{ \underline{\zeta} \} \cos \vartheta + |\underline{\zeta}|^2 \cos^2 \vartheta}. \quad (3.6)$$

3.2.1 Examples of Wall Impedances

We can define some extreme cases for perpendicular incidence. Real materials approximate these conditions quite well. A heavy concrete wall, for instance, represents a “hard wall” and the ocean surface represents a “soft wall” for underwater sound. Some other examples are listed as follows (see Kuttruff 2007).

Matched wall $\underline{Z} = Z_0; \underline{R} = 0; \alpha = 1$

Hard wall $\underline{Z} = \infty; \underline{R} = 1; \alpha = 0$

Soft wall $\underline{Z} = 0; \underline{R} = -1; \alpha = 0$

Mass Layer

At $x = 0$ we assume a layer of mass per surface area of m'' . The reaction of this layer is exclusively inertia. Accordingly the layer is characterized by neglected internal and mounting stiffness and losses. The force $p - p'$ on a surface area of 1 m^2 excites the layer to vibrations with the velocity $v' = v, p$, v denotes the pressure and particle velocity on the incident side and p', v' on the transmission side of the mass layer (Fig. 3.2).

One can easily show that

$$Z = \left(\frac{p}{v} \right)_w = Z_0 + j\omega m''. \quad (3.7)$$

The absorption coefficient is

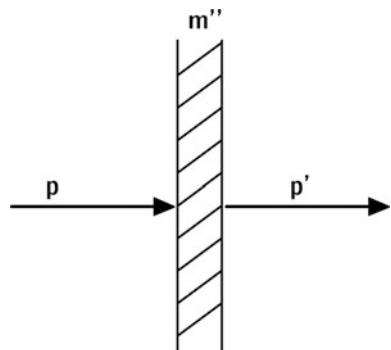
$$\alpha = \frac{1}{1 + \left(\frac{\omega m''}{2Z_0} \right)^2}. \quad (3.8)$$

If $\omega m'' \gg 2Z_0$, the equation can be simplified to

$$\alpha = \left| \frac{p'}{p} \right|^2 \approx \left(\frac{2Z_0}{\omega m''} \right)^2. \quad (3.9)$$

For instance, assuming even a lightweight element like a 6 mm glass pane with $m'' = 15 \text{ kg/m}^2$, the term in brackets in Eq. (3.8) already exceeds a value of 10 at frequencies above 30 Hz.

Fig. 3.2 Sound hitting a mass layer



Mass Layer in Front of Hard Wall

In the next section the combined impedance from a mass layer mounted with an air gap to the rigid wall is discussed. At first, the air gap alone is analyzed. Defining a new reference plane for $x = 0$ results in an air gap. It is necessary to shift the reference plane if another object like porous fabric is placed at $x = 0$ (see below) (Fig. 3.3).

Before returning to $x = 0$ the reflected sound wave travels twice the distance d . The reflection factor is thus

$$\underline{R} = |\underline{R}|e^{-2jkd}, |\underline{R}| = 1 \quad (3.10)$$

and

$$\underline{Z} = -jZ_0 \cot kd \quad (3.11)$$

when $kd \ll 1$, meaning the air gap is much smaller than the wavelength, we can approximate the cotangent function by $\cot kd \approx 1/kd$

$$\underline{Z} \approx \frac{\rho_0 c}{jkd} = \frac{Z_0 c}{j\omega d}. \quad (3.12)$$

In this frequency range, the air gap apparently reacts as a spring with the stiffness per m^2

$$s'' = \frac{1}{n''} = \frac{Z_0 c}{d} \quad (3.13)$$

Fig. 3.3 Sound incident on virtual air gap in front of a hard wall

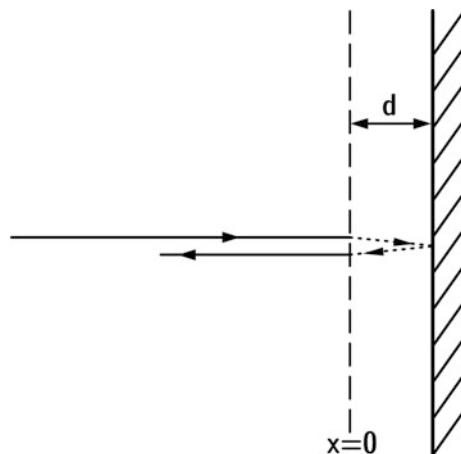
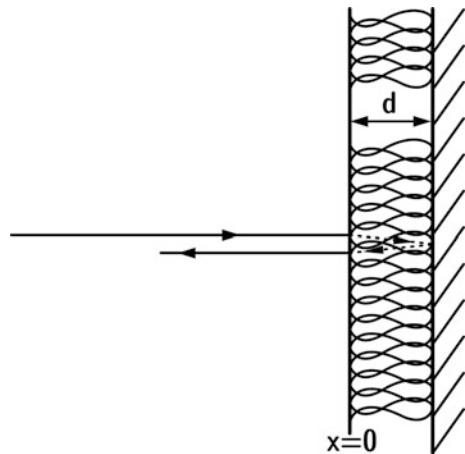


Fig. 3.4 Sound hitting a mass layer with air gap and porous filling



We now add the mass layer at $x = 0$ by adding its impedance $j\omega m''$. Losses are added by a flow resistivity per m^2 , w'' , representing porous material placed in the air gap (Fig. 3.4).

$$\underline{Z} = w' + j \left(\omega m'' - \frac{Z_0 c}{\omega d} \right) \quad (3.14)$$

The reader may recognize that this impedance belongs to a resonator with resonance frequency

$$\omega_0 = c \sqrt{\frac{\rho_0}{m'' d}} \quad (3.15)$$

Porous Layer in Front of a Hard Wall

The impedance of the air gap is now combined with a purely resistive component, w'' (Fig. 3.5)

$$\underline{Z} = w'' - j Z_0 \cot kd. \quad (3.16)$$

Fig. 3.5 Sound hitting a curtain in front of a wall

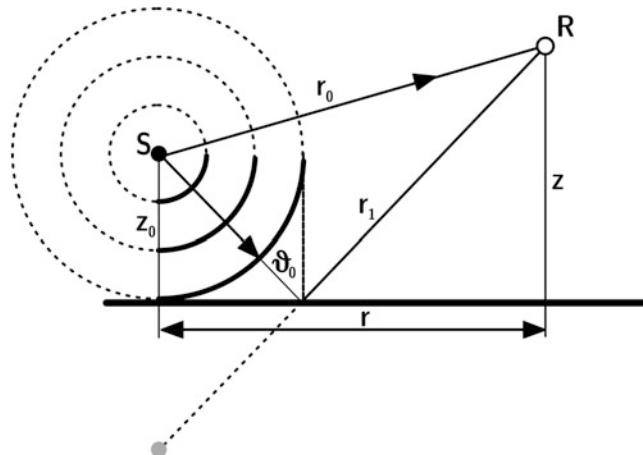
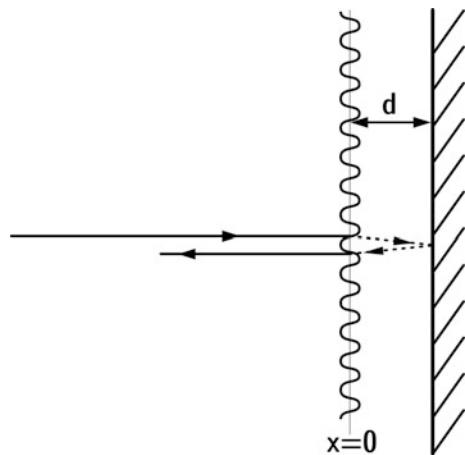


Fig. 3.6 Spherical wave reflection above impedance plane

3.3 Spherical Wave Above Impedance Plane

A point source at S radiates a spherical wave (Fig. 3.6). The total field at R contains a contribution from the direct sound travelling along the vector \vec{r}_0 and another component reflected from the plane. A plane wave incidence and a spherical wave incidence differ significantly as the impedance plane is hit at various angles and the total reflection must be integrated:

$$p = \frac{j\omega\rho_0\hat{Q}e^{-jkr_0}}{4\pi r_0} - \frac{\omega^2\rho_0\hat{Q}}{4\pi c} \int_{\vartheta} J_0(kr \sin \vartheta) e^{-jk(z+z_0)\cos \vartheta} \underline{R}(\vartheta) \sin \vartheta d\vartheta \quad (3.17)$$

The first term represents the direct sound and the latter the contribution of the reflection. Note that the reflection term is an integral over various angles between the wavefront and the surface. An approximation of the entire expression assuming a constant angle of incidence, ϑ_0 , is

$$p = \frac{j\omega\rho_0\hat{Q}e^{-jkr_0}}{4\pi r_0} + \frac{j\omega\rho_0\hat{Q}e^{-jkr_1}}{4\pi r_1} \underline{R}(\vartheta_0). \quad (3.18)$$

This equation assumes a constant angle and, thus, a plane wave. Now, the contribution of the reflection can be related to another point source, called “image source” (see also Sect. 11.4), which is apparently located below the surface radiating a spherical wave whose amplitude is reduced by $\underline{R}(\vartheta_0)$. Even though a spherical wave is present, the reflection is calculated for a constant angle, ϑ_0 . The sound field model is apparently switched to a plane wave,² when reflection occurs. This “image source model” is even accurate if $R = 1$ or $R = -1$. For other reflection factors, the approximation was deemed sufficiently accurate in case of z, z_0 and $r \gg \lambda$ (Suh and Nelson 1999), see also Sect. 11.4.3.

In a reflecting plane, plane waves or spherical waves can be divided into elementary sources (Huygens principle). Together these sources together build up an interference field which produces a plane or spherical wave. This model provides the basis for the simplicity in Eq. (3.18) (which is in fact exact in case $|R| = 1$).

If, however, the plane has discontinuities in terms of its geometry or impedance, the Huygens superposition will be disturbed and the reflected field shows scattering and diffraction. Scattering is related to the reflection by an object, while diffraction is related to the boundary of an object.

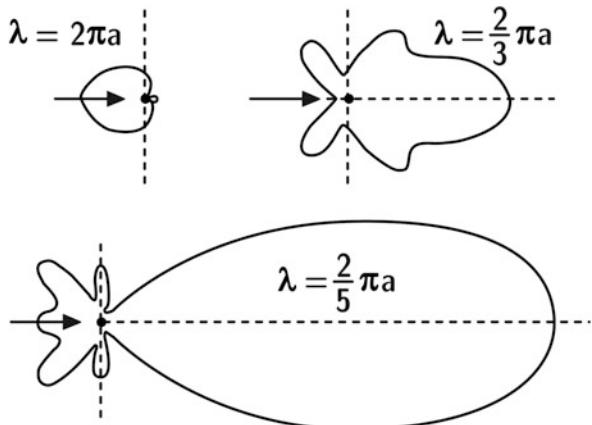
3.4 Scattering

3.4.1 Object Scattering

Sound waves may hit obstacles. Depending on the size of these objects compared with the wavelength, the scattered field has large amplitudes in forward direction (“forward scattering”), in reverse direction (“reflection”) or any in other direction following a specific distribution. The exact formulation and solution of the scattered field amplitude is a difficult problem, except for academic cases where objects like spheres, cylinders, etc. are studied (Fig. 3.7).

²plane wave approximation.

Fig. 3.7 Sound scattering at a sphere with radius a (see Morse and Ingard 1968)



An efficient strategy to address the problem in practical cases is to map and to compare the scattered field to an equivalent field created by a spherical scatterer. This approach makes it possible to define the scattering cross section. With reference to an incidence plane wave with intensity I_0 the scattering cross section is related to the power of the scattered field, P_s , ($\lambda \ll a$)

$$Q = \frac{P_s}{2I_0}. \quad (3.19)$$

In a more general approach, the theoretical model is based on superposition of the undisturbed incident field, p_0 , and the scattered wave, p_s ,

$$p = p_0 + p_s. \quad (3.20)$$

p_s must fulfill the boundary condition at the object's surface. With $Z = \infty$ the normal component of the particle velocity must be zero, and, thus, the normal component of the pressure gradient, too. From this follows

$$\frac{\partial p_s}{\partial n} = \frac{\partial p_0}{\partial n} \quad (3.21)$$

and for the particle velocity of the scattered wave

$$(v_s)_n = \frac{1}{j\omega\rho_0} \frac{\partial p_s}{\partial n}. \quad (3.22)$$

The radiated field generated by this velocity distribution is to be calculated by using the models described in [Sect. 2.5](#) for an equivalent radiation problem.

3.4.2 Surface Scattering

Sound reflection at rough, corrugated walls is also described by scattering. We still assume a large wall, but its surface corrugations in length and depth are not small compared with the wavelength. A plane wave incident on the wall will interact so that the local phases of the elementary (Huygens) sources form a complicated total field of scattered sound.

As illustrated in Fig. 3.8, the wall can be assumed smooth at low frequencies if the depth, h , and the length, a , of the corrugation profile is significantly smaller than $\lambda/2$. When corrugation dimensions are in the order of magnitude of the wavelength, a complicated scattered field will develop. At high frequencies the fine structure of the corrugations will lead to a specular type of reflection again.

Note that scattering from a rough surface may lead to sound paths with oblique angles of incidence and reflection and with retarded arrival, compared with the specular sound path (Fig. 3.9).

Analytic or numerical solutions are available for some corrugation types. The bandwidth of the sound waves determines whether the total scattered field at the observation point can be approximated by energetic formulations or if distinct spectral and directional scattering lobes will occur (Cox and D'Antonio 2004). Scattering theory is best understood when it is expanded into spatial wave

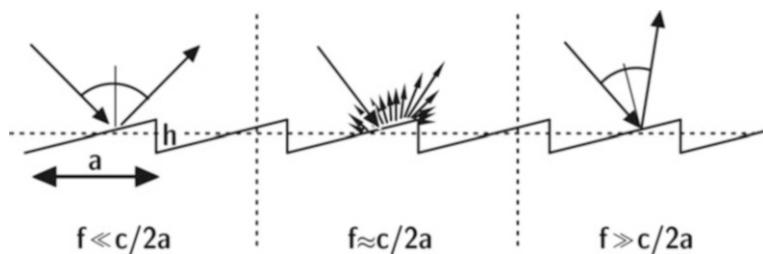


Fig. 3.8 Scattering caused by surface corrugations

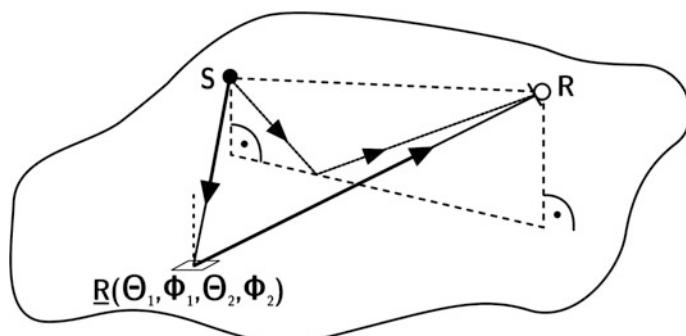
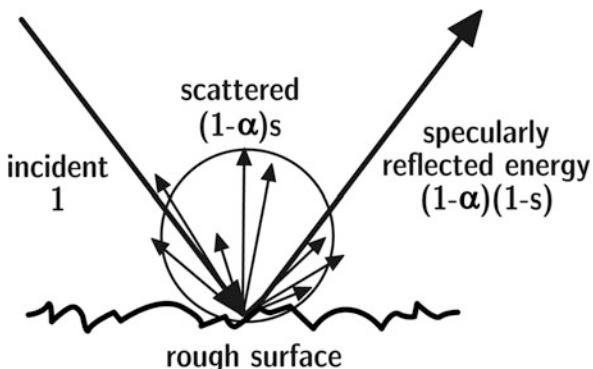


Fig. 3.9 Delayed sound paths from surface scattering

Fig. 3.10 Energy reflected from a corrugated surface into a scattered and specularly reflected portion. Definition of the total reflected energy $(1 - \alpha)$, the scattered energy $(1 - \alpha)s$ and specularly reflected energy $(1 - \alpha)(1 - s)$ (Vorländer and Mommertz 2000)



decomposition, the zero order component (zero-order lobe) representing the specular reflection component. The higher-order lobes direct the sound to non-specular directions.

The energies of reflections are normalized with respect to the incident plane wave, as shown in Fig. 3.10.

$$E_{\text{spec}} = (1 - \alpha)(1 - s) \equiv (1 - a), E_{\text{total}} = (1 - \alpha) \quad (3.23)$$

a is the “specular absorption coefficient”. It is an apparent absorption coefficient because the energy is scattered rather than being absorbed. From these equations the energy portion scattered, s , can be determined by

$$s = \frac{a - \alpha}{1 - \alpha} = 1 - \frac{E_{\text{spec}}}{E_{\text{total}}} \quad (3.24)$$

Furthermore, measurement data is available which serves as input data for simulation software (see Sect. 11.3 and Annex). It describes the energetic amount of scattering (scattering coefficient) compared with the zero-order scattering lobe (specular component) (Fig. 3.11).

The uniformity of the directional scattering distribution (diffusion coefficient) is also of interest, but this should not be mixed up with the scattering coefficient (Cox and D’Antonio 2004; Cox et al. 2006). The approach taken to define the directional distribution of the sound scattered from the surface or the object can be compared to the test that is used to assess the uniformity of loudspeaker radiation. Thus, a free-field polar response must be calculated or measured. The diffusion coefficient is then a single figure describing the uniformity of the polar response. If the energy is scattered uniformly in all directions, then the diffusion coefficient is one. If all the energy is scattered in one direction, then the diffusion coefficient is zero. The diffusion coefficient is usually determined in one-third octave bands and is frequency dependent.

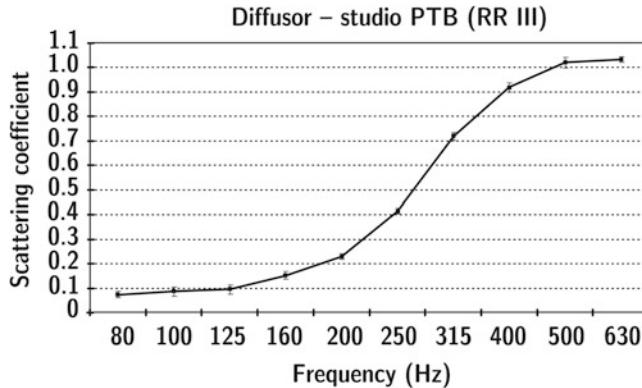


Fig. 3.11 Random-incidence scattering coefficients measured on a scale-model sample of a diffusing ceiling structure (PTB, Round Robin III, used in Bork 2005a)

The limiting case is the ideal diffuse reflection according to Lambert's cosine law. The intensity of a Lambert scatterer depends on the cosine of the scattering angle, ϑ , and the distance from the wall element dS . It is independent of the angle of incidence.

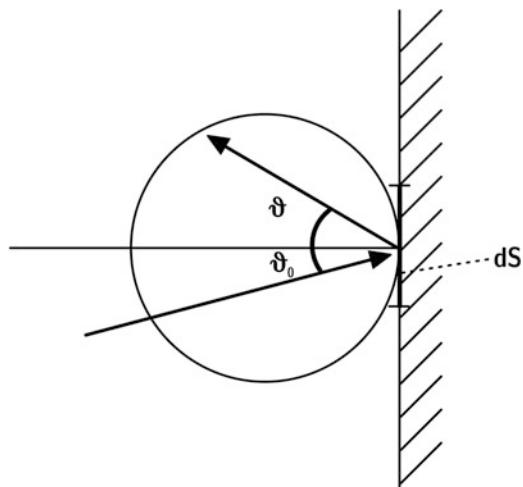
$$\left| \vec{I}(\vartheta) \right| = (1 - \alpha) \frac{B dS}{\pi r^2} \cos \vartheta \quad (3.25)$$

B denotes the irradiation strength on the wall; α the part of the incident energy BdS which is not reflected from the wall. This kind of scattering distribution creates a constant illumination effect on a detector with a fixed sensor area, like a membrane or, in the optical analogy, like a camera or an eye with fixed aperture. A white sheet of paper thus seems to have a brightness independent of the observation angle. Another example of a Lambert scatterer is the light-reflecting moon which looks more like a disc than a sphere, since the light scattered from the surface near the moon's apparent midpoint (at normal direction towards the observer) has the same effective brightness than the light scattered from the apparent circumference, with a observation angle of almost $\pi/2$ (Fig. 3.12).

3.5 Diffraction

Diffraction occurs at objects with free edges, at corners and edges in a room, or at boundaries between materials with two different impedances. The diffraction wave is apparently radiated from the edges or perimeter of an object. Its intensity is negligibly small if the object is small compared with the wavelength. In this case the incident wave remains unaffected. As the object gets larger compared to the wavelength, first a shadow region appears which subsequently becomes clearer and

Fig. 3.12 Probability distribution of scattered sound (Lambert's law)



sharper. The shadow results from a total cancellation of the incident wave by the diffraction wave.

As for the scattering, the calculation of the diffracted field is analytically possible for simple geometrical obstacles (spheres, circular discs, cylinders, free edges on a screen, slits or holes in a screen, etc.). As the sizes of many objects encountered in daily life are in the order of magnitude of the wavelength, diffraction is easily noticeable. It is, however, hard to predict its effect quantitatively by calculation. Diffraction influences the binaural hearing, the sound transmission through doors or windows when they are not completely sealed and the orchestra sound from an orchestra pit in an opera house.

The mathematical basis of the diffraction phenomenon has already started in the far past by Huygens, Fresnel, and Kirchhoff. While first formulations were initially promoted by the optical characteristics of electro-magnetic waves, in 1896 Sommerfeld also dealt with the question of describing the diffraction around objects that are, due to their geometrical dimension, settled in the range of wavelengths. The geometric theory of diffraction was formulated by (Keller 1962), which is closely linked to the description of “rays”, which split-up close to edges of diffraction. The publication of (Biot and Tolstoy 1957) can be counted as a milestone within the “unified” theory of diffraction because first they describe the exact composition of the direction of arrival through a frequency dependent transfer path along the wave front in a closed form, see 3.4.1. This achievement cleared the way for a series of comparisons and verifications of the approaches of the different theories with measurements. Medwin expedited the Biot-Tolstoy method further (Medwin 1981). In his publication on the diffraction of sound in the acoustical shadow zone, the author could present a closed form of the Biot-Tolstoy formula, which can be applied to attenuate sound barriers with arbitrary angle of aperture and which, in contrast to the original view, were finite in their dimension (Biot-Tolstoy-Medwin approach).

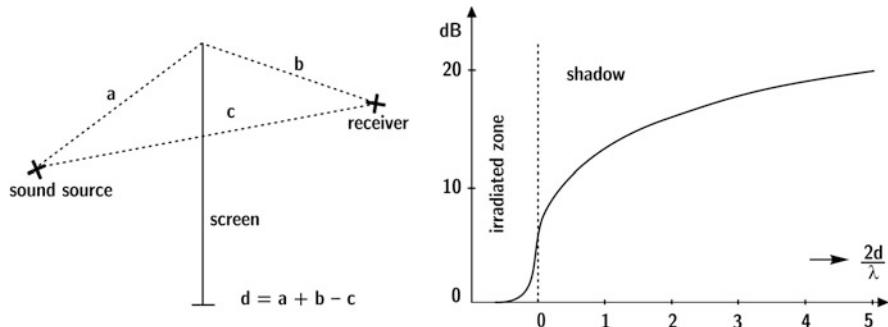


Fig. 3.13 Estimating the insertion loss of a screen

For acoustics in practise, especially Maekawa's work (Maekawa 1968) is very important, which, through experimental ascertainment, frequency dependently determines the diffraction of sound by an infinite half-plane approximating the reduction of the sound energy. The result is solely based on geometrical operations to identify the source's way around the boundary to the receiver. The approach was then extended and applied to all kinds of transmission paths such as sets of houses or noise barriers, whose amount of energy is subsequently accumulated as issued by ISO 9613-2. To establish simplified procedures with reasonable mistakes of approximation, modifications that combine theory and empiricism also rely on the approach of the specification of shadowing through sound barriers, for example that by (Pierce 1974).

Diffraction models must be taken into account when the sound propagation is, for instance, predicted for large distances in urban areas or in open-plan offices. Noise barriers are a typical example for the application of engineering models of diffraction. With Maekawa's model, the diffracted wave from the edge and corresponding insertion loss of a vertical screen can be approximated by using the detour, d , of the diffraction (see Fig. 3.13).

$$\Delta L \approx 10 \log \left(2\pi^2 \frac{d}{\lambda} \right) \quad (3.26)$$

3.5.1 Unified Theory of Diffraction

While Maekawa's by now well-established model, usually referred to as "rubber band model", is widely used in simulation software. (Svensson et al. 1999) could significantly ameliorate the mathematical description of the effects of diffraction in the range of time and frequency. His work on the analytical secondary source model of diffraction of sound through a closed form description – which is based on the

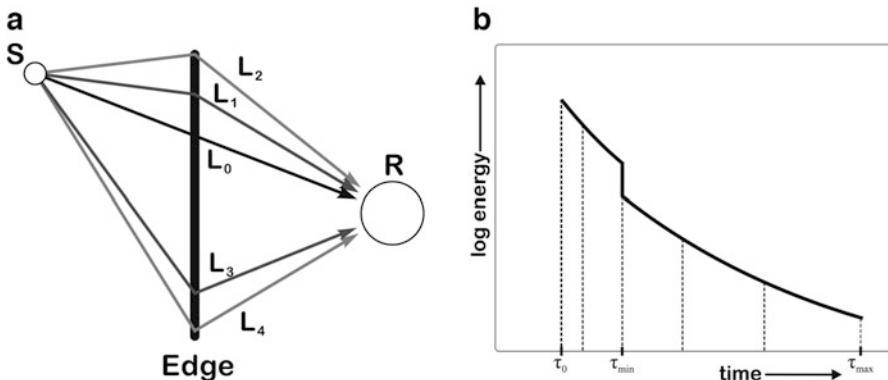


Fig. 3.14 (a) Sound paths from S to R via an edge. L_0 describes the shortest path over the edge, the other paths have longer detours over the edge. (b) Energetic edge diffraction impulse response. τ_0 is the delay of the shortest path, τ_{\min} and τ_{\max} describe the delays over the edge's starting point and end point (after Schröder 2011)

Biot-Tolstoy-Medwin method and which includes finite edges through a line integral on its apex – is another milestone in the development of diffraction, leading to the formulation of the Unified Theory of Diffraction, UTD.

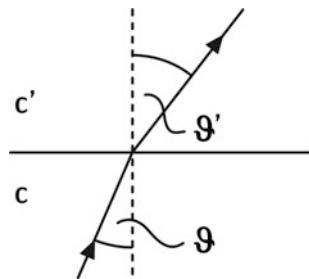
The key of UTD is that the edge can be thought of as a line of secondary point sources. The paths over the edge are taken into account, which results in the integral of all secondary sources along the edge line. The total time-discrete and decaying impulse response is then determined by integrating the contributions each path over the finite edge (Fig. 3.14).

Up to today, this achievement can be seen as the exact and elegant solution for the description of directivity around a finite diffraction edge that abdicates discrete scanning and develops a continuous impulse response. Several publications followed the Biot-Tolstoy-Medwin, now extended by Svensson, which for instance successfully integrated the approach into room acoustic simulation by means of algorithms such as mirror sound sources and ray-based approaches. (Calamia and Svensson 2007) developed a discrete procedure based on the continuous integral form that diminishes the computing time, yet maintaining a similar quality.

3.6 Refraction

Fermat's principle says that sound waves take the path with the shortest travel time. For transmission of a plane sound wave from air with characteristic impedance Z_0 into another medium with characteristic impedance Z' , we use the refraction index

Fig. 3.15 Refraction at the boundary between two media



$$n = \frac{c'}{c} = \frac{\sin \theta'}{\sin \theta} \quad (3.27)$$

for calculation of the geometric conditions (Fig. 3.15).

The amplitude of the refracted wave into the medium with Z' follows from

$$T = \frac{2Z'}{Z' + Z_0}, \quad (3.28)$$

while for the reflection factor Eq. (3.3) still holds.

This approach can also be used for derivations of curved transmission paths in layered media. If the sound speed changes gradually, for instance due to changing temperatures at various atmospheric elevations or different temperature and salt concentration in the ocean, the effect can be described by using small layers of constant sound speed. For the boundary between two adjacent layers with sound speed c and $c + dc$

$$\frac{c + dc}{c} = \frac{\sin(\vartheta + d\vartheta)}{\sin \vartheta} \approx 1 + \frac{\cos \vartheta}{\sin \vartheta} d\vartheta. \quad (3.29)$$

This effect corresponds to a curved sound path with reciprocal radius

$$\frac{1}{r_{\text{curvature}}} = \frac{1}{c} \frac{\partial c}{\partial n}, \quad (3.30)$$

with n denoting the normal direction of the sound wave. The curvature is the bigger the larger the sound speed gradient is in the direction normal to the propagation. This effect can lead to curved sound rays in the atmosphere. Under certain weather conditions (temperature increasing with height), upwards radiated sound may be bent down to reach the ground again. The long-distance sound propagation is affected greatly. The same can be observed with wind speed profiles. When it comes to outdoor sound propagation, refraction is therefore an essential aspect of sound field modelling.

3.7 Attenuation

Another effect of long distance sound propagation is attenuation. It should be noted that long distance sound propagation may occur outdoors as well as indoors. If a sound wave is observed during its propagation for some seconds, it is clear that it has travelled several hundred metres, independent of whether the problem occurs outdoors or indoors, the latter involving numerous wall reflections.

Several attenuation effects lead to a complex wave number, \underline{k}' , and to an exponential decrease of the sound pressure and intensity, described by

$$\underline{p}(x, t) = \hat{p} e^{-\frac{m}{2}x} e^{j(\omega t - \underline{k}' x)} = \hat{p} e^{j(\omega t - \underline{k}' x)}, \quad (3.31)$$

with m denoting the energetic attenuation coefficient and

$$\underline{k}' = \frac{\omega}{c} - j \frac{m}{2} \quad (3.32)$$

The intensity along the x coordinate in a plane wave is

$$I(x) = I_0 e^{-mx} \quad (3.33)$$

or expressed in effective level loss,

$$D = 4.34 m \frac{dB}{m}, \quad (3.34)$$

with the level loss after propagation over the distance x

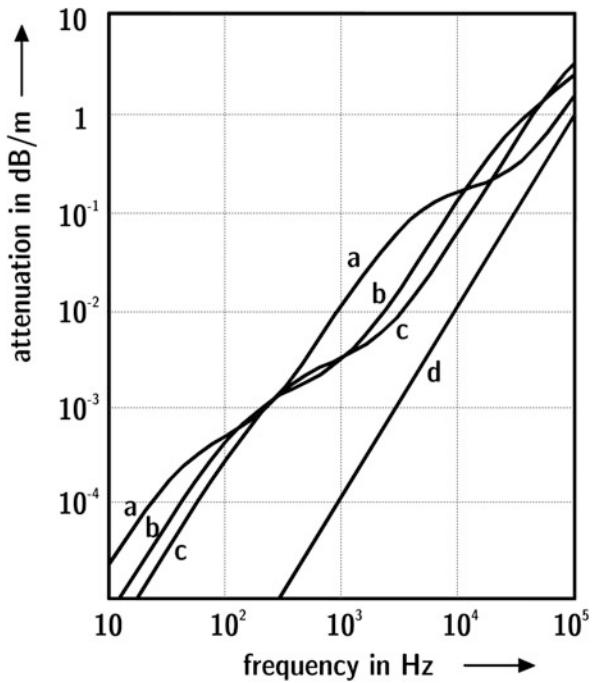
$$L = L_0 - D \cdot x. \quad (3.35)$$

Attenuation is caused by viscosity, heat conduction and thermal relaxation (Bass et al. 1995). All effects irreversibly extract energy from the sound wave and feed other energy reservoirs, for instance translational, rotational or vibratory modes of water molecules. The amount of water in the medium – humidity – has a crucial influence on the attenuation (Fig. 3.16).

3.8 Doppler Effect

Non-experts know very little about acoustics. One acoustic effect, however, is very well known: the Doppler effect. The Doppler effect is the frequency shift perceived when police cars or fire-brigades are passing by at high speeds blowing their horns. Moving sound sources or receivers cause a change in the received rate of sound

Fig. 3.16 Typical attenuation curves (see Kuttruff 2007) of humid air
 (a) 10%, (b) 40%, (c) 80%,
 (d) classical theory ($\sim \omega^2$)



pressure maxima and minima and, thus of frequency at the receiver. In contrast to electromagnetic waves, the acoustic Doppler effect depends on the actual movement and not just on the relative movement. If the receiver is moving with velocity V towards the source, the received frequency is higher than the radiated. Inserting $x = x_0 - Vt$ into a harmonic sound pressure equation yields

$$\underline{p} = \hat{p} e^{j[(\omega + kV)t - kx_0]}. \quad (3.36)$$

The perceived frequency is thus

$$f' = \frac{\omega + kV}{2\pi} = f \left(1 + \frac{V}{c} \right). \quad (3.37)$$

If, however, the sound source is moving relative to the medium in direction of the receiver, at first another sound speed is to be accounted for, $c' = c - V$ and $k' = \omega / (c - V)$. This, together with the changed distance $x = x_0 - Vt$, yields

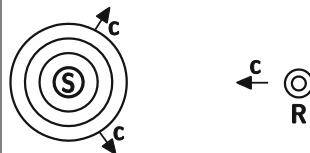
$$\underline{p} = \hat{p} e^{j[(\omega + k'V)t - k'x_0]} \quad (3.38)$$

with the effective frequency

Table 3.2 Four cases of Doppler effect at relative speed c

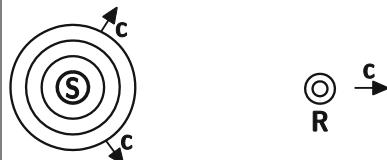
1. Receiver moving towards the source

$$f' = f \left(1 + \frac{V}{c}\right) \xrightarrow{V \rightarrow c} 2f$$



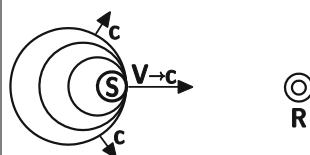
2. Receiver moving away from the source

$$f' = f \left(1 - \frac{V}{c}\right) \xrightarrow{V \rightarrow c} 0$$



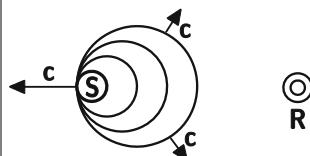
3. Source moving towards the receiver

$$f' = \frac{f}{1 - \frac{V}{c}} \xrightarrow{V \rightarrow c} \infty$$



4. Source moving away from the receiver

$$f' = \frac{f}{1 + \frac{V}{c}} \xrightarrow{V \rightarrow c} f/2$$



$$f' = \frac{\omega + k'V}{2\pi} = f \left(1 + \frac{V}{c - V}\right) = \frac{f}{1 - \frac{V}{c}}. \quad (3.39)$$

In the latter case only movement with $V < c$ leads to a registration of regular harmonic sound at the receiver. If $V > c$, the sound signal at the receiver is compressed to a shock wave. The difference between actual and relative movement can be better understood by discussing the extreme cases ($V \rightarrow c$) with respect to a harmonic signal. The rate of received pressure maxima determines the received frequency (Table 3.2).

Chapter 4

Sound Fields in Cavities and in Rooms



4.1 Introduction

The basics of sound wave radiation and propagation were discussed in the preceding chapters. Free-field propagation can be studied and simulated with these mathematical models. Numerous problems, however, are related to sound in cavities. This is true for small appliances or vehicles as well as for larger volumes such as living rooms or kitchens, or even large spaces like concert halls and opera houses. In our daily lives, we can analyse by listening whether we are in a room or not, and we can estimate the room volume and amount of absorption from to the auditory impression of the room response. Nevertheless, room acoustical fields are extremely complicated and show a detailed fine structure, depending predominantly on the size of the room. It is therefore appropriate to study at first the physical room acoustic effects based on wave theory. Then, it will be shown that perhaps not all physical details of wave fields are relevant for hearing in rooms. Approximations and considering energy models are also very useful to describe room acoustic effects and for room acoustic modelling. In this chapter, therefore, we extend the theoretical description by alternative models of statistical acoustics, diffuse sound fields and reverberation theory.

4.2 Cavities

In cavities with dimensions smaller than the wavelength, a harmonic vibration deflection creates a harmonic pressure variation (Fig. 4.1).

In a linear formulation of small amplitudes, the sound pressure amplitude is proportional to the relative volume variation.

Fig. 4.1 Small cavity excited by piston

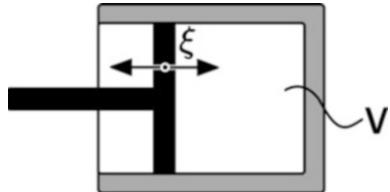
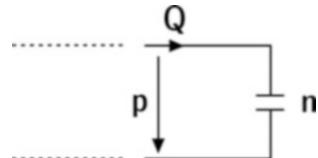


Fig. 4.2 Input impedance of a small cavity



$$\frac{\hat{p}}{p_0} = \frac{\hat{V}}{V_0} \quad (4.1)$$

The volume variation, for instance, can be created by a piston with an area S which vibrates with an amplitude ξ . Due to its small dimensions, sound pressure and particle velocity are constant in the cavity. The equation of the sound pressure in relation to the particle velocity and also the impedance, p/v , can be easily calculated.

$$p = \frac{p_0}{V_0} S \xi = \frac{\rho_0 c^2}{j\omega V_0} S v = \frac{\rho_0 c^2}{j\omega V_0} Q \quad (4.2)$$

The acoustic impedance of the cavity reacting to the piston is thus spring-like (see also [Eq. \(3.13\)](#))

$$\underline{Z} = \frac{1}{j\omega n} \quad (4.3)$$

with

$$n = \frac{V_0}{\rho_0 c^2 S} \quad (4.4)$$

According to this elementary relation, the sound pressure signal created by a volume source can be modelled by using a simple spring two-port ([Fig. 4.2](#)).

4.3 Modes

When the cavity is not small compared with the wavelength, a correct calculation of sound fields in cavities or rooms must be based on the wave Eq. (1.14). With a proper representation of boundary conditions in terms of its geometry and impedances, it can be solved, at least in principle. It must be decided in which domain the results shall be discussed. It is suitable to study the sound field in space and the spatial sound pressure distribution to determine its modes. Using a harmonic volume source excitation $q = \hat{q}e^{j\omega t}$, the wave equation can be transformed into the Helmholtz equation (stationary case)

$$\Delta p + k^2 p = -jkZ_0 q. \quad (4.5)$$

The eigenvalues, k , of the Helmholtz equation must then be determined so that they fulfil the differential equation and the boundary conditions. The solution is a discrete series of k_{lmn} related to discrete eigenfrequencies, f_{lmn} :

$$f_{lmn} = \frac{c}{2\pi} k_{lmn}. \quad (4.6)$$

Each eigenfrequency can be assigned to a specific standing waveform: a mode. This solution can be found easily in case of mathematically elementary geometrical structures like rectangular boxes, cylinders or spheres. For example in case of the rectangular box with its dimensions L_x , L_y and L_z and $Z = \infty$ throughout the boundaries we find a set of orthogonal mode functions

$$\psi_{lmn}(x, y, z) = \cos\left(\frac{l\pi x}{L_x}\right) \cos\left(\frac{m\pi y}{L_y}\right) \cos\left(\frac{n\pi z}{L_z}\right) \quad (4.7)$$

and

$$f_{lmn} = \frac{c}{2} \sqrt{\left(\frac{l}{L_x}\right)^2 + \left(\frac{m}{L_y}\right)^2 + \left(\frac{n}{L_z}\right)^2} \quad (4.8)$$

with l , m and $n \in \mathbb{N}_0$.

In room acoustics, it is well known that the density of modes (eigenfrequencies) increases with frequency, independent of the room shape.

$$\frac{dN_f}{df} \approx 4\pi V \frac{f^2}{c^3} \quad (4.9)$$

with V denoting the room volume.

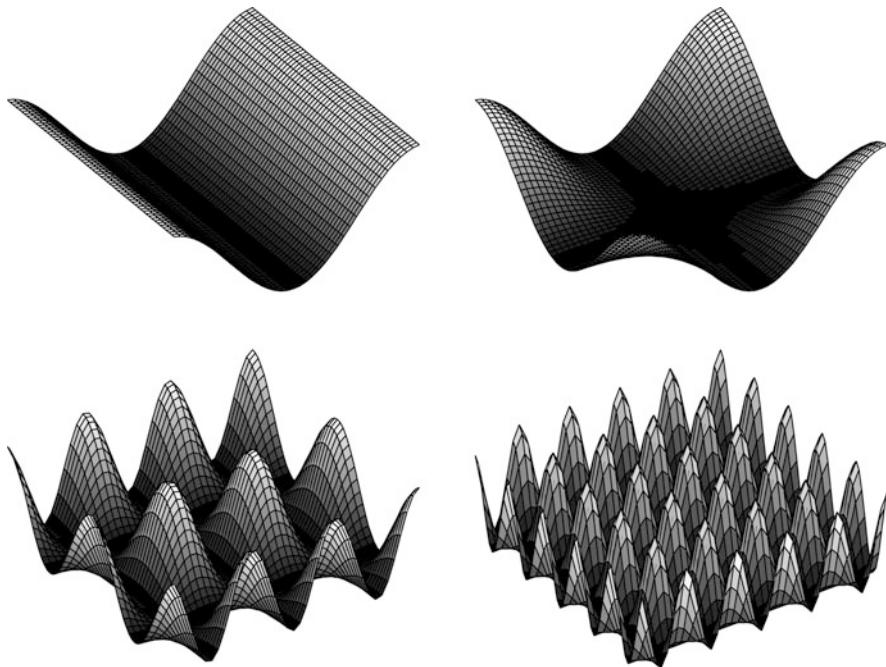


Fig. 4.3 Sound pressure distribution in 2D modal fields: $(l,m) = (1,0), (1,1), (3,2)$, and $(5,5)$

If a source is placed at an arbitrary location in the enclosure, several modes are excited in the forced response and participate in the total sound field (Fig. 4.3). The forced response of the enclosure depends on the modal overlap and the modal contributions at the certain frequency of excitation. The total field is given by superposition (Morse and Ingard 1968; Kuttruff 2016)

$$p(\vec{r}) = jkqZ_0 \sum_l \sum_m \sum_n \frac{\psi_{lmn}(\vec{r}) \psi_{lmn}(\vec{r}_s)}{K_{lmn}(k^2 - k_{lmn}^2)} \quad (4.10)$$

With \vec{r} and \vec{r}_s denoting vectors to the receiver and the (point) source positions, respectively.

4.3.1 Boundary Conditions

Boundary conditions of any kind of wall impedance do not affect the validity of the concept of modes. The modes, however, become wider in bandwidth and their overlap regions increase accordingly. The eigenfrequencies will also be slightly shifted and the local sound pressure field modified in shape and effective bandwidth.

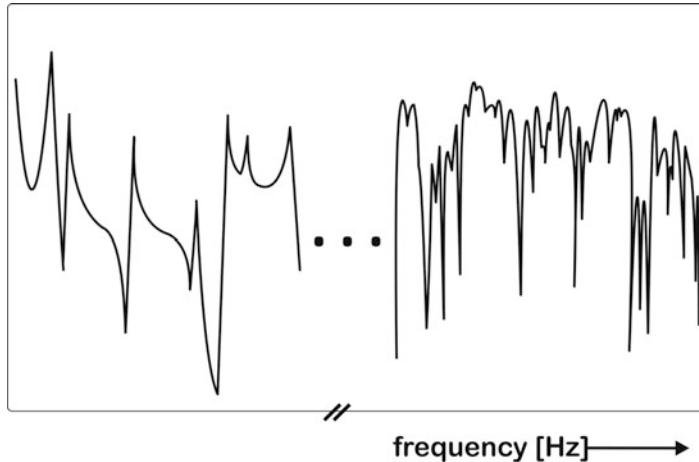


Fig. 4.4 Overlapping room modes at low and high frequencies

Normal modes and tangential modes affect the sound pressure field in a specific way. With uniform distribution of small finite wall admittances, the solution can be given by an approximation which is near to the solution without losses (see Chap. 9.4 in (Morse and Ingard 1968) for further details).

The formal solution can be also related to a complex wave number. If we use

$$\underline{k}_n = \frac{\omega_n}{c} + j \frac{\delta_n}{c} \quad (4.11)$$

Equation (4.10) changes into

$$p(\vec{r}) = j\omega qc^2 \sum_l \sum_m \sum_n \frac{\psi_{lmn}(\vec{r}) \psi_{lmn}(\vec{r}_s)}{(\omega^2 - \omega_{lmn}^2 - 2j\delta_{lmn}\omega_{lmn})}. \quad (4.12)$$

The content of Eq. (4.12) is a sequence of resonance peaks with overlap. The amount of overlap depends on the spacing of resonances and their bandwidth. If the mean resonance frequency spacing (see Eq. (4.9)) is large compared with the mean half-width, the modal overlap will be small and the characterization of the sound field in the cavity can be studied mode by mode independently (left part of Fig. 4.4).

If, however, the resonances overlap heavily, i.e. when the frequency spacing is small compared with the mean half-width, a complicated superposition occurs (right part of Fig. 4.4). According to measurements by (Kuttruff and Thiele 1954) and theoretical explanations by (Schroeder 1954; Schroeder and Kuttruff 1962), frequency curves in rooms are not suitable for solving practical problems in room acoustics. The statistical distributions of minima and maxima are just related to a general average quantity – the so-called “reverberation time”, T (see Sect. 4.5.1). The frequency range, in Hz, where this overlap occurs, starts above

$$f_s \gg \sqrt{\frac{c^3}{4V\langle\delta_{lmn}\rangle}} \approx 1200\sqrt{\frac{T}{V}} \quad (4.13)$$

with T denoting the reverberation time in s and V the room volume in m^3 . f_s is called Schroeder frequency.

Above the frequency limit¹, numerous resonances with different phases interfere in a quasi-stochastic process and produce a result of characteristic statistical features of the room transfer function (sound pressure ratio between two points in the room). The sound pressure magnitudes, for instance, follow a Rayleigh distribution.

It is crucial to recognize that the superposition result is dependent on the phase mixing of the resonances. If small details in the cavity or the room are changed (a window or a door is opened, a person is moving, air condition system changes the temperature), the speed of sound, the phases will be affected in a somewhat arbitrary way, and the result of the room transfer function may change dramatically in detail. The listening experience, however, is not as sensitive to small changes or temperature shifts. The exact fine structure of the room transfer function has, therefore, no relevance for the perception of sound in rooms. An exception is the discussion of feedback in microphone-loudspeaker systems in rooms where the exact phase and amplitude of the room transfer function is crucial to describe the feedback condition.

Our listening experience in rooms has hardly anything to do with varying sound pressure at different locations. On the contrary, room recognition is related to the subjective listening impression at one point in the room, the receiver location. And at that location, the room impression is given mainly by a time-domain effect: reverberation. The reverberation process can best be described by using the concept of geometrical room acoustics.

4.4 Geometrical Acoustics

First of all we have to discuss the concept of a “sound ray” or “sound particle”. A sound ray is thin line of sound energy transport. Accordingly, we can imagine that a particle carries the energy of the sound wave. At this point, we accept that sound rays or particles are incoherent and cannot interfere with wave effects. A ray is an energy carrier representing a spherical wave with, however, an infinitely small opening angle, $d\Omega$. From the basic distance, law follows that the intensity of the ray decreases with $1/r^2$ (r = distance from the ray origin) (Fig. 4.5).

Superposition of two rays is calculated by adding the sound pressures with amplitudes \hat{p}_1 and \hat{p}_2 . With a certain difference in path length, Δx , corresponding with a relative phase shift, the superposition of the broadband sound pressures simplifies to energy addition,

¹Which is usually very low (< 100 Hz).

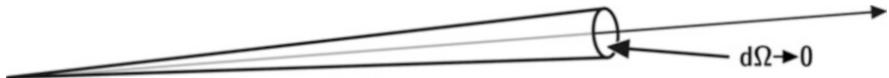


Fig. 4.5 Definition of a ray

$$|p|^2 \approx |\hat{p}_1|^2 + |\hat{p}_2|^2. \quad (4.14)$$

This can be put down to the result of integration over frequencies (broadband signals), where the latter term vanishes (see also 1.7):

$$|p|^2 = |\hat{p}_1 + \hat{p}_2 e^{jk\Delta x}|^2 = |\hat{p}_1|^2 + |\hat{p}_2|^2 + 2|\hat{p}_1||\hat{p}_2| \cos\left(\omega \frac{\Delta x}{c}\right) \quad (4.15)$$

Accordingly, we can directly deal with sound intensities as primary carriers of ray contributions:

$$I = |\vec{I}| = |\vec{I}_1| + |\vec{I}_2|. \quad (4.16)$$

The rays, too, can be related to energy density. The superposition in this case reads

$$w = w_1 + w_2, \quad (4.17)$$

with $I = cw$ in a situation where the ray represents a straight line perpendicular to a quasi-plane wave. Special models of geometrical acoustics are equivalent to those in ray optics. They are based on imaging, phase superposition and scattering, see Chap. 11.

4.5 Statistical Reverberation Theory

The basic assumption of statistical reverberation theory is the concept of a diffuse field. The concept of room sound fields is based on the assumption that sound travels in all directions with same probability and intensity. Accordingly the distribution of directions of sound incidence is uniform at any location in the room. Real sound fields have been analysed in this respect. It can be shown that the concept of a diffuse field is a good approximation in case of ordinary room dimensions and well distributed absorption on the room boundaries. In special cases such as corridors or very large factory halls with concentrated absorption, diffuse sound fields will not be present.

The diffuse sound incidence also holds for the interaction of sound with the boundaries. According to fundamental calculations (Kuttruff 2000), the irradiation strength of the diffuse sound incident on a room boundary is

$$B = \frac{c}{4} w. \quad (4.18)$$

The effective absorption coefficient, α_{diffuse} , is then

$$\alpha_{\text{diffuse}} = \frac{I_{\text{abs}}}{B} = \int_0^{\pi/2} \alpha(\theta) \sin 2\theta \, d\theta \quad (4.19)$$

with $\alpha(\theta)$ denoting the angular-dependent absorption coefficient. In case of locally reacting absorbers $Z(\theta) = \text{const.} = Z$. For real Z holds

$$Z = \frac{1 + \sqrt{1 - \alpha_0}}{1 - \sqrt{1 - \alpha_0}} \quad (4.20)$$

with α_0 = absorption coefficient for normal incidence, and according to Eqs.(3.3) and (3.4) (Cops and Mynke 1973)

$$\alpha_{\text{diffuse}} = 8 \left[\frac{1 - \sqrt{1 - \alpha_0}}{1 + \sqrt{1 - \alpha_0}} \right]^2 \left[\frac{2}{1 - \sqrt{1 - \alpha_0}} - \frac{1 - \sqrt{1 - \alpha_0}}{2} + 2 \ln \frac{1 - \sqrt{1 - \alpha_0}}{2} \right] \quad (4.21)$$

Finally, the diffuse sound field will be created by numerous reflections of sound waves or rays hitting walls. The mean free path describing this process is dependent on the ratio of the volume to the surface area. It is also equivalent to equations describing the radiation of heat energy from a heated body: The larger the boundary is, compared with the volume, the greater the energy exchange between the boundary and the exterior space. Thus,

$$\bar{n} = \frac{cS}{4V} \quad (4.22)$$

gives the mean reflection rate (between two successive reflections), or

$$\bar{l} = \frac{4V}{S} \quad (4.23)$$

the mean free path.

4.5.1 Reverberation

When sound power is fed continuously into a room and the source is suddenly switched off, the sound will not stop rapidly but decay. This observation is of importance when it comes to assessing the adequacy of the room for performance of speech or music. It is called reverberation. Reverberation is caused by sound reflections interacting with the room boundaries and the energy transfer to the boundaries according to energy absorption. To obtain a derivation of the reverberation equation, we study a room with homogenous distribution of absorption, α , and a diffuse sound field. The total sound field consists of sound rays of energy e_0 emitted from a source. Once in a while, they hit walls and suffer an average energy loss² of α . The remaining energy is $e_0(1 - \alpha)$. After n reflections (after a time $t = n/\bar{n}$), the remaining energy of one ray is

$$e(t) \approx e_0(1 - \alpha)^{\bar{n}t} = e_0 e^{\bar{n}t \ln(1 - \alpha)} \quad (t \geq 0). \quad (4.24)$$

The same principle can be observed for the whole population of rays and the energy density in the room.

$$w(t) \approx w_0(1 - \alpha)^{\bar{n}t} = w_0 e^{\bar{n}t \ln(1 - \alpha)} \quad (t \geq 0). \quad (4.25)$$

According to the definition of the sound pressure level, the decay of sound energy density can be expressed in terms of a linear level decay (Fig. 4.6).

$$L(t) = L_0 + 4.34 \bar{n}t \ln(1 - \alpha). \quad (4.26)$$

The reverberation is usually described by a specific decay time, namely, the time required for a level decrease of 60 dB. This time is referred to as reverberation time, T .

$$T = -\frac{60}{4.34 \bar{n} \ln(1 - \alpha)} \quad (4.27)$$

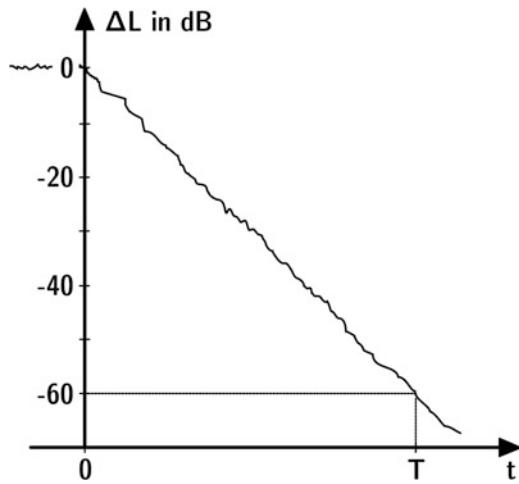
or written as Eyring's equation ($c = 340$ m/s)

$$T = -\frac{24 \ln 10}{c} \frac{V}{S \ln(1 - \alpha)} = 0.16 \frac{s}{m} \frac{V}{-S \ln(1 - \alpha)} \quad (4.28)$$

which can be approximated for $\alpha \ll 1$ (the famous Sabine's equation)

²the average α for the diffuse field

Fig. 4.6 Decay curve.
Level drop $\Delta L = L_0 - L(t)$
after switch-off at $t = 0$



$$T = 0.16 \frac{V}{S\alpha} = 0.16 \frac{V}{A} \quad (4.29)$$

(V in m^3 , S , A in m^2 , T in s). If the absorption is distributed locally, the mean effective absorption coefficient can be estimated by accounting for the surface areas of the materials of distinct absorption.

$$\bar{\alpha} = \frac{A}{S} = \frac{1}{S} \sum_{i=1}^N S_i \alpha_i \quad S = \sum_{i=1}^N S_i \quad A = \sum_{i=1}^N S_i \alpha_i \quad (4.30)$$

The quantity A is denoted equivalent absorption area.

Now, we have to keep in mind that a sound ray may travel in a room from reflection to reflection over a rather long total distance. If we assume a travel time of 2 s until its energy is negligible (at reverberation time with level loss of 60 dB, for instance), the total path length is 680 m. Over this long distance of sound propagation in air, the air attenuation will also contribute to energy losses.

According to Eq. (3.33), the intensity or energy density loss is given by $e^{-mc t}$. The energy loss of sound rays, therefore, can be expanded to

$$w(t) = w_0 e^{[\bar{n}t \ln(1-\alpha) - mc]t} \quad (4.31)$$

which modifies the result of Eyring's and Sabine's reverberation formulae as follows

$$T = 0.16 \frac{V}{-S \ln(1-\alpha) + 4mV} \quad (4.32)$$

$$T = 0.16 \frac{V}{A} \quad (4.33)$$

with

$$A = \sum_{i=1}^N S_i \alpha_i + 4mV. \quad (4.34)$$

4.5.2 Steady-State Energy Density and Level

Imagine a sound source radiating a sound power P in a steady state. Sound rays are emitted which travel between the room boundaries and contribute to the total energy density. The first-order reflected sound and each reflected component will permanently be present. The total field is the sum of all reflected components. At any time, the temporal series of reflections as described by Eq. (4.31) radiated from the source add up with corresponding (reverse) delay. Components of late arrival were radiated earlier, just accounting for the travel time t' for each contribution, so that all coincide at any arrival time t . The steady-state energy in the room is, thus, calculated by integrating Eq. (4.31), starting from the shortest reflection travel time, t_0 . With correct sound power normalization where $e_0 = Pdt$, it reads (in this case $P(t) = \text{const}$)

$$w(t) = \frac{1}{V} \int_{t_0}^{\infty} P(t-t') e^{[\bar{n}t \ln(1-\alpha)-mc]t} dt'. \quad (4.35)$$

The solution of this integral is

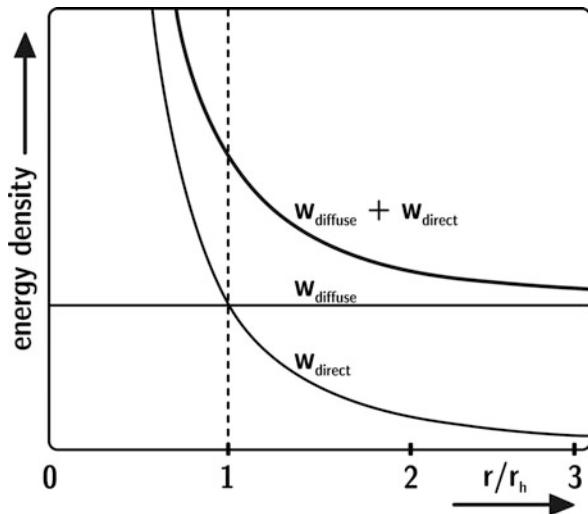
$$w_{\text{diffuse}} = \frac{4P}{cA} e^{-A/S} \quad (4.36)$$

if we assume an average first-order reflection delay $t_0 = 1/\bar{n}$. The equivalent absorption area, A , in this equation represents the same quantity as used in the corresponding reverberation formula. Either it can be interpreted as

$$A_{\text{Sab}} = \sum_{i=1}^N S_i \alpha_i + 4mV. \quad (4.37)$$

(Sabine approximation for $\alpha \ll 1$), or more generally

Fig. 4.7 Sound energy density in rooms with diffuse field



$$A_{Eyr} = -S \ln(1 - \bar{\alpha}) + 4mV, \bar{\alpha} = \frac{1}{S} \sum_{i=1}^N S_i \alpha_i \quad (4.38)$$

In level notation and with reference to the sound power level of the source, the diffuse field sound pressure level reads

$$\begin{aligned} L_{\text{diffuse}} &= L_w - 10 \log A + 6 - 4.34 \frac{A}{S} \\ L_{\text{diffuse}} &= L_w + 10 \log T + 14 - 4.34 \frac{A}{S}. \end{aligned} \quad (4.39)$$

The energy density accounts for the room reflections. We must take the primary sound, called “direct sound”, into account to obtain the total energy density in steady state (Fig. 4.7).

The sound source will generally radiate a spherical wave, either omnidirectional or with specific directivity, $\Gamma(\theta, \varphi)$. According to Eq. (2.28) the direct sound contributes with

$$w_{\text{direct}} = \frac{P}{4\pi c r^2} |\Gamma|_{\max=1}^2. \quad (4.40)$$

The total sound energy density is composed of the direct and the reverberant field. The distance of equal contribution of each field is the reverberation distance, r_{rev} .

$$r_{\text{rev}} \approx \frac{1}{4} \sqrt{\frac{A}{\pi}} \Gamma(\vartheta, \varphi) \approx 0.057 \sqrt{\frac{V}{T}} \Gamma(\vartheta, \varphi) \quad (4.41)$$

and the ratio of the reverberant (diffuse) field to the direct field energy is

$$\frac{w_{\text{diffuse}}}{w_{\text{direct}}} = \left(\frac{r}{r_{\text{rev}}} \right)^2. \quad (4.42)$$

Non-Diffuse Room Sound Fields

When it comes to non-diffuse room sound fields which are found in flat or long rooms or in coupled rooms, the sound pressure level can be calculated based on the sound power. The room surfaces, their absorption and scattering and, furthermore, the obstacle scattering of room fittings have crucial influence on the sound decay and level. Both quantities depend on the distance between source and receiver (Figs. 4.8 and 4.9).

It is important to know that in these rooms neither the decay curve is straight nor the room level is constant. For more detailed information, see (Kuttruff 2016).

The table below summarizes equations available for predictions of so-called sound propagation curves in rooms for some typical cases; see also (VDI 3760). Particularly for fitted rooms see (Hodgson 1983; Ondet and Barby 1988, 1989). Sound propagation curves are plots of the sound pressure level, L , minus the sound power level, L_w , as a function of the distance between the source and the receiver. The equations listed in the table are useful for evaluation of simulation models such as ray tracing (see Chap. 11). Furthermore, with these equations enable us to model the sound field in a room by using a simple distance-dependent direct sound (with or without directional characteristics of the source) and an artificial reverberation process to simulate a room response. The level balance between those two components is based on Eq. (4.39) and on the direct sound level as listed in Table 3.1. Later it will be shown that such an approach will lead to artificial sound effects sufficient for computer games, for instance. The details of specific room acoustics as related to the specific physical phenomena and to aspects of perception, however, are not taken into account sufficiently, and better simulation models are required (Table 4.1).

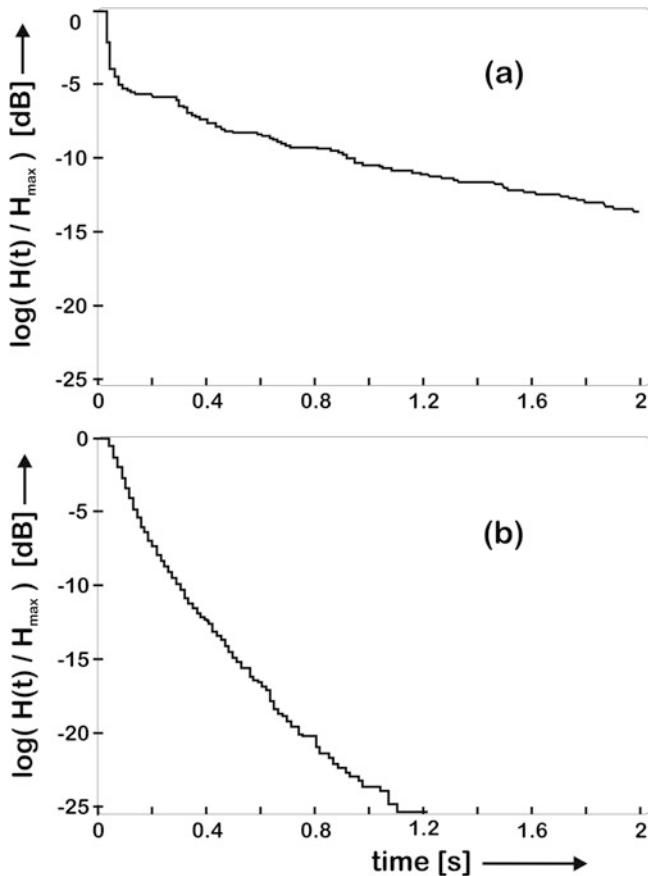


Fig. 4.8 Decay curves in a flat room ($100 \times 100 \times 4 \text{ m}^3$, $\alpha = 0.1$) with specularly (a) and diffusely reflecting (b) floor and ceiling (after Vorländer 1988)

Fig. 4.9 Sound level in a flat room with height a and diffusely reflecting floor and ceiling. (see Kuttruff 1985; Vorländer 1988). Analytic results in the solid curve compared with ray tracing results are shown. Level plot relative to direct sound in distance $a/2$. The side walls are located at $r/a = 12.5$

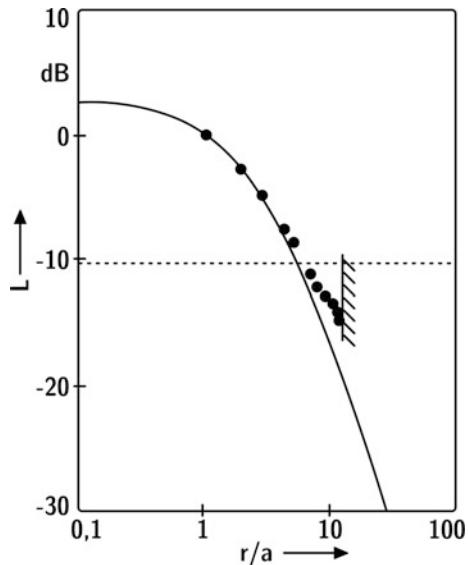


Table 4.1 Equations for calculation of sound propagation curves in non-diffuse spaces (after (Kuttruff 1985)), in all equations and figures: $\rho = 1-\alpha$

Room type	
Flat room (height $a \ll$ width, length), with specularly reflecting floor (index f) and ceiling (index c), empty. Dashed line: Free field	$w = \frac{P}{4\pi c} \sum_{k=1}^{\infty} \left(\frac{1/\rho_c + 1/\rho_f}{r_{2k-1}^2} + \frac{2}{r_{2k}^2} \right) (\rho_c \rho_f)^k$ <p>With $r_k = \sqrt{k^2 a^2 + r^2}$ and r = source-receiver distance.</p>
As above, $r \gg a$	$w \approx \frac{P}{4\pi c r^2} \frac{\rho_c + \rho_f + 2\rho_c \rho_f}{1 - \rho_c \rho_f}$

Chapter 5

Structure-Borne Sound



5.1 Introduction

The field of structure-borne sound is more complex than fluid acoustics due to the manifold possible degrees of freedom for force and torque and corresponding elastic reactions. In contrast to fluids, solids can react with elastic forces to pure deformation. The number of possible waveforms is therefore higher. And these waveforms can interact and exchange energy, depending on the geometric and elastic coupling in the specific shape of the solid body. Structure-borne sound plays a very important role in building acoustics, in noise control and in vehicle acoustics, for instance. The characterization of structure-borne sources is a very complex field, too. The dynamic behaviour of the source coupled to the structure must be known. Even in the geometrically simple case of a point-like force source, the elementary parameters of the source can be complex since they are related to real and ideal force sources and inner impedances, similar to electric models of voltage sources acting on electric networks. The power output then not only depends on the source itself but also on the kind of attached structure and the impedance coupling between source and structure.

5.2 Waves in Solid Media

In general, structure-borne sound in unbounded elastic media can be separated into normal (compression) strain effects and tangential (shear) strain effects (Cremer et al. 1973). The elastic constants for these strains are denoted σ and τ , the indices indicating the direction of forces and displacements, for instance, σ_{xx} and τ_{xy} , respectively. The following equations can be interpreted as special cases of Hooke's law in solid structures. We introduce the displacement vector, \vec{s} , and its components

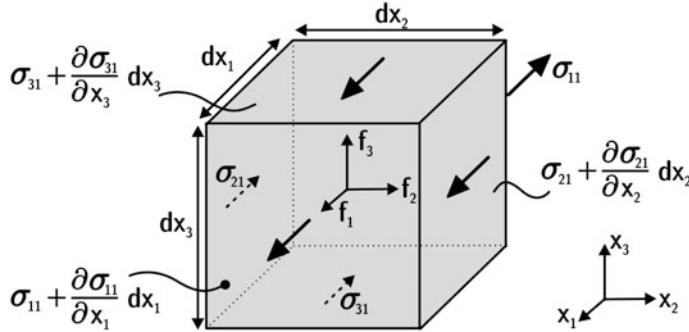


Fig. 5.1 Notation of coordinates in a solid medium

ξ, η, ζ , all being functions of the position, x_1, x_2, x_3 , in Cartesian coordinates (more details for example in Möser 2004; Kuttruff 2007).

$$\vec{s}(x_1, x_2, x_3) = \begin{pmatrix} \xi(x_1, x_2, x_3) \\ \eta(x_1, x_2, x_3) \\ \zeta(x_1, x_2, x_3) \end{pmatrix} \quad (5.1)$$

$$\sigma_{x_1 x_1} = 2\mu \frac{\partial \xi}{\partial x_1} + \lambda \operatorname{div} \vec{s}, \tau_{x_1 x_2} = \mu \left(\frac{\partial \xi}{\partial x_2} + \frac{\partial \eta}{\partial x_1} \right) \quad (5.2)$$

and so on; see Fig. 5.1. The constants μ and λ are called Lamé elastic constants:

$$2\mu = \frac{E}{1 + \gamma}, \lambda = \frac{\gamma E}{(1 + \gamma)(1 - 2\gamma)} \quad (5.3)$$

E denotes the Young's modulus of elasticity and $\gamma =$ Poisson ratio (lateral contraction). μ is also called shear modulus. By setting the elastic forces equal to the inertia, we obtain a set of wave equations which can be discussed for cases of constant conditions within planes of (y, z). This consideration yields plane-wave equations, the first of which characterizes longitudinal waves with displacement in x direction. The local displacements are related to pure compression, just as in fluid media, and the sound speed is

$$c_L = \sqrt{\frac{2\mu + \lambda}{\rho}}. \quad (5.4)$$

Another equation characterizes transversal waves (shear waves) with the sound speed.

$$c_T = \sqrt{\frac{\mu}{\rho}}. \quad (5.5)$$

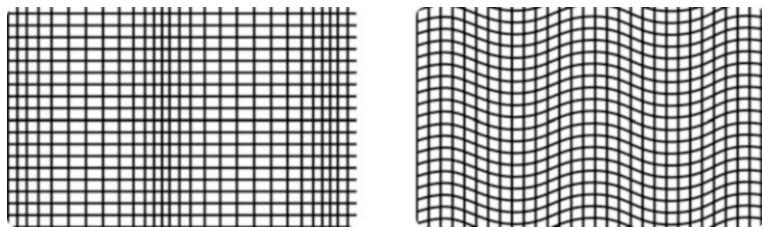


Fig. 5.2 Longitudinal (left) and transversal (right) waves in solid media

Table 5.1 Material data for structure-borne sound

Material	ρ (10^3 kg/m 3)	γ	E (10 9 N/m 2)	c_L (m/s)	c_T (m/s)
Aluminium	2.7	0.355	70	6420	3040
Stainless steel	7.9	0.30	200	5790	3100
Copper	8.9	0.37	125	5010	2270
Brass	8.6	0.374	105	4700	2110
Lead	11.3	0.43	16	1960	690
Glazing	3.9	0.224	54	3980	2380
Plexiglass	1.2	0.40	4	2680	1100

As the medium reactive force and the elasticity are lower for shear forces than for compression forces, the speed of transversal waves is lower than the speed of longitudinal waves (Table 5.1). Figure 5.2 illustrates the direction of the displacement in these examples for longitudinal and transversal waves.

5.3 Waves on Plates and Their Radiation

The wave forms in plates and beams are quite similar to those in unbounded media. The elastic conditions at the boundary, however, yield deviations from the purely longitudinal or transversal movement. The boundary conditions produce slight components of transversal movement on longitudinal waves and vice versa. Nevertheless we might identify the elementary wave types as quasi-longitudinal and quasi-transversal (Fig. 5.3).

The quasi-transversal waves, called bending waves, are particularly important since they interact with airborne sound in a very effective way. Problems of coupled airborne and structure-borne sound modelling therefore involve bending wave theory directly (Fig. 5.4).

Sound is radiated at an angle of ϑ from a large thin ($h \ll \lambda$) plate with bending waves

$$\sin \vartheta = \frac{\lambda}{\lambda_B} = \frac{c}{c_B}. \quad (5.6)$$

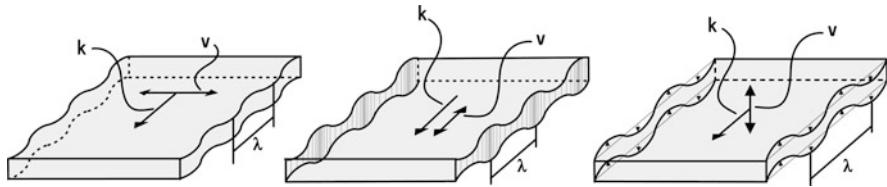
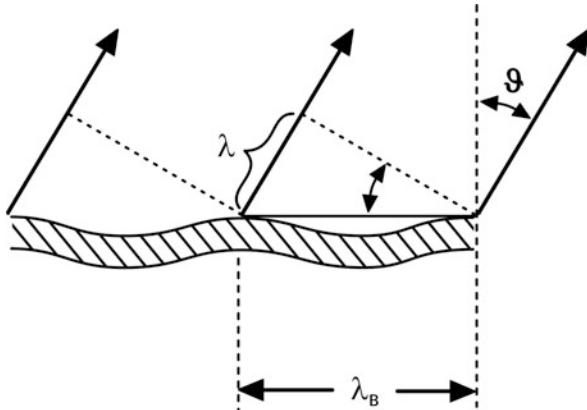


Fig. 5.3 Waves on plates

Fig. 5.4 Bending wave interaction with airborne sound



This equation can only hold if the airborne sound speed, c , is smaller than c_B . The point is that the bending wave speed is dependent on frequency. Approximations of the transversal wave equation for thin plates yield the dispersion relation

$$\omega = k_B^2 \sqrt[4]{\frac{B}{\rho}}. \quad (5.7)$$

with B denoting the bending stiffness of the plate,

$$B = \frac{h^2}{12} \frac{E}{1 - \gamma^2}, \quad (5.8)$$

ρ the density and h the thickness. From the dispersion relation, we can derive two wave speeds: the group velocity

$$c'_B = \frac{d\omega}{dk_B} = 2\sqrt{\omega} \sqrt[4]{\frac{B}{\rho}} \quad (5.9)$$

which denotes the speed of the envelope of a gated pulse (see Fig. 5.5). The phase velocity is relevant for phase coupling with the airborne sound wave. It is the speed of a zero of the carrier signal.

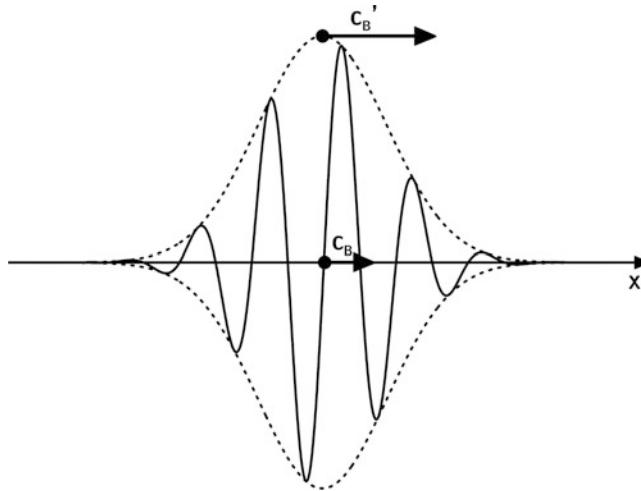


Fig. 5.5 Group velocity and phase velocity in bending waves, $c'_B = 2 c_B$

$$c_B = \frac{\omega}{k_B} = \sqrt{\omega} \sqrt[4]{\frac{\rho}{B}} \quad (5.10)$$

The frequency regime where the phase velocity, c_B , is larger than the sound speed in free air, c , is given for

$$f > f_c = \frac{c^2}{2\pi} \sqrt{\frac{\rho}{B'}} \quad (5.11)$$

which also is a requirement for $\sin\theta \leq 1$ in Eq. (5.6).

Due to the coincidence effect, sound radiation is, thus, very effective above the critical frequency, f_c . It can also be written as

$$f_c = \frac{c^2}{2\pi} \sqrt{\frac{m''}{B'}} \quad (5.12)$$

with m'' denoting the mass per unit area in kg/m^2 and B' in Nm the bending stiffness normalized to the thickness h .

The radiated sound power can be calculated from the intensity. With a given plane airborne sound wave coupled with the bending wave (Fig. 5.4), we can identify that the plate velocity component $j\omega\zeta$ is identical with the velocity of the airborne wave, $v \cos\theta$. This yields

$$\left| \vec{I} \right| = \frac{1}{2} \rho_0 c |v|^2 = 2\pi^2 \rho_0 c |\zeta|^2 \frac{f^3}{f - f_c}. \quad (5.13)$$

Below f_c the radiated sound field has a very small intensity. The movement of the plate leads to near-field effects but to only negligible airborne sound due to acoustic short circuits between neighbouring plate areas. These findings can be transferred to the case of plates with finite size. At the plate boundaries, the local air movement in the short circuits will not find antiphase neighbouring zones. Hence the boundary zones of the plate will act as a line source, the total sound radiation of which depends on the shape of the plate and the local boundary condition of the plate support.

An important quantity to describe the total plate as an airborne sound source is the radiation efficiency. It is defined to be the ratio, σ , of the radiated sound power, P , to the maximal possible sound power of a matched source of area S and effective plate velocity, \tilde{v}^2 ,

$$\sigma = \frac{P}{\rho_0 c S \tilde{v}^2}. \quad (5.14)$$

5.3.1 Finite-Size Plates

First of all we have to take a look at the differential equations and wave types for unbounded cases to study elastic wave on finite-size plates (Fig. 5.3). The boundary condition is then introduced for the plate support at its perimeter, for example, “simply supported”, “clamped” or “free” (Cremer et al. 1973).

The example of a rectangular plate supported at the edges leads to eigenfunctions related to the bending wave equation. The velocity and, thus, the input impedance seen from the source at the force injection point depend strongly on the location. In case of a simply supported plate, the velocity and the acceleration at the edges must vanish. The eigenfunctions accordingly are similar to the case of airborne sound with hard boundary conditions ($v = 0$). And also similar to room modes, at higher frequencies the plate modes overlap and form a statistical energy system.

This modal system can be excited by airborne sound by applying a forced response and a resonant response. In forced response, the plate follows the airborne excitation in wavelength, phase and angle. In resonant response, the modal pattern of bending waves comes into play, which happens, of course, above critical frequency.

Another aspect of the radiation can be expressed in terms of the radiation efficiency and its dependence on the plate aperture. The latter can also be related to the radiation impedance of the plate (Wallace 1972).

5.3.2 Internal Losses and Structural Reverberation Time

Damping of waves on plates can be caused by friction and visco-elastic effects in the material. Another factor of energy loss in the plate wave is given by sound radiation into the fluid surrounding the plate. This effect can accordingly be accounted for by the radiation efficiency. A third component is energy flow through the plate boundaries, at least in non-ideal mounting conditions or junctions, like fixed, free or simply supported. In practical cases, plates are part of a larger system or construction, connected by junctions. The outgoing flow of vibration energy, therefore, must be interpreted as a kind of loss.

All types of losses affect the modal behaviour and the width of the resonance peaks. The absorption of a modal field of plate waves can also be expressed in the time domain by the structural reverberation time, T_s .

$$T_s = \frac{2.2}{f \eta} \quad (5.15)$$

with f denoting the frequency and η the total loss factor. The total loss factor, η_{tot} , consists of the three components listed above, internal, radiation and coupling losses. In case of a four-sided plate, it reads

$$\eta_{\text{tot}} = \eta_{\text{int}} + \frac{\rho_0 c \sigma}{\pi f m''} + \frac{c}{\pi^2 S \sqrt{f f_c}} \sum_{k=1}^4 l_k \alpha_k. \quad (5.16)$$

In this equation the internal losses, η_{int} , the radiation efficiency, σ , the surface area of the plate, S , the frequency f and the critical frequency f_c , and l_k and α_k the length and the plate wave absorption coefficient of the edge k , respectively, are used.

5.4 Vibration Transmission Over Junctions

Therefore the velocity amplitude of bending waves of plates interacting with air is a crucial quantity. When it comes to complex structures, these plates or plate elements are integrated into a system of adjoining plates, beams, cavities, etc., for instance, in a vehicle, ship or building. Plates meet at line junctions. Classical junction types are the cross-junction, \perp ; the L-junction, L ; and the T-junction, T . The connection between the joints, whether rigid or elastic, plays an important role in the energy transmission over the junction. Furthermore, the junction is a location of discontinuity for the plate wave which will lead to a change in wave impedance and, thus, to a reflection and transmission. Energy will be transferred and exchanged according to

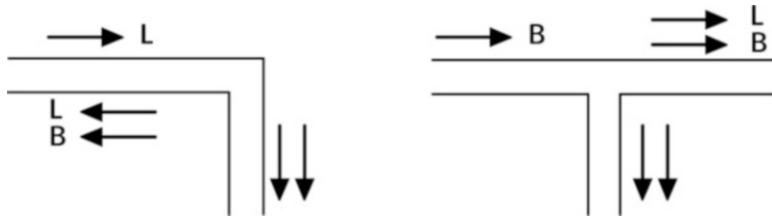


Fig. 5.6 Examples of wave type conversion of longitudinal (L) and bending (B) waves at junctions

the degrees of freedom of each displacement wave type. A pure bending wave travelling into a junction will result into a mixed bending/longitudinal wave in the reflection and in the transmitted domain, too (Fig. 5.6).

An important parameter describing the energy transmission of bending waves over a junction is the vibration reduction index, K_{ij} . It is defined as the structure-borne sound energy passing the junction:

$$K_{ij} = \frac{D_{v,ij} + D_{v,ji}}{2} + 10 \log \frac{l_{ij}}{\sqrt{\alpha_i \alpha_j}} \quad (5.17)$$

$D_{v,ij}$ denotes the vibration level difference between the plate on the incident side, i, and the plate on the transmission side, j, over the coupling length l_{ij} . α_i and α_j are the equivalent absorption lengths of the plate i and j, respectively, representing the internal, junction and radiation losses of the plates.

The relatively simple equations listed above may be applied to the case of homogenous thin plate structures. In practice, however, multi-layered plates and junctions with point or line contacts, frames with lightweight plates and silicone-filled gaps must often be simulated. For these cases, an analytical solution cannot be found, nor can it be obtained easily by numerical simulation. Instead, models that include thickness effects and compression-to-shear wave conversion must be used to describe the junction transmission loss of the specific case.

Chapter 6

Psychoacoustics



6.1 Introduction

A proper physical foundation of acoustics is essential to understand the generation and transmission of sound. As far as simulation and auralization are concerned, however, human beings as the final recipient of sound are the final authorities on sound quality, comfort or annoyance. Basic knowledge about the human hearing system, its anatomy and neurophysiology, thus, provides us with a deeper understanding of how auralization must be organized and programmed, how exact it must be, which acoustic aspects are essential and which can be neglected. The term psychoacoustics involves the description of human hearing and its modelling. Psychoacoustic model functions try to extract from physical data like sound pressure time functions or spectra the characteristic data related to specific hearing dimensions, like loudness, sharpness, impulsiveness or spaciousness. These dimensions are straightforward physically based descriptors, which as a set form the “character” of sound. Finally, with inclusion of neural top-down processes, based on our memory, previous experiences and expectations, the “quality” of sound can be judged, and its contribution to comfort or annoyance can be evaluated. To do this, however, the context of the sound event and a reference for judgement must be taken into account (Blauert 2005).

6.2 Anatomy of the Peripheral Hearing System

Knowledge of the peripheral hearing organ and corresponding signal processing leads to models of perception of frequency range and resolution, temporal resolution, loudness, modulation, speech intelligibility and simultaneous and temporal masking. Models of central processes, particularly involving both input channels – left and

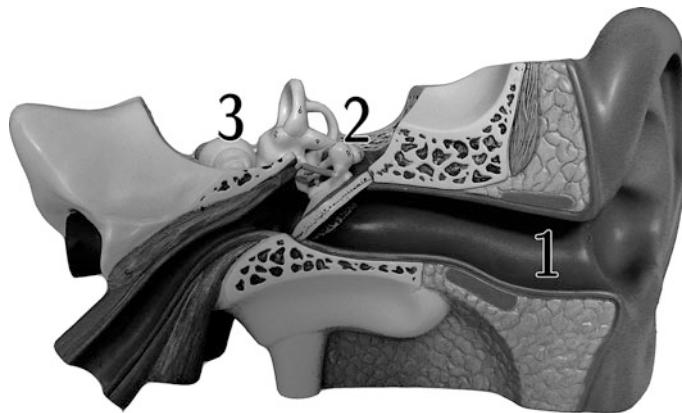


Fig. 6.1 Human hearing organ

right ear – lead to special, so-called “binaural” and “interaural” cues like source localization, precedence effect and spaciousness (see below).

Binaural hearing (Sect. 6.4) is a specific discipline in psychoacoustics. It is essential for a realistic or at least plausible auralization result. The consequences of binaural hearing are a directional localization and distance perception. Furthermore, binaural hearing makes a very effective noise suppression possible which enables us to understand speech in noise environments, even in rooms and under conditions of concurrent speech signals (“cocktail party effect”).

If sound is to be measured, simulated or auralized binaurally, it must be represented for both ears separately and accurately. This requirement follows from a system-theoretical description of the outer ear and the transmission of sound from the undisturbed sound field¹, as such, into the ear canal and to the eardrum, to the middle and inner ear and so forth.

The peripheral hearing system consists of outer (1), middle (2) and inner (3) ear (Fig. 6.1). The outer ear includes the pinna and the ear canal. The latter is a tube of about 2.7 cm length and 6 to 8 mm diameter. At its end, it is terminated by the eardrum, a piece of skin of 0.8 cm^2 area and 0.1 mm thickness. The ear canal is roughly a $\lambda/4$ -resonator with a resonance frequency of about 3 kHz. In this frequency range, the hearing sensitivity has its maximum. Except for spatial hearing at higher frequencies, the pinna is of no influence.

The middle ear (2) consists of the middle ear bones malleus, incus and stapes. They are coupled to mechanical transmitter levers. Airborne sound from the ear canal must be transmitted into the fluid-filled inner ear. The impedance match provided by the middle ear improves the efficiency of sound transmission significantly. The main effect of the middle ear, however, is not a force transducer lever,

¹With reference to the sound field with the head absent.

but a pressure/force transducer due to the area ratio of the eardrum and the oval window. The oval window is the connection between the stapes and the inner ear.

The inner ear (3) consists of the cochlea and the semicircular canals which contain the human organ of equilibrium. The cochlea is shaped like an upward-growing spiral with 2.5 twists and a length of 3 cm. It is placed into the petrosal, an extremely hard bone. The cochlea is divided into two parts by several membranes. The basilar membrane plays a central role in the transduction of vibro-acoustic stimuli into neural spike rate signals. It varies significantly in terms of width and stiffness along its length. The stiffness in the part near the oval window is greater than towards the end.

The fluid in the cochlea is excited by the stapes acting on the oval window. The sound wave in the fluid excites the basilar membrane to transversal waves. The better the impedance match between the travelling wave and the local mass and stiffness of the basilar membrane, the higher the amplitude of these transversal waves. Therefore, the vibration amplitude on the basilar membrane varies with frequency along its length. The phenomenon is called frequency-space-transformation.

The local displacements on the basilar membrane (3) are detected by sensory cells in the organ of Corti. They are grouped into outer and inner hair cells. The latter are connected to the auditory nerve which leads to the auditory cortex in the brain. Each nerve cell responds to a certain frequency, according to its location of connection to the inner hair cells. An important feature of the Corti organ is feedback. The outer hair cells have little influence on direct sound transduction into the auditory cortex. Instead, they play an active role in increasing the effective vibration amplitude of the inner hair cells. By this process it is possible to increase sensitivity by several tens of decibels and, at the same time, the frequency selectivity.

The inner ear can also be excited by direct structural vibration (bone conduction). The hearing threshold for airborne-excited bone conduction of the head, however, is 40 to 50 dB above the normal air conduction threshold. The situation changes, of course, when the head bones are excited by a force directly, with bone conduction transducers (or by a dentist drill).

6.3 Psychoacoustic Characterization

Detailed knowledge is available about the function and physiology of the peripheral hearing system, including the outer, middle and inner ear. Stimuli with certain characteristics will lead to specific impressions. Hearing models that include the hair cell stimulation and feedback describe several aspects of peripheral sound processing and of sound character. They can be used to develop a set of parameters beyond the purely physical sound level in decibels.

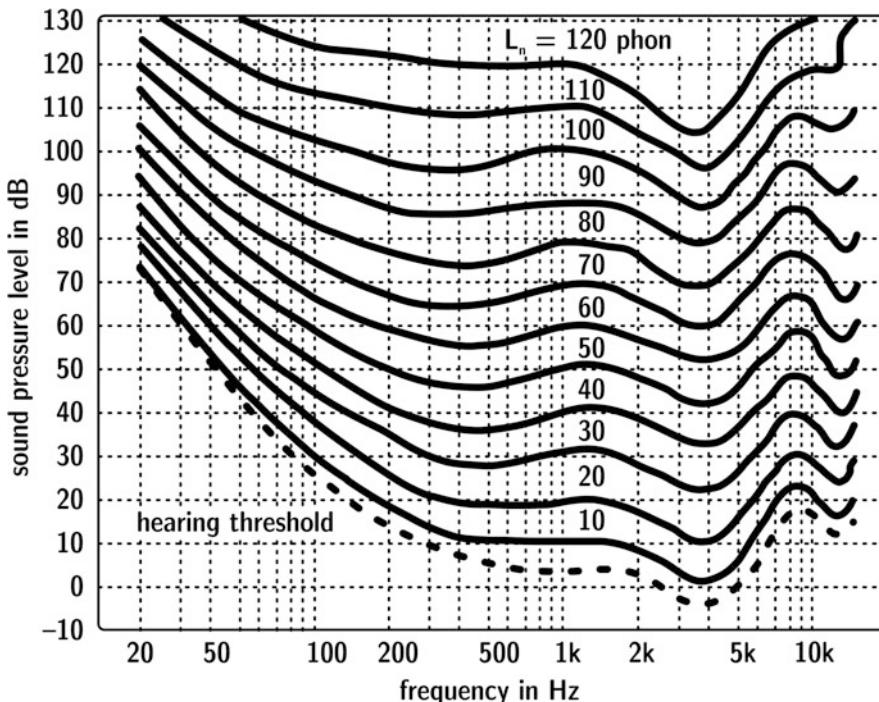


Fig. 6.2 Equal loudness level contours

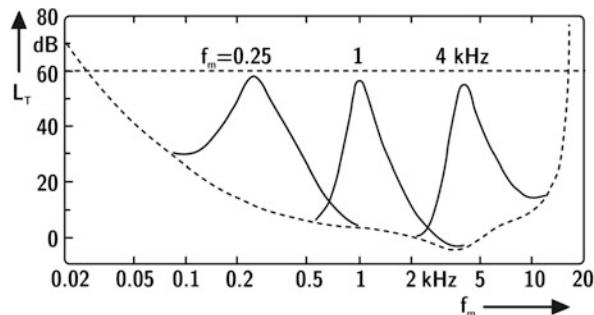
6.3.1 Loudness

A first-level approximation for many human sensory perceptions is that they are proportional to the relative change in a measurable physical stimulus (law of Weber and Fechner). The loudness impression should then be represented by a logarithmic scale of sound pressure, thus, in decibel scale. However, the decibel scale is not exactly a subjectively linear scale since doubling the subjective loudness impression requires an increase of 10 dB (in narrow band).

Equal loudness in wide frequency ranges is linked to a frequency-dependent physical stimulus of sound pressure level. The audible range can hence be described by a set of curves representing equal loudness level contours for pure tones. An old unit used in this respect is the “phon” (Fig. 6.2). For a pure tone of 1 kHz, the decibel value is equivalent to phon value.

A better estimation of the subjective impression of loudness is given by the scale in units of sone. 1 sone corresponds to the loudness of a 1 kHz pure tone at 40 dB sound pressure level. Based on these contours, filters were developed which approximately account for the “frequency response” of the ear. A smooth filter response at low levels was implemented as the so-called A-weighting, and the level measured with A-filter is expressed in dB(A). B, C and D filters were defined as well to express

Fig. 6.3 Masking curves (shifted hearing threshold).
Maskers are narrowband noise at midband frequencies of 250 Hz, 1 kHz and 4 kHz, each with 60 dB sound pressure level. The test tone is a pure tone (after Zwicker and Fastl 1999)



the ear's frequency sensitivity at higher levels. Due to the nonlinear behaviour of the hearing system (which is the reason for the different weighting curves at different levels), one closed linear technical system is not sufficient to model the subjective impressions. Another consequence of nonlinear cochlea mechanics is the so-called masking.

Sound in one frequency band may mask sound in adjacent bands. This phenomenon is called "simultaneous masking". Determining the A-weighted sound pressure is not sufficient to describe the perceived loudness. Better results are obtained by dividing the sound spectrum into critical bands before calculating the specific loudness in each band and to account for the masking curves between the bands (Zwicker and Fastl 1999).

It is more common for lower frequencies to mask higher frequencies although higher frequencies can also mask higher frequencies. Thus the masking curves shown in Fig. 6.3 are not symmetric. These masking curves represent the minimum sound pressure level which is required to perceive a tone, while another tone or masking sound is present. It is a threshold as well but connected to simultaneous masking sound (tone, noise, etc.).

For approximations of the exact masking curves, standardized (ISO 532B) curves are available to evaluate the specific (in narrow bands) and the total loudness. Modern instrumentation is capable of calculating the total loudness by planimetry.

6.3.2 *Temporal Masking*

When it comes to rapidly changing transient sound like a series of impulses, for instance, masking occurs as well. After exciting the ear with an impulse, a certain time is required to reload the charge difference for the electrostatic potential in the hair cells. This effect is called "post-masking". It can be easily understood by this cochlear effect. Pre-masking also exists. In this case, a previous impulsive sound may be masked by a loud sound arriving later. This effect, of course, contradicts cause and effect and can only be explained by a higher-level neural process in sound

evaluation in the brain. During the sound event processing, the priority obviously is shifted towards a more “important” sound event, in spite of the ongoing processing of a previous, less intensive sound.

6.3.3 Time-Varying Loudness

The current standards and loudness models are based on stationary sounds. Fluctuating sound is in fact more common than steady state sound. The hearing organ discriminates loudness in time intervals of 10 ms. In psychoacoustic applications, this effect is relevant for modulated sounds or impulsive sounds including music and speech. Standard models were developed in the revision of ISO 532 based on the German standard DIN 45631/A1:2010. In part 2 of ISO 532, the American ANSI S3.4-2007 standard for calculating stationary loudness according to (Moore and Glasberg 1996) is included. The procedure of loudness calculation strictly includes standardized temporal and spectral filter effects on the waveform of the time signal. The results are specific and total loudness vs. time functions.

6.3.4 Sharpness

A significant subjective impression and independent dimension of sound character is the sensation of sharpness. It is primarily determined by the balance between low and high spectral components, and it increases when more high frequency components are present. Narrowband noise in the critical band of 1 kHz at a sound pressure level of 60 dB serves as reference. This sound event will produce a sharpness of “1 acum”. The sharpness of arbitrary signals is obtained by calculating a kind of centre of gravity of the specific loudness.

$$S = 0.11 \frac{\int_0^{24\text{ Bark}} N' g(z) z dz}{\int_0^{24\text{Bark}} N' dz} \text{ acum} \quad (6.1)$$

with S denoting the sharpness and N' the specific loudness in Bark scale (z). $g(z)$ is a weighting factor.

6.3.5 Fluctuation Strength

Slow amplitude modulations are perceived as fluctuations. Fluctuations should not be faster than 20 Hz in order to be identified as a modulation of a clearly perceivable steady sound. Modulations above 20 Hz cause a sensation of roughness (see below).

At about 4 Hz, the ear has largest sensitivity to amplitude modulations. Thus the reference signal for fluctuation strength is a pure tone at 1 kHz and 60 dB with an amplitude modulation of 100% and 4 Hz. This corresponds to a fluctuation strength of 1 vacil, according to the following equation:

$$F = \frac{0.008 \int_0^{24\text{Bark}} (\Delta L / \text{dB}) / \text{Bark} dz}{\left(\frac{f}{4\text{Hz}}\right) + \left(\frac{4\text{Hz}}{f_{\text{mod}}}\right)} \text{ vacil} \quad (6.2)$$

F is the fluctuation strength and ΔL the spectral masking pattern evaluated from the temporal evolution of the masking curve.

6.3.6 Roughness

Fast amplitude modulations can no longer be perceived as fluctuations. Their effect can rather be described by the sensation of roughness. To get a feeling for the impression of roughness, imagine the voice of Louis Armstrong. Roughness is a significant effect that differs from other psychoacoustic effects. It requires fast amplitude modulations (15 Hz to 300 Hz). The reference signal is a 1 kHz pure tone and level of 60 dB, modulated at 70 Hz by 100%. This stimulus corresponds to 1 asper.

6.3.7 Tonality, Pitch and Pitch Strength

Sounds often have a tonal character. In spectral representation (see also 7.3), they contain discrete frequency components, which correspond to periodicities in time domain. Examples for periodic signals are vowels and technical sounds from machinery with rotating processes (usually described by rotation speed in rpm).

Tonal components are recognized immediately. The human hearing system extracts tonal components from a complex mixed sound event with great efficiency. When humans listen to a complex sound event, the attention is directed towards the tonal component as well. Tonality can be described by the prominence ratio or by spectral level difference between tonal and broadband components (DIN 45681). The tonal character determined by the level difference between adjacent frequency bands in the spectrum is emphasized by adding a constant tonality penalty, up to 6 dB, depending on the type of sound.

Pitch is a subjectively perceived quantity, and it is extremely difficult to characterize it in a general way (Terhardt 1974). It is easily understood when pitches of tones are compared. The pitch of a 1 kHz pure tone is expected to be clearly defined and, thus, considered to be “strong”. Other sounds like narrowband noise or high-

pass noise limited at 1 kHz have a certain tonal character, but not the same strength as a pure tone. The fact that the above-mentioned sounds are not periodic at all is rather confusing. It is therefore necessary to study the pitch perception by using models of neural processing, spectral merging of tones complexes, formation of spectral regions of high response in the cochlea, influences of signal duration, their phases and several other factors.

The parameter “pitch strength” can be determined experimentally by using a magnitude scale (Fastl and Stoll 1997). Pitch strength can also be modelled by a frequency transformation model of the peripheral hearing organ (cochlea) and by evaluating the result concerning the maximum spectral amplitude (pitch) and its variance (pitch strength). For details and current results, see (Fruhmann 2004a, b).

6.4 Binaural Hearing

Humans beings can localize sound due to their binaural hearing (Blauert 1996). We assume a person in a free sound field of a plane wave incident from a particular direction. The plane wave is disturbed by reflection and diffraction at the head and torso. The sound signal travelling with the plane wave is thus influenced by linear distortions. These distortions are dependent on direction. In the case of spherical waves, distortions are dependent on the distance, too. The sound signals arrive with differences in time and amplitude at both ears. It takes longer for a sound to travel from a source located at the side of the head to the contralateral ear², and it suffers frequency-dependent damping due to diffraction and absorption. Both effects are noticeable as differences between the ear signals, namely, as interaural time differences, ITD and interaural level differences, ILD.

In the median plane (the vertical plane between the ears along the fontanella), ILD and ITD are small. Therefore the precision of localisation is less than with clear interaural cues. Nevertheless humans can tell the difference between frontal, up, or back direction, based on their evaluation of the monaural cues. Monaural cues are identical at both ears and represent the spectral differences with reference to a free sound field or with reference to a specific direction, usually the frontal incidence.

Head-related transfer functions are defined as a complete and formal way of describing the linear sound distortion caused by the head and torso.

6.4.1 Head-Related Transfer Functions

The monaural and binaural cues are introduced into the eardrum sound pressure by diffraction of sound incident at the head and torso. Usually a plane wave is used as

²The ear in the shadow zone of the head.

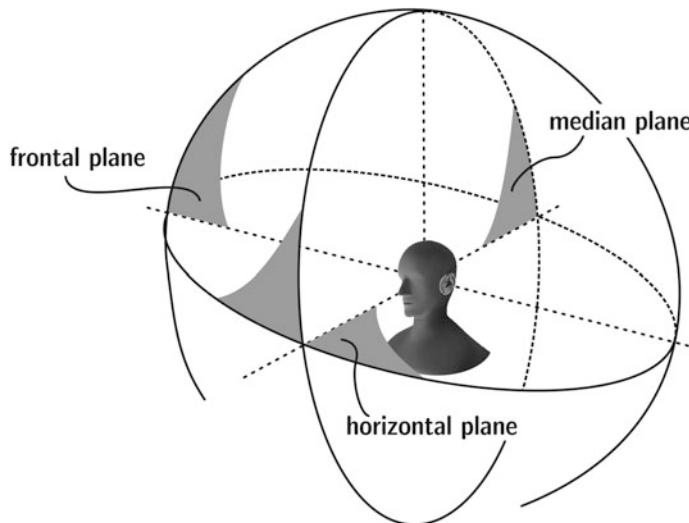


Fig. 6.4 Head-related coordinate system

reference. The amount of diffraction is described by the head-related transfer function, HRTF, the details of which were discovered and described in pioneering work by (Shaw 1982). It is defined by the sound pressure measured at the eardrum or at the ear canal entrance and divided by the sound pressure measured with a microphone at the centre of the head with the head absent. Accordingly, HRTF are dependent on the direction of sound incidence (Fig. 6.4).

The principal components of HRTF are above 200 Hz, where the linear sound field distortion due to diffraction becomes significant. The head and shoulder affect the sound transmission into the ear canal at mid frequencies, whereas the pinna contributes to distortions in the higher frequency range (above 3 kHz). It should be mentioned that HRTF are dependent on the individual anatomical dimensions.

A specific coordinate system is required for an exact definition of the direction of sound incidence. The horizontal plane is described using an azimuth angle, φ , counterclockwise between 0 degrees (front direction) and 360 degrees and using a polar angle, θ , between 0 degrees (front direction) and 90 degrees (upwards the vertical plane). The polar angle can also be extended towards negative values, in order to represent sound incidence from the lower hemisphere. For example, $(\varphi, \theta) = (+90, -10)$ degrees denotes a sound incident from the left just below the horizontal plane.

Figure 6.5 illustrates several typical HRTF. They were measured with four test subjects, frontal incidence. Figure 6.6 shows an example of a HRTF and the corresponding temporal function, the head-related impulse response. The direction of sound incidence is from the left-hand side. In the frequency responses, this fact can be identified from the level difference between the curves for the left (black) and

Fig. 6.5 Head-related transfer functions of four human test subjects, frontal incidence

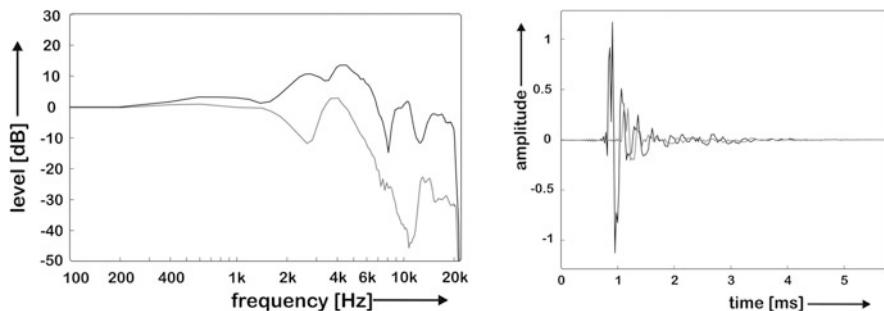
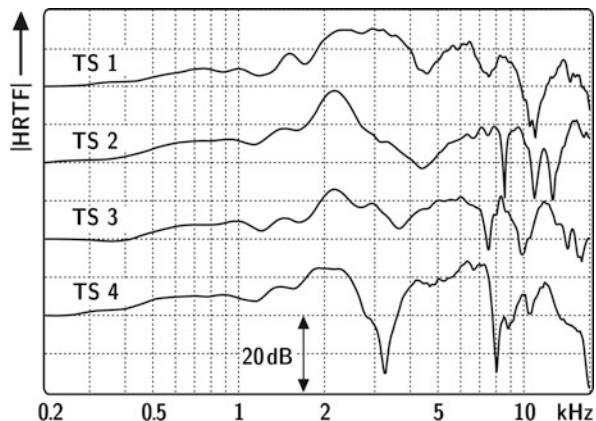


Fig. 6.6 Head-related transfer function, HRTF, (left) for sound incidence from the left and corresponding head-related impulse response HRIR (right)

the right (grey) ear. Furthermore the impulse responses highlight the fact that the sound arrives at the left ear first and with greater amplitude, while at the right ear, the sound is arriving delayed and reduced by shadowing (diffraction).

The just noticeable differences (also called localization blur) depend on the signal presented and on the direction of sound incidence. See Chap. 2 in (Blauert 1996) for more details. The jnd in the horizontal plane are the smallest. In approximation, the order of magnitude of the localization performance of humans is 1 degree in frontal direction ($\varphi = 0^\circ$), 10 degrees on right/left axis ($\varphi = 90^\circ$ or 270°) and 5 degrees in back direction ($\varphi = 180^\circ$). In the median plane at $\theta = 90^\circ$, the jnd is about 20 degrees.

The reason for a more accurate localization in the horizontal plane is that interaural time differences are available, whereas in the median plane just monaural (spectral) cues are fed into the binaural hearing system. Spatial hearing improves the performance because of interaural differences which offer great advantages when it comes to localizing sources, hearing in noisy environments in general and noise suppression in particular (precedence effect (Blauert 1996) and “Cocktail party effect”). The precedence effect (law of the first wavefront) can be explained by

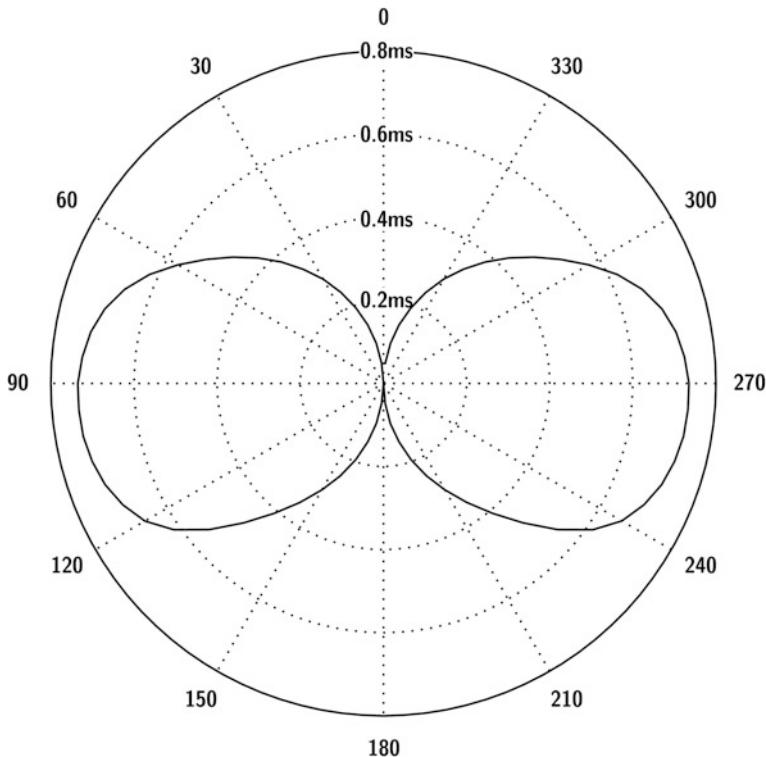


Fig. 6.7 Example of ITD for sound incidence in the horizontal plane

temporal effects neural processing with preference of early sound incidence. The cocktail party effect is a binaural effect of noise suppression due to spatial separation of sound sources from several directions or from speech babble in diffuse fields. The main effect is that speech intelligibility of our specific communication partner is enhanced, as if the speech level was increased (binaural intelligibility level difference) (Fig. 6.7)³.

6.4.2 Artificial Heads

In the 1960s, the pioneering work done by (Burkhard and Sachs 1975) and (Shaw 1982) was presented. In their studies, they aimed at creating a standard dummy head and standard ear simulators for measurements and simulations. KEMAR was introduced in 1972. At first it was mainly used for measurements of hearing aids under in

³The specific communication partner needs not to be in front direction.

situ conditions. KEMAR, however, was used as reference dummy head in a wider sense as well. Nowadays it still serves as one of the reference databases for HRTF⁴.

Together with the properties of head and torso, the conditions of the ear canal must be described, i.e. the open-ear condition with impedance simulator (ear canal plus eardrum) or blocked ear canal.

The main problem of research and development that needed to be solved at that time was the definition and creation of an average head representing an “average listener”. The aim, of course, was to find the most appropriate head and torso dimension for standardization. The result, however, might not be the optimum in the performance for localisation tests with an arbitrary individual listener (see below). This problem was supposed to be solved with various kinds of geometric or structural averaging.

Further research aimed at a technical description of the head and torso geometry, however, with the same acoustic behaviour as natural replicas of torso and pinna. (Genuit 1984) developed a mathematical diffraction model based on elliptical and cylindrical elements. The pinna is also simplified to the cavum conchae and the asymmetric ear canal entrance. The exact location of the ear canal entrance point, mostly used as reference point for measuring sources close to the ear, is very important to obtain correct localisation cues.

When using the standard dummy heads for recordings and for research that focuses on perception of direction and distance, a rather large portion of the test subject population reports on disturbances of the listening experience. The test results usually suffer from front-back confusion and large uncertainties in localisation in the median plane. Multidimensional analysis (modulus, phase, frequency, ILD, ITD, etc.) comparing HRTF of individuals with the dummy heads reveals the reasons. In some cases, the performance of the test subject and the dummy head are similar and matching quite well, while others' ears differ tremendously from the dummy head ears.

Approaches aimed at developing best-match heads were published by (Schmitz 1995) and by (Christensen et al. 2000). The procedure was not to measure and average the head and torso dimensions but the HRTF of a large population of test subjects and to use this database of HRTF in localization listening tests. The individual HRTF with the best performance in localisation tests among the largest group of test subjects is obviously the “best choice HRTF” and, accordingly, belongs to the individual person with best matching head and torso geometry. Please note that this approach is not based on a geometric or structural average but on a selected HRTF including significant HRTF details like peaks and notches.

⁴HRTF data can be found on the Internet, for example:

KEMAR: <http://sound.media.mit.edu/resources/KEMAR.html>

Spatially Oriented Format for Acoustics: <https://www.sofaconventions.org/mediawiki/index.php/Files>

ITA dummy head: <https://www.akustik.rwth-aachen.de/go/id/pein>

ITA individuals: <https://www.akustik.rwth-aachen.de/go/id/lslv>

Recent results showed that a well-selected human head (fitted with probe microphones) is superior to dummy heads. Dummy heads which were created from an individual selection process and a copy of an individual human (rather than from an average) are almost as good as human heads.

The selection of test subject population represented by any dummy head is a crucial process. All studies presented so far suffer from a somewhat arbitrary choice and limited number of subjects. The most comprehensive study in this respect is still (Burkhard and Sachs 1975). Today, it might be argued that the human population changed over the last three decades (usually the standard group of normal hearing subjects consists of adults of age 18–25, who are possibly taller than they used to be a couple of years ago). And one might raise the question whether one standard dummy head might be sufficient to cover all human beings or not.

Today, dummy heads and binaural technology are of high quality. International standards for telecommunication technology and for measuring hearing aids are available. In recent years, many studies dealing with this state-of-the-art technology, that is also used for applications in research and development, were published. Apart from that dummy heads were also used for recordings and reproduction for home entertainment applications. This achieved, however, only modest commercial success. Several other fields such as 3D audio, virtual environments, headphone development and sound quality could be created and extended with high relevance to applied acoustics and audio engineering. However, some questions still remain concerning the applicability of average head dimensions for both sexes, humans from all over the world, and particularly for children⁵.

Individual HRTF

The data collection of individual HRTF is a challenge concerning the long duration of data acquisition. The conventional measurement of HRTF is to move a loudspeaker at the desired position around the listeners head. A measurement procedure using the conventional method is very time consuming. Therefore, the requirements for an individual HRTF measurement are high-speed data acquisition with still broadband frequency spectrum and high spatial resolution. With new signal processing available and appropriate loudspeaker arrays, the sequential mechanical scanning can be replaced by fast methods of measurement signal processing (Majdak et al. 2007; Masiero et al. 2011; Masiero 2013).

For example, an HRTF arc consisting of an array of 64 loudspeakers (designed to act almost like an array of point sources) is placed around the listener's head (Fig. 6.8). The challenges of such a measurement is centring of the head and post processing of the HRTF data for any kind of subsequent application such as interpolation or distance extrapolation.

⁵Which is, for example, relevant for fitting hearing aids.

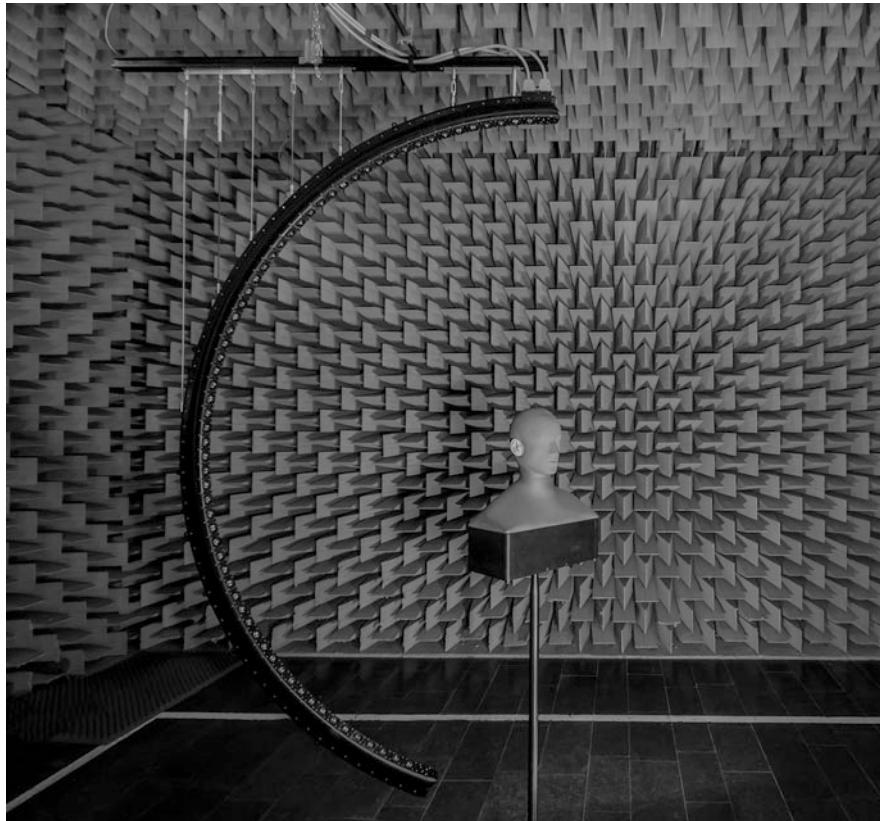


Fig. 6.8 Loudspeaker arc for measuring HRTF (Richter and Fels 2019)

Head Movements

Head movements are essential if a front-back confusion affects a precise localization of the source (Börger et al. 1977; Mackensen 2003). In virtual sound environments, it must therefore be possible to move and rotate the head slightly to enhance the localization precision, see 0. Current research thus focuses on the dynamic features in HRTF by looking at relative head and torso movement, like in normal behaviour to enhance sound localization accuracy. The degree of freedom is then extended towards a larger number of spherical coordinate systems related to head orientation in horizontal and vertical axes and the torso orientation (Moldrzyk et al. 2004).

6.5 Hearing in Rooms

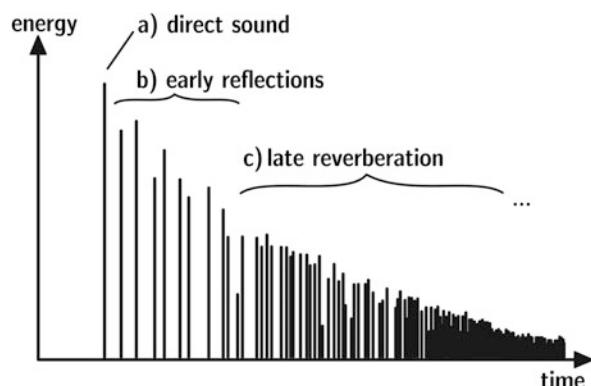
The psychological aspect of room acoustics is a very complex problem. On the one hand, the sound field in rooms is very complicated, and on the other hand the subjective impressions of listeners in rooms are multi-dimensional and influenced by many factors, some of which are non-acoustical. Psychological room acoustics tries to bridge a gap between those subjective impressions and objectively measurable data. The most dominant effect of hearing in rooms is the sensation of reverberation. Reverberation time, thus, both its average value and its frequency dependence is the most important quantity for acoustic characterization of rooms. Apart from reverberation, there are other specific quantities that are important for description of the overall subjective effect in rooms. Unlike reverberation, they depend on the specific listener position in the room.

According to the concept of geometrical acoustics, the reverberation in a room is the result of sound reflected numerous times at the room boundaries. A listener, thus, notices not only the direct sound but a series of delayed reflections, too. The response of the room to an excitation with an ideal impulse, therefore, serves well as a basis for an interpretation of room acoustics. A typical example for a room impulse response (“energy time curve”) is depicted in Fig. 6.9.

The components in the energy time curve are arranged in a characteristic series of pulses. The pulse density increases to with t^2 along the time axis, while the pulse energy decreases with t^{-2} and with absorption. In short-time averages, the quadratic time functions are cancelled, and the net energy loss is given by the exponential decay due to absorption.

The first impulse, the direct sound, determines the perceived direction of sound incidence (precedence effect). Reflections as secondary components are delayed due to the longer path of sound propagation. Even in the case of secondary sound with higher level than the direct sound (up to 10 dB), which might occur with secondary loudspeaker sound or when the direct sound is blocked by a barrier, the localisation

Fig. 6.9 Energetic room impulse response



will still be determined by the first arriving component (Haas effect). However, the reflections within an interval of so-called “early reflections” contribute in a specific way to the direct sound impression. They enhance the loudness and support the intelligibility of speech, the clarity of music and the impression of the auditory source width. All reflections delayed by more than 50–80 ms build up the reverberation in its more specific meaning. Important aspects of these late reflections are that they create impressions of reverberance and listener envelopment.

6.5.1 Reverberance

The sensation of reverberance must be analysed with regard to the presented signal. Music and speech do have transient components. However, the mean syllable duration and also the mean tempo of music makes it impossible to follow the reverberation over a level drop of 60 dB. Instead, the early part of the reverberation has more significance for the perceived reverberance than the later part. That is why several reverberation times were introduced which evaluate early parts of the energy impulse response⁶ curve (Fig. 6.9).

The average decay in the impulse response envelope, however, is not in accordance with the original definition of the reverberation time introduced by W.C. Sabine. Reverberation time was defined with regard to a switch-off of a so far steady-state signal (originally an organ pipe). An impulse response, in contrast, is related to a short pulse excitation.

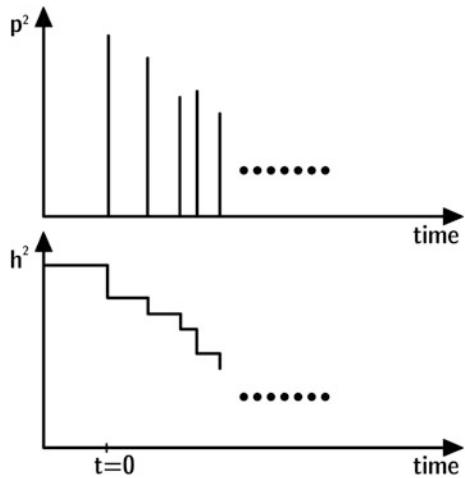
The way to transfer results from impulse excitation to steady-state signals with switch-off is the integrated impulse response method. An excitation of the room by steady-state signal can be described as permanent excitation with densely repeated pulses. Thus, the response to steady-state excitation is a sum of all components of the impulse decay. In a stationary case, the direct sound is present all the time (because it is created steadily), and each of the early reflections is present permanently (because it was radiated in the past, just to arrive in time with the direct sound). The same is true for all reflections. Retardation and coincidence⁷ of all components in the end yield an integration of the impulse response. The value of the integral is the steady-state energy of the continuous excitation.

When modelling the switch-off, the energy components (contained in the impulse response) must be subtracted according to their delay. The first “missing” sound after switch-off is the direct sound, while all others are still present in the total sound field, since they were radiated before switch-off time. Afterwards, however, the energy from the temporal series of reflections will vanish one by one. This process is

⁶The term “impulse response” will be introduced and based on signal theory described in Sect. (7.3.2).

⁷Delayed components vs. coincident components.

Fig. 6.10 Top: Energetic impulse response, bottom: Integrated Impulse Response. Note the temporal coincidence of impulses and subtraction steps



modelled by the so-called Integrated Impulse Response (Schroeder 1965). It consists of two steps (see Fig. 6.10):

- (a) Integration of the energetic impulse response, $p^2(t)$

$$C = h^2(t < 0) = N_0 \int_0^\infty p^2(\tau) d\tau \quad (6.3)$$

- (b) Switch-off at $t = 0$: Subtraction

$$\begin{aligned} h^2(t) &= C - N_0 \int_0^t p^2(\tau) d\tau = N_0 \left[\int_0^\infty p^2(\tau) d\tau - \int_0^t p^2(\tau) d\tau \right] \\ &= N_0 \int_t^\infty p^2(\tau) d\tau \end{aligned} \quad (6.4)$$

The reverberation time⁸ can now be calculated based on the decay curve, h^2 . The perceived reverberance which is correlated with the Early Decay Time, EDT, can now be evaluated as well. It is defined as six times the time interval corresponding to the average decay between 0 dB and –10 dB (Atal et al. 1965).

⁸For measurements, the standard reverberation time is based on a logarithmic slope regression between –5 dB and –35 dB.

6.5.2 Strength

As explained in the previous section, the total energy for the steady-state case is contained in the integral of the impulse response. The main energy, however, is concentrated in the early part. For example, the energy contained in the reverberation tail later than half of the reverberation time contributes only 22% of the total energy. The total energy density expressed in decibels is hence too small by just 1 dB (which corresponds to the just audible difference), if the second half of the reverberation tail is neglected.

The total energy can be expressed independent of the sound power of the source. By choosing a reference distance in free sound field, the parameter “strength”, G , is introduced. The reference distance is 10 m. If we assume an average distance in a room of 10 m as well, G denotes the gain of the room, compared with the free field propagation of direct sound.

$$G = 10 \log \frac{\int_0^\infty p^2(t) dt}{\int_0^\infty p_{10m}^2(t) dt} \quad (6.5)$$

$p(t)$ denotes the sound pressure of the room impulse response and $p_{10m}(t)$ the reference sound pressure in 10 m distance with the same source in free field. The integration limit ($t = 0$) is defined to be the arrival time of the direct sound (including the direct sound).

6.5.3 Speech Intelligibility and Transparency

As the early reflections support speech intelligibility, their energy sum can be used to characterize the syllable intelligibility. To simplify things, the lower integration limit (arrival time of the direct sound) is set to zero. This is used to introduce the parameter “definition”, D , for speech

$$D = \frac{\int_0^{50ms} p^2(t) dt}{\int_0^\infty p^2(t) dt} \quad (6.6)$$

A similar expression is obtained by introducing the “clarity” index, C (Reichardt et al. 1974). It is defined as C_{80} for music and C_{50} for speech. With C_{80} we can correlate the impression of transparency in fast pieces of music, the ability to recognise musical details behind the “curtain” of reverberation.

$$C_{80} = 10 \log \frac{\int_0^{80ms} p^2(t) dt}{\int_{80ms}^{\infty} p^2(t) dt} \quad (6.7)$$

$$C_{50} = 10 \log \frac{\int_0^{50ms} p^2(t) dt}{\int_{50ms}^{\infty} p^2(t) dt} \quad (6.8)$$

C_{50} , by the way, is related to D by the equation

$$C_{50} = 10 \log \left(\frac{D}{1-D} \right) \quad (6.9)$$

The balance between early and late parts of the impulse response can also be expressed by using the first moment of the impulse response so that we can define the “centre time”, t_S .

$$t_S = \frac{\int_0^{\infty} t p^2(t) dt}{\int_0^{\infty} p^2(t) dt} \quad (6.10)$$

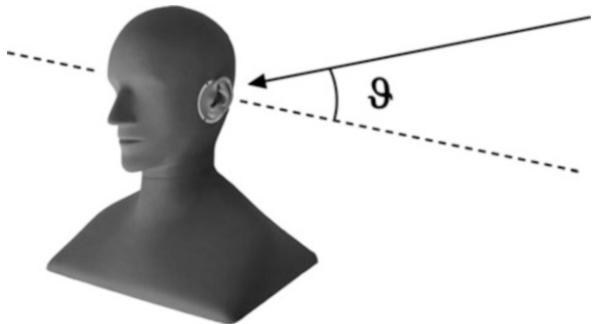
Definition, clarity and centre time are interrelated and correspond to the clarity of speech and music. Which parameter is in fact best is not clear. Extensive listening tests and questionnaires used in the laboratory and in music halls have shown that all quantities are useful to describe this specific auditory impression of transparency related to the linguistically or musically basic elements like syllables or notes.

Another well-known approach to describe speech intelligibility is the evaluation of the modulation of the speech signal (Houtgast and Steeneken 1973). The characteristic modulation in speech signals can be affected by reverberation and by background noise. The so-called speech transmission index, STI, expresses the degree of changes in modulation depth caused by reverberation and by noise. The basic function in this evaluation is the modulation transfer function, MTF. The MTF is a ratio of the spectra of the envelope time signal of the original and the disturbed signal. Values smaller than 1 indicate a reduction of speech intelligibility. The spectral differences can, furthermore, be discussed with regard to broadband or high-frequency components. While background noise affects the STI in all frequency bands, reverberation affects mainly high frequencies.

6.5.4 Spatial Impression

The subjective impression of spaciousness is linked to lateral reflections (Barron 1971). Purely energetic integration is, thus, not sufficient to characterize this effect. Furthermore, early and late lateral reflections create two kinds of spatial impression.

Fig. 6.11 Sound incidence from lateral reflections



The early part, up to 80 ms, contributes to “ASW”, the auditory source width. As described above, source localization in complex sound fields is evaluated by the human hearing system so that the first arriving sound event determines the perceived direction of sound incidence (precedence effect). Early lateral reflections add some uncertainty to the localization. The source is not localized at an exact position but related to an extended source with a characteristic width, ASW. The objective parameter which correlates well with this impression is the early lateral fraction, LF.

$$LF = \frac{\int_{5ms}^{80ms} p_{\text{grad}}^2(t) dt}{\int_0^{80ms} p^2(t) dt} \quad (6.11)$$

with p_{grad} denoting the sound pressure impulse response obtained using a gradient (figure-of-eight) microphone oriented to the horizontal axis through the ears of the listener (Fig. 6.11).

Another dimension of spatial impression is the listener envelopment, LEV. It is caused by late lateral reflections. The impression of envelopment is described by the objective parameter of late lateral strength, LG.

$$LG = 10 \log \frac{\int_{80ms}^{\infty} p_{\text{grad}}^2(t) dt}{\int_0^{\infty} p_{10m}^2(t) dt} \quad (6.12)$$

When sound arrives at the listener’s ears from lateral directions, it will lead to a loss in correlation between in the binaural pattern. Particularly interaural differences occur. Hence the interaural correlation can be used as well to describe spatial measures. The interaural cross-correlation function between the sound pressure signals at the right and the left ear of a listener or a dummy head, p_r and p_l , respectively, is defined as follows.

$$\text{IACF}_{t1,t2}(\tau) = \frac{\int_{t_1}^{t_2} p_l(t) \cdot p_r(t + \tau) dt}{\sqrt{\int_{t_1}^{t_2} p_l^2(t) dt \int_{t_1}^{t_2} p_r^2(t) dt}} \quad (6.13)$$

Its maximum between $-1 \text{ ms} < \tau < 1 \text{ ms}$ is denoted “IACC”, interaural cross-correlation coefficient,

$$\text{IACC}_{t1,t2} = \max |\text{IACF}_{t1,t2}(\tau)|. \quad (6.14)$$

Using the integration interval in Eq. (6.13), early and late lateral reflections can be treated separately, which thus leads to an additional independent objective parameter for ASW and LEV.

A simulation and auralization of room sound fields should produce as many cues of the above-described kind as possible. These parameters of hearing in rooms have been studied in laboratory tests and in real rooms. These tests have shown the extent to which human listeners can tell different situations apart. Like in other psycho-acoustic areas, the just noticeable differences, jnd, provide important information about the necessary precision and simplification allowable in computer models. In the standard ISO 3382-1, which describes measurement procedures in room acoustics, a table of jnd is given (Table 6.1).

6.5.5 Spatial Variations in a Room

The energy time curve is not one individual characteristic function of the room, but it depends on the position in the room. Accordingly the listening impression differs from position to position in the room. This fact is important for any kind of room simulation and auralization. The spatial resolution of a simulation must represent the natural listening experience related to noticeable differences from one position to another.

The physical approach to local variations is given in the application of the correlation function (coherence) between two room impulse responses. In an ideal

Table 6.1 Just noticeable differences of subjective room acoustical impressions (see Vorländer 1995; ISO3382-1))

Subjective listening aspect	Acoustic quantity	Main frequency range (Hz)	jnd
Level	G (dB)	500 to 1000	1 dB
Reverberance	EDT (s)	500 to 1000	5%
Clarity, Definition	C ₈₀ (dB)	500 to 1000	1 dB
	D	500 to 1000	0.05
	TS (s)	500 to 1000	10 ms
ASW	LF	125 to 1000	0.05
LEV	LG (d)	125 to 1000	1

diffuse sound field in a room, this coherence between the two sound pressure signals, p_1 and p_2 , at a distance d is known to be (Kuttruff 2016)

$$\Psi_{p1,p2}(x) = \frac{\sin kd}{kd} \quad (6.15)$$

If we assume that our sound signal has a frequency content dominant at low to mid frequencies between 50 Hz and 500 Hz, the physical coherence function tells us that it is not necessary to discuss a local spatial resolution smaller than 10 cm.

The listening experience in rooms is even more robust. The total auditory impression in a concert hall or opera house is not different from one seat to the next. In a scale of a two-seat distance or more we would expect to notice differences, but not on a smaller scale. The coverage of the room acoustical quantities described above and the necessary spatial resolution of simulation and auralization algorithms is, at least typically, in the order of magnitude of 1 m.

6.5.6 Estimation of the Monaural Subjective Parameters

Before starting to discuss a detailed sound field simulation in Chap. 11, we have to analyse the order of magnitude of the subjective parameters in a room of volume, V , surface, S , and reverberation time, T . This kind of estimation yields no exact result but only the order of magnitude of room acoustic parameters. We might use this information to define expectation values and a reference for checking the plausibility of simulation results. If an auralization is acceptable with just an average plausible room effect, the controls of delay lines and reverberation processors can be adjusted to match their artificial impulse response with regard to the estimated room acoustic parameters as well.

For an approximate approach, assuming a diffuse sound field and the prerequisite exponential decay function (Sect. 4.5.1), the strength and other parameters in a room can be roughly estimated. The lower integration limit of the series of reflections is set as an average to $t_0 = 1/\bar{n}$. This choice is reasonable since the mean time delay of a first reflection from any room point to any receiving point is proportional to the mean free path. The direct sound of a monopole source is added explicitly. The distance between source and receiver is denoted by r (in m), the room has a volume of V (in m^3), a total surface of S (in m^2), a reverberation time of T (in s) and an equivalent absorption area of A (in m^2). In accordance with Eq. (4.39), we find for the strength

$$G = 10 \log \left(\frac{1}{r^2} + 310 \frac{T}{V} \right) + 4.34 \frac{A}{S} + 20$$

$$G = 10 \log \left(\frac{1}{r^2} + \frac{50}{A} \right) + 4.34 \frac{A}{S} + 20 \quad (6.16)$$

or as an approximation for large distances r

$$G = 37 - 10 \log A + 4.34 \frac{A}{S}, r \gg \sqrt{A}/7 \quad (6.17)$$

Furthermore, the early (within 50 ms) and late energy density can be calculated as follows, similar to Barron's revised theory (Barron 2000).

$$w_{\text{late}} = \frac{1}{V} \int_{0.05s+1/n}^{\infty} e^{-13.8t/T} dt = \frac{T}{13.8V} e^{-(\frac{0.69}{T} + A/S)} \quad (6.18)$$

$$\begin{aligned} w_{\text{early}} &= \frac{1}{V} \int_{1/\bar{n}}^{1/\bar{n}+0.05s} e^{-13.8t/T} dt + \frac{1}{4\pi c r^2} \\ &= \frac{T}{13.8V} \left(e^{-A/S} - e^{-(\frac{0.69}{T} + A/S)} \right) + \frac{1}{4\pi c r^2} \end{aligned} \quad (6.19)$$

Accordingly, definition, clarity and centre time result in

$$\begin{aligned} D &= \frac{\frac{T}{13.8V} \left(e^{-A/S} - e^{-(\frac{0.69}{T} + A/S)} \right) + \frac{1}{4\pi c r^2}}{\frac{T}{13.8V} e^{-A/S} + \frac{1}{4\pi c r^2}} \\ &\approx \frac{\frac{T}{13.8V} \left(1 - e^{-0.69/T} \right) + \frac{1}{4\pi c r^2}}{\frac{T}{13.8V} + \frac{1}{4\pi c r^2}} \end{aligned} \quad (6.20)$$

$$\begin{aligned} C_{80} &= 10 \log \frac{\frac{T}{13.8V} \left(e^{-A/S} - e^{-(\frac{1.104}{T} + A/S)} \right) + \frac{1}{4\pi c r^2}}{\frac{T}{13.8V} e^{-(\frac{1.104}{T} + A/S)}} \\ &\approx 10 \log \left(e^{\frac{1.104}{T}} \left(1 + \frac{13.8V}{4\pi c r^2 T} \right) - 1 \right) \text{ for } A \ll S \text{ or } \bar{\alpha} \ll 1 \end{aligned} \quad (6.21)$$

$$\begin{aligned} t_S &= \frac{T}{13.8} \left(\frac{A}{S} + 1 \right) \\ &\approx \frac{T}{13.8} \text{ for } A \ll S \text{ or } \bar{\alpha} \ll 1, \end{aligned} \quad (6.22)$$

the latter with the influence of the direct field neglected.

Chapter 7

Signal Processing for Auralization



7.1 Introduction

Signal theory provides the basis for coupling of computer simulation of sound sources and sound propagation. It is an important tool for bridging the gap between computer data and audible sound, which can be played through the soundcard of a PC, for example. Such tools are the classical digital audio processing components of sampling, analogue-digital conversion and transforming, filtering and analysing sound signals, and they are introduced in the context of signal processing in virtual acoustics technology.

The fundamentals and techniques of theoretical and engineering acoustics as introduced in the preceding chapters, with all available analytic calculation models, allow the prediction of generation and radiation of sound. Psychoacoustics provides performance models of the human auditory system and a focused evaluation of the technical character of sound. Accordingly, numerous methods for analyses of acoustic signals are available. As described in the previous chapter, they can be based on signal theory, wave field physics or psychoacoustics. Sound can yield information on sound energy (level), spectral information including masking, temporal attributes, spatial cues and specific parameters related to room acoustics. This set of analytic tools should be sufficient for all kinds of acoustical problems. Or is it not?

The crucial point is that any figure extracted from acoustic signals can indeed represent an average impression, at least approximately. The subjective sound event as such is, however, only covered by a full auditory experience. The perception, the impression, the interpretation and the meaning of sound are not covered by this technical approach. The full characterization and interpretation of sound, in the end, can only be achieved when hearing and other senses are directly involved. Therefore the technique of auralization offers an important addition to acoustic analysis and synthesis, prediction and rating. It involves the listener directly without the need to

explain the meaning of acoustic events verbally. It represents an important component of multimodal sensation and corresponding psychological effects.

7.2 The Concept of Auralization

Auralization is the technique to create audible sound files from numerical (simulated, measured, synthesized) data.

The principle of auralization is illustrated in Fig. 7.1. It shows the basic elements of sound generation, transmission, radiation and reproduction. The figure indicates that the coupling between the blocks requires attention. In room acoustics, for instance, we rarely find an effect of feedback to the source. The radiation impedance is typically not affected by the room. Nevertheless the source, if it is a person, will adapt his or her singing or musical playing based on the room response. This is, however, not a problem of physical feedback, but of psychological response. In a purely physical sense, the signal flow can be modelled in the forward direction only. In contrast, in problems of structure-borne sound, the situation changes completely. The vibration velocity and displacement in beams and plates depend on the kind of source and the contact admittance of the components; see the back arrow in Fig. 7.1.

If the interface between the source signal and the transmission chain is clearly defined in a robust way, the acoustic situation can be transformed into a signal flow model. “Robust” in this respect means that the interface will transfer the same velocity or pressure when sources or transmission elements are changed. The signal flow model is typically represented by a two-port model, the components of which are determined by simulation or measurement. If the transfer functions of the elements are obtained by calculation or measurements, then the signal transmitted in the structure, duct, room or free field can be processed by convolution.

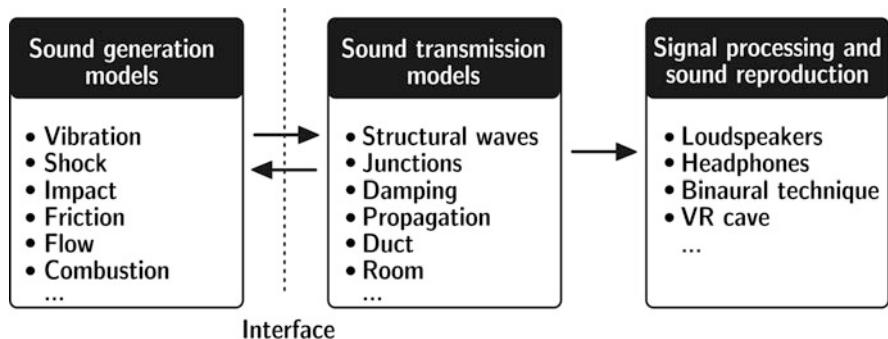


Fig. 7.1 Principle of auralization

This looks simple at first glance, but the task of generating an appropriate filter becomes more difficult when more details are taken into account. For more detailed illustration, some examples are given in the following paragraph. Obviously the auditory quality requirements of the signal used in a listening test should be high: the bandwidth and the corresponding sampling rate, the colouration and the corresponding quality of the reproduction system, relevance of the direction of sound incidence, perceived distance of the sound event, a specific room impression, source characteristics and movement of the source or the receiver, just to name a few elements.

The technique of auralization and its result, a sound file, must take all these aspects into account, depending on the specific application. A basic task in this respect is the identification of relevant signal paths, the degrees of freedom of vibration in structural paths and the identification of interfaces between sound and vibration (Fig. 7.2).

A historic example was mentioned in the preface. In 1929, Spandöck and colleagues tried to model a room for speech and music performance in Munich. The basic idea was to use a 1:10 scale model of the room under test, to play music and speech into the model at scaled frequencies, to record the result in the scale model and to reproduce it by rescaling the frequency content of the signal down to the real scale.

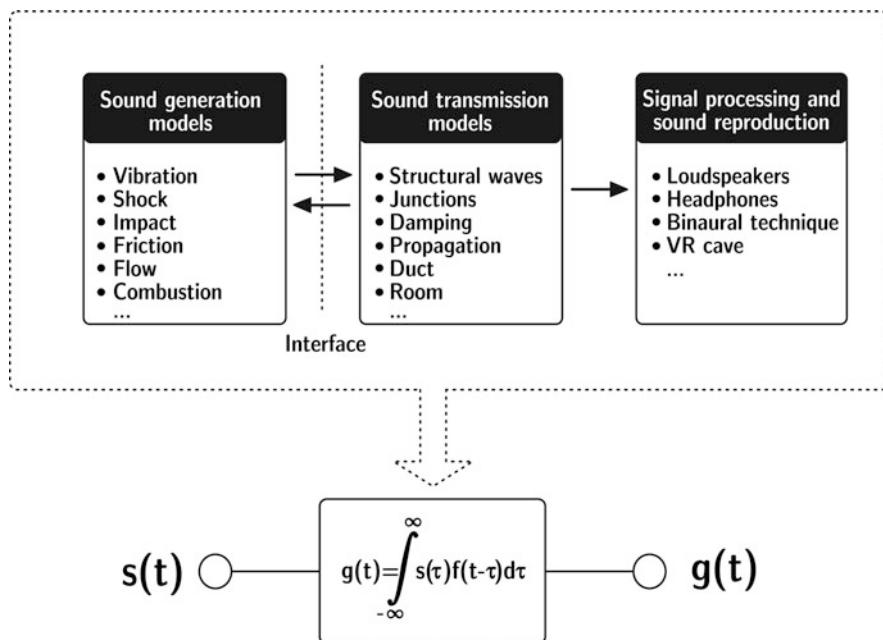


Fig. 7.2 Convolution of source signal $s(t)$ with a filter impulse response $f(t)$ to obtain a receiver signal $g(t)$

Today, with powerful computers available, the components of the auralization are usually obtained by computer simulation. Nevertheless, some problems in acoustics and vibration may exceed feasibility. Measurements of sources and/or transfer paths are an indispensable prerequisite for an auralization for industrial application or for research. Any kind of determination of sound and vibration transfer functions from the source(s) to the receiver can be integrated into the concept of auralization.

Before we concentrate on specific models for simulations of acoustics and vibration in the next chapter, the technique of auralization shall be further introduced in an overview.

Starting with the source description, a primary signal is created or recorded. This primary signal may represent a volume flow of a point source, the sound power and directivity of an extended source or of distributed sources or the blocked force output or the free velocity of a vibration source, for instance. The primary signal must be made available in amplitude scale, in units of sound pressure or volume flow, for instance. Then the primary sound can be fed into the transmission path. The result will be a transmitted sound pressure signal which can be considered as perceivable and ready for sound reproduction (for instance, over headphones). The steps that are required for a proper auralization are performed by using tools of the field from signal processing. The transfer function obtained by simulation (or measurement) is, accordingly, interpreted as the transfer function of a “filter”.

The procedure of convolution is the basis of signal analysis and processing. It is related to linear time-invariant systems.

7.3 Fundamentals of Signal Processing

Nearly all sound-transmitting systems in acoustics can be approximated by linear time-invariant systems. By definition, these systems transmit sound in a repeatable way, independent of the actual starting time of the acoustic excitation. With the term linearity, we describe the fact that linear superposition holds.

7.3.1 Signals and Systems

A so-called signal in the sense of signal theory is the time-dependent function of a scalar physical quantity. In our case it might be sound pressure, vibration velocity or a similar signal.¹ We denote this function with $s(t)$ in the analogue (real) world and $s(n)$ in the digital representation in the computer, respectively. This signal can be recorded or simulated, transmitted over a system, changed in some way by a system and finally received by a sensor or a human being. A linear system affects signals in a

¹Output from any kind of sensor

linear way, which means that signal superposition can be treated as linear combination. Amplification just results in an amplitude change. For any transmission² (transformation Tr) of a signal fed into a system, the following holds

$$\text{TR} \left\{ \sum_i (a_i \cdot s_i(t)) \right\} = \sum_i (a_i \cdot \text{TR} \{s_i(t)\}) = \sum_i (a_i \cdot g_i(t)) \quad (7.1)$$

with s_i denoting the input signals and g_i the output signals, $i = 1, 2, 3, \dots$. We can assume that amplifications, delays, filtering or summations behave as system transformations. The equation means that the transmission of a linear combination of input signals ($a_i s_i(t)$) is equal to the sum of the combined output signals.

Furthermore, the specific behaviour of a system is important. It is time invariant, if for any time shift holds

$$\text{TR}\{s(t - t_0)\} = g(t - t_0) \quad (7.2)$$

By far most systems in acoustics in fact show this behaviour. A loudspeaker radiates a sound pressure proportional to the input current, at least when driven in linear mode (nonlinearities are well-known in loudspeakers, of course, but this happens only at very high sound levels). The variations of a system in time are mostly negligible too (the loudspeaker might change due to heating of the voice coil, but usually this can be neglected in steady state, or we can assume slow variations).

Linearity and time invariance are combined in the expression LTI system. LTI systems can be described with respect to their reaction to signals in time domain and in frequency domain. This reaction is uniquely represented by the impulse response (in time domain) or the stationary transfer function (in frequency domain).

7.3.2 Impulse Response and Transfer Function

An LTI system fed with an input signal $s(t)$ will yield an output signal $g(t)$ with

$$g(t) = \int_{-\infty}^{\infty} s(\tau)h(t - \tau)d\tau = s(t) * h(t). \quad (7.3)$$

$h(t)$ is the impulse response of the system. The operation denotes a convolution integral. This general equation is the basis for all theoretical considerations of LTI systems (Fig. 7.3).

²A “transmission” in a general sense could represent sound propagation in fluid media, transduction (in electroacoustics) or propagation/damping/insulation of sound and vibration in complex structures.

Fig. 7.3 Processing of source signal $s(t)$ with a filter impulse response $h(t)$ to obtain a receiver signal $g(t)$

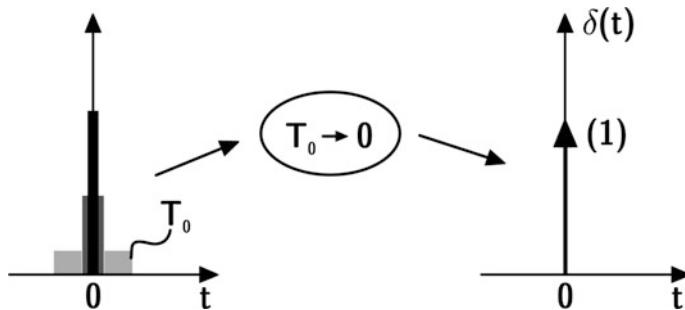
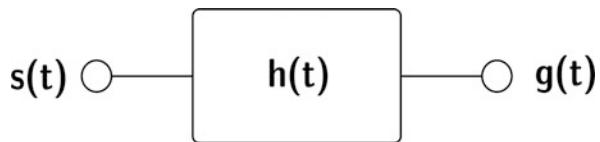


Fig. 7.4 Dirac pulse

It allows in particular the construction of filters. In some examples, it is a direct measure for the system characteristics, for instance, in room acoustics. The Dirac pulse, $\delta(t)$, plays a specific role. It can be intuitively explained by considering the approximation of a set of rectangular pulses of equal area, the width of which tends to zero and the height to infinity (Fig. 7.4):

$$\delta(t) = \lim_{T_0 \rightarrow 0} \frac{1}{T_0} \text{rect}\left(\frac{t}{T_0}\right) \quad (7.4)$$

The Dirac pulse is the impulse response of an ideal transmission system without linear distortions. In this case the output signal is identical with the input signal:

$$g(t) = \int_{-\infty}^{\infty} s(\tau) \delta(t - \tau) d\tau = s(t) \quad (7.5)$$

The convolution algebra for Dirac pulses is very simple. We will need the following examples of rules for Dirac pulses later, particularly for constructing auralization filters:

Multiplication with a Factor (Amplification)

$$a \cdot \delta(t) * s(t) = a \cdot s(t) \quad (7.6)$$

Time Shift (Propagation Path, Delay Line)

$$\delta(t - t_0) * s(t) = s(t - t_0) \quad (7.7)$$

Integration (Step Function $\varepsilon(t)$)

$$\int_{-\infty}^t \delta(\tau) d\tau = \varepsilon(t) \quad (7.8)$$

Excitation of a System with a Dirac Pulse

$$h(t) * \delta(t) = \int_{-\infty}^{\infty} h(\tau) \delta(t - \tau) d\tau = h(t) \quad (7.9)$$

The system performance can also be described by the stationary transfer function, $\underline{S}(f)$. It can be expressed in terms of components' real and imaginary parts ($\text{Re}\{\underline{S}(f)\}$ and $\text{Im}\{\underline{S}(f)\}$) or in an equivalent form as modulus and phase ($|\underline{S}(f)|$ and $\varphi(f)$).

$$\underline{S}(f) = \text{Re}\{\underline{S}(f)\} + j \text{Im}\{\underline{S}(f)\} = |\underline{S}(f)| \cdot e^{j\varphi(f)} \quad (7.10)$$

If the signal modification caused by a system is to be determined, the linear distortion of harmonic input signals is of particular interest. By discussing the damping, delay or amplification of harmonic signals, we can characterize the system by the ratio of the output, $\underline{G}(f)$, and the input, $\underline{S}(f)$. As it is already complex, the steady-state transfer function (related to harmonic signals) $\underline{H}(f)$ is defined as

$$\underline{H}(f) = \frac{\underline{G}(f)}{\underline{S}(f)}. \quad (7.11)$$

In an experiment, we excite the system directly with a pure tone, equivalent to an infinite stationary harmonic signal, provided the system is responding in steady state. Determination of the response amplitude and phase and calculation according to Eq. (7.11) yields the transfer function at this frequency. Repetition of this procedure in certain frequency steps provides a sample of the transfer function (Fig. 7.5).

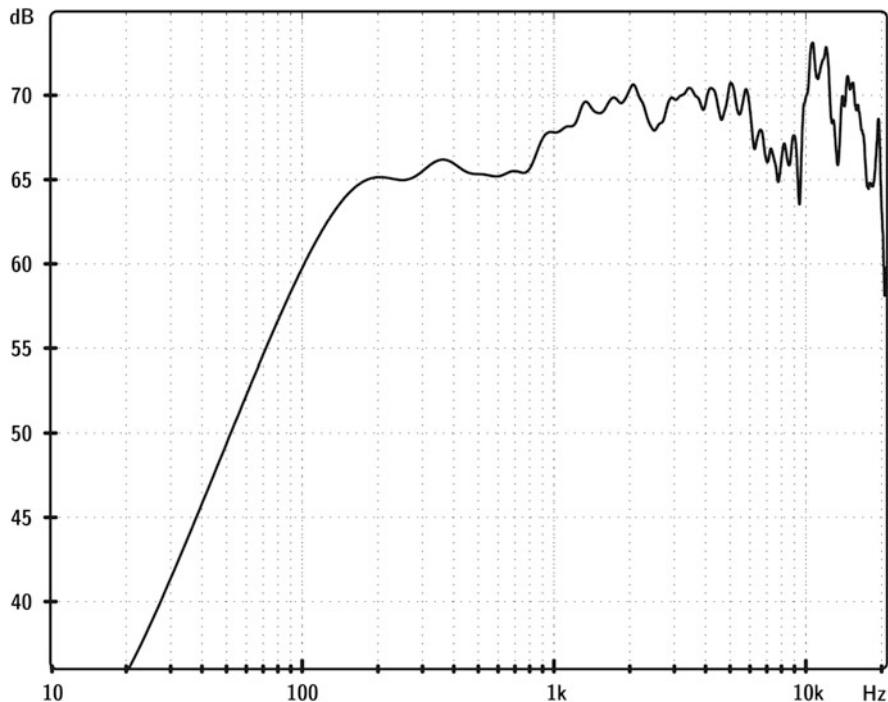
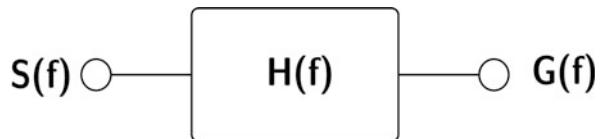


Fig. 7.5 Example of a loudspeaker sensitivity transfer function, $\underline{H}(f)$

Fig. 7.6 Processing of source signal $\underline{S}(f)$ with a stationary transfer function $\underline{H}(f)$ to obtain a receiver signal $\underline{G}(f)$



Accordingly the signal flow expressed in frequency domain reads

$$\underline{G}(f) = \underline{S}(f) \cdot \underline{H}(f). \quad (7.12)$$

As will be explained in the next section, this equation must be interpreted as equivalent to Eq. (7.3) (Fig. 7.6).

7.4 Fourier Transformation

The impulse response of a system and its steady-state transfer function are linked by Fourier transformation one to one:

$$\mathcal{F}\{h(t)\} = \underline{H}(f) \quad (7.13)$$

Thus, the LTI system can both be described in the time or frequency domain. The signal flow through LTI systems can, therefore, be studied in time domain and frequency domain, and all results can be related to the corresponding function in the other domain, too.

The Fourier transformation is the fundamental algorithm to change the interpretation of signal flow from time signals to spectra and vice versa. As illustrated in Fig. 7.7, the Fourier transformation can be applied at any stage of signal transmission. Even the temporal calculation process, the convolution integral, can be “transformed” into the frequency domain, thus giving a multiplication. This is not surprising since the Fourier transformation is known in mathematics as the key to solve integrals of convolution type.

The calculation rule of Fourier transformation for converting between impulse response and steady-state transfer function is

$$\underline{H}(f) = \int_{-\infty}^{\infty} h(t) \cdot e^{-j2\pi ft} dt, \quad (7.14)$$

$$h(t) = \int_{-\infty}^{\infty} \underline{H}(f) \cdot e^{j2\pi ft} df. \quad (7.15)$$

In the case of transforming signals, it reads

$$\underline{S}(f) = \int_{-\infty}^{\infty} s(t) \cdot e^{-j2\pi ft} dt, \quad (7.16)$$

$$s(t) = \int_{-\infty}^{\infty} \underline{S}(f) \cdot e^{j2\pi ft} df. \quad (7.17)$$

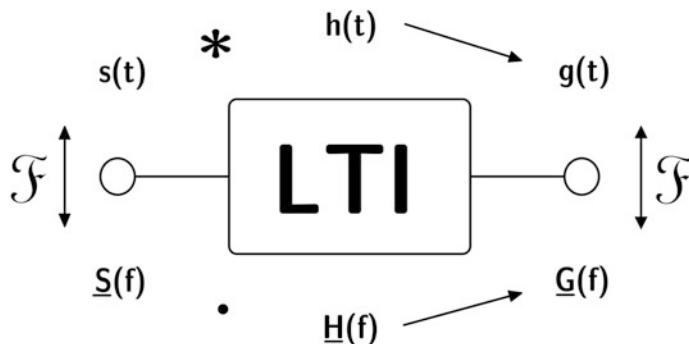


Fig. 7.7 Input and output signals of LTI systems

The Dirac pulse is to be mentioned again. It is the identity function of convolution, since its spectrum is 1, the neutral element of multiplication. The last equation can also be interpreted as a definition of the Dirac pulse.

$$\int_{-\infty}^{\infty} \delta(t) \cdot e^{-j2\pi ft} dt = 1 \quad (7.18)$$

$$\delta(t) = \int_{-\infty}^{\infty} e^{j2\pi ft} df \quad (7.19)$$

So far the fundamentals of signal processing related to acoustic systems have been introduced. For a deeper understanding, however, these basics must be adapted to processing in digital computers. The most important aspect is thus the consideration of discrete signal processing and the proper representation of continuous functions by sampling.

7.5 Analogue-to-Digital Conversion

In order to feed signals into a computer memory and to process it, the analogue signals must be digitized. By using an A/D converter, the analogue time functions $s(t)$ are quantized according to their amplitude (in the end represented by an electric voltage) in several steps and sampled in time, thus yielding a discrete series of scaled binary data.

The precision of quantization depends on the chosen amplitude resolution. The range of numbers that is used is normalized and transformed into an appropriate binary format. The full amplitude scale of the A/D converter is then related to n bits, allowing the analogue signal to be expressed in $2^n/2$ different values between zero and \pm full scale (assuming AC signals with an average close to zero). With a resolution of 16 bit, this is related to 65,536 integer numbers between $-32,768$ and $+32,767$, mapped to a voltage between $-U_{\max}$ und $+U_{\max}$.

Considering arbitrary signals, the approximation uncertainties caused by quantization are distributed stochastically. Since the smallest voltage step is $U_{\max}/2^n$, the level of the expected (rms) quantization noise is given by

$$N_{\text{quant}} = -20 \log 2^n \approx -6n \quad . \quad (7.20)$$

Typical for sound in the hearing range are sampling rates of 40–50 kHz and quantization of 16 bit, in measurement or sound recording hardware also up to 24 bit. Dynamic ranges caused by hardware limitations are thus available that have same range as the best transducers, condenser microphones, with about 130 dB between full scale and quantization noise.

The clock frequency of sampling (sampling frequency) depends on the frequency content of the signal (see below). Taking into account an adequate depth of

discretization, the samples represent an exact image of the analogue signal. In order to modify the signal, however, the discrete form allows much more flexible and elegant solutions of processing (filtering, analysis, amplification, delay, etc.). These modifications can now be carried out as mathematical operations.

According to the theory of linear time-invariant systems, the sampling process can be described as follows: an analogue signal³ $s(t)$ is sampled at times nT with $n = 0, 1, 2, \dots$ and $T = 1/f_{\text{Sample}}$, and instantaneous voltage is measured at each sample. This process corresponds to a multiplication of the analogue signal with a series of Dirac pulses,

$$\text{III}(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT), \quad (7.21)$$

and the sampled signal reads

$$s(n) = s(t) \cdot \text{III}(t)|_{t=nT} = \sum_{n=-\infty}^{\infty} s(nT) \delta(t - nT) \Big|_{t=nT}. \quad (7.22)$$

At least two samples must cover one period of the harmonic signal, as illustrated in Fig. 7.8, in order to exclude any ambiguity, the so-called aliasing. Otherwise, harmonic signals with integer frequency multiples will lead to the same correspondence between the samples and the analogue signal. The complete spectrum is, thus, a series of repeated spectra on the frequency axis.

We can thus identify sampling in time domain (multiplication in the time domain) as the convolution of spectra in the frequency domain. According to one of the main rules of Fourier transformation, we can express

$$s\left(\frac{t}{T}\right) \cdot \text{III}\left(\frac{t}{T}\right) \quad \text{O---●} \quad S(Tf) * \text{III}(Tf) \quad (7.23)$$

with

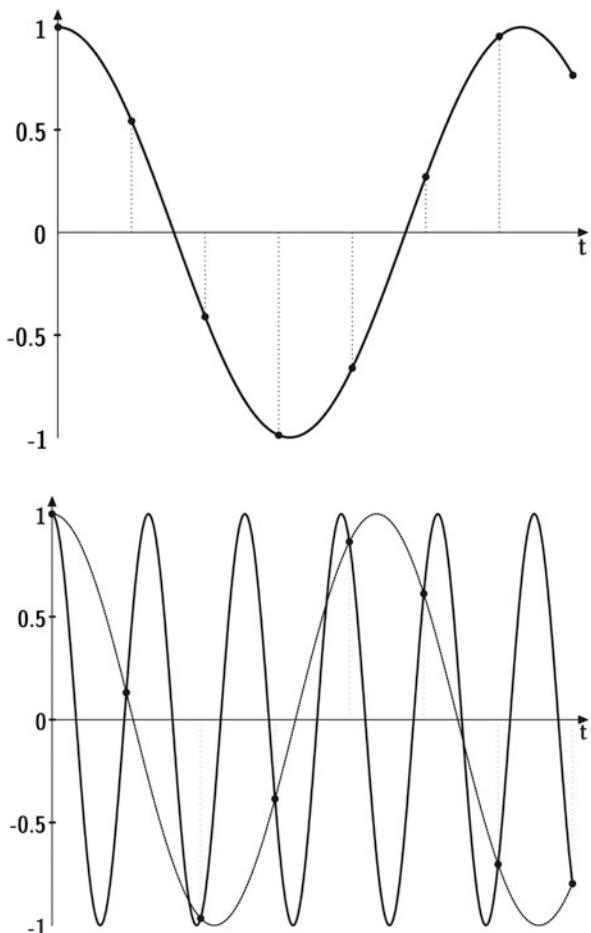
$$\text{III}(Tf) = \int_{-\infty}^{\infty} \text{III}\left(\frac{t}{T}\right) \cdot e^{-j2\pi ft} dt \quad (7.24)$$

denoting the Fourier transform of the Dirac series. Time and frequency axes are normalized to the sampling rate $1/T$. Note the inverse relationship between $\text{III}(t/T)$ and $\text{III}(Tf)$. A narrow Dirac sequence in time domain corresponds to a wide series of spectral lines in frequency domain.

Provided we can cut the original spectrum from the series, the original signal is constructed unambiguously. This can be achieved by applying a low-pass filter (see Fig. 7.9) truncating the spectrum at f_{max} . Accordingly the distance of the centre of the

³To be typically considered preconditioned in Volt units at the input of the A/D device (sound card).

Fig. 7.8 Top: sampling of a signal. Bottom: ambiguity of the discrete samples matching to sinusoidals



alias spectra must be larger than $2 f_{\max}$. This fact is expressed in the sampling theorem.

$$f_{\text{sample}} \geq 2f_{\max} \quad (7.25)$$

7.6 Discrete Fourier Transformation (DFT)

In case of sampled signals, the question of an efficient Fourier transformation remains. The calculation algorithm for the discrete Fourier transformation is (compare Eq. (7.16))

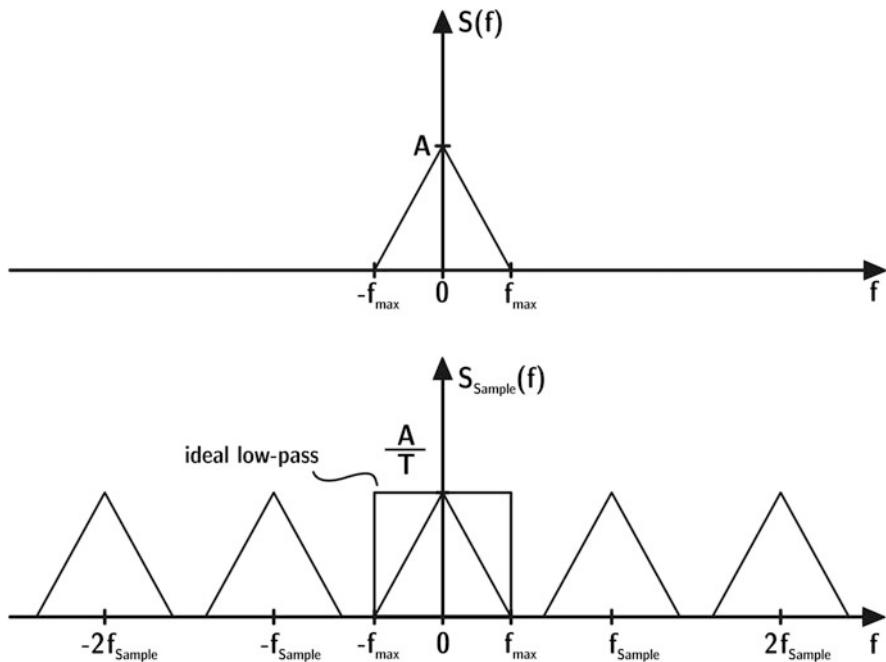


Fig. 7.9 Reconstruction of the analogue signal

$$\underline{S}(k) = \frac{1}{N} \sum_{n=0}^{N-1} s(n) e^{-\frac{j2\pi nk}{N}}; \quad k = 0, 1, \dots, N-1 \quad (7.26)$$

The variable n represents the time domain and k the frequency domain. To solve the sum, N^2 (complex) multiplications are required.

Due to sampling, the spectrum of a sampled signal will be periodic (Eq. (7.23)) and continuous. But in digital representation, the spectrum can only be stored in digital form at certain frequency lines.⁴ The discrete spectrum is, thus, a line spectrum. A line spectrum such as this, however, is strictly related to periodic time signals, even when the original signal is not periodic. Apparently there is a contradiction between (analytic) Fourier transformation and discrete Fourier transformation, DFT. But this conflict can be solved in the same way as spectral aliasing was solved, by ensuring a sufficient distance between the temporal periods (see Fig. 7.10).

⁴We cannot store continuous data in the computer memory.

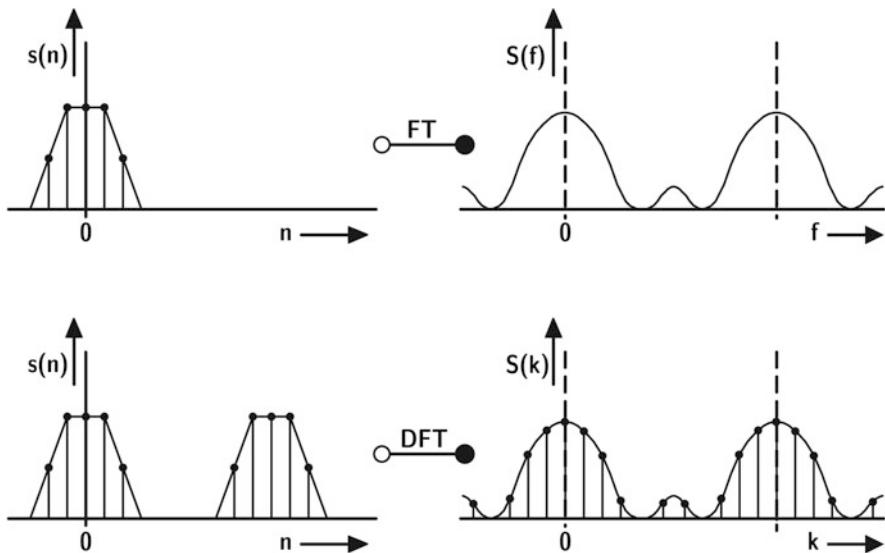


Fig. 7.10 Sampling and processing of a signal, $s(n)$ in top left. Top right: corresponding theoretically continuous spectrum, $S(f)$; bottom right, numerical (discrete) spectrum $S(k)$; bottom left, the periodic signal corresponding to the discrete spectrum $s(n)$ (after Lüke 1999)

7.7 Fast Fourier Transformation (FFT)

A special version of DFT is the so-called fast Fourier transformation, FFT. It is one of the key algorithms in virtual acoustics, in acoustic measurements, in speech and image processing and in other fields. It is not an approximation, but a numerically exact solution of Eq. (7.26). However, it can be applied only in block lengths of

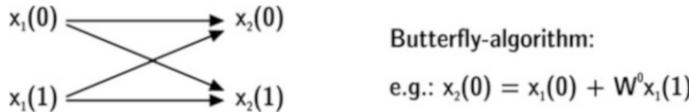
$$N = 2^m, \quad (4, 8, 16, 32, 64, \dots) \quad (7.27)$$

The reason for the accelerated calculation is preprocessing with the result of pre-sorting symmetrical terms and reduction of the necessary processing steps to a small fraction.

The algorithm expressed in Eq. (7.26) is arranged in a linear equation system in matrix formulation, here illustrated in an example with $N = 4$.

$$\begin{pmatrix} S(0) \\ S(1) \\ S(2) \\ S(3) \end{pmatrix} = \begin{pmatrix} W^0 & W^0 & W^0 & W^0 \\ W^0 & W^1 & W^2 & W^3 \\ W^0 & W^2 & W^4 & W^6 \\ W^0 & W^3 & W^6 & W^9 \end{pmatrix} \begin{pmatrix} s(0) \\ s(1) \\ s(2) \\ s(3) \end{pmatrix}, \quad (7.28)$$

with

**Fig. 7.11** FFT butterfly

$$W = e^{-j2\pi/N} . \quad (7.29)$$

Note the high symmetry in the complex phase function, W , which divides the complex plane into N segments. W raised to the power of n corresponds to a rotation and imaging of W into itself, if $2\pi/N$ produces circular symmetry of a half, quarter, eighth, etc. The core of FFT is thus the transformation of the matrix into a matrix of symmetry. This is achieved by a so-called bit reversal, a specific interchange of columns and rows, so that quadratic blocks of zeros ($2 \times 2, 4 \times 4, 8 \times 8, \dots$) are created. Of course, all multiplication terms involving zeros can be omitted. In our example the transformed matrix is

$$\begin{pmatrix} x_2(0) \\ x_2(1) \\ x_2(2) \\ x_2(3) \end{pmatrix} = \begin{pmatrix} 1 & W^0 & 0 & 0 \\ 1 & W^2 & 0 & 0 \\ 0 & 0 & 1 & W^1 \\ 0 & 0 & 1 & W^3 \end{pmatrix} \begin{pmatrix} x_1(0) \\ x_1(1) \\ x_1(2) \\ x_1(3) \end{pmatrix}, \quad (7.30)$$

with x_1 and x_2 denoting the temporal and spectral vectors, respectively, after matrix conversion. For instance, the calculation of two vector elements of x_2 reduces to $x_2(0) = x_1(0) + W^0 x_1(1)$ and $x_2(1) = x_1(0) + W^2 x_1(1)$. All other product terms are zero.

It is worth mentioning that the remaining terms create links between neighboured vector elements to two others. This fact can be used to express the process in a butterfly algorithm (Fig. 7.11).

The solution of an $m \times m$ matrix can finally be found by a cascade of m butterflies, which reduces the required number of multiplications from N^2 to $N \log_2(N/2)$, for example, for $N = 4096$ by a factor of 372 from 16,777,216 down to 45,056.

7.7.1 Sources of Errors, Leakage and Time Windows

At the given boundary conditions, several sources of errors are possible. At first, we have to keep in mind that the FFT as a special form of the DFT is related to periodic signals. If the signal to be transformed is indeed periodic, the block length (time frame) of the DFT or FFT must correspond to an even number of periods, so that the

continuation at the end is exactly the same as at the beginning of the block. Otherwise the forced periodicity of the DFT creates a discontinuity, and the Fourier transform is related to this discontinuous signal.

If the DFT or FFT block length exactly matches an even number of periods, this error is avoided. This can be accomplished by manual or automatic period identification and sampling rate conversion (Fig. 7.12).

Another approximate method is window technique. A window is applied by multiplication of a window function to the time frame. The window function acts like a pass filter, however, in time domain. A symmetric window reduces early and late components in the signal and lets the midtime part pass unchanged. Basically the window reduces the leakage effect by reducing the relative amplitude of the discontinuity. Windowing corresponds to a convolution of the signal spectrum with the window spectrum. Windows can therefore be optimized, based on temporal and spectral features (Table 7.1).

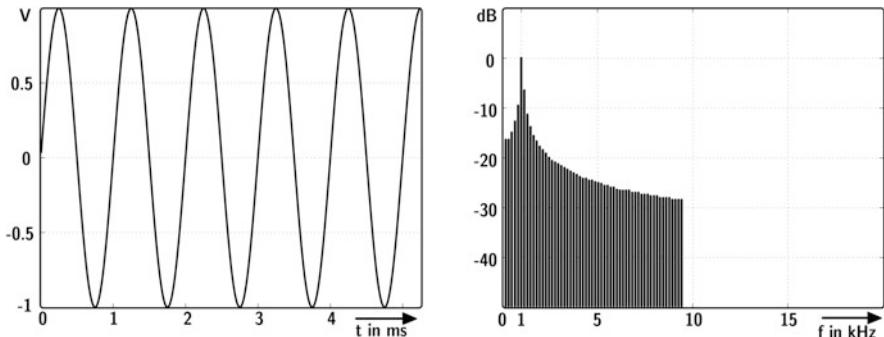


Fig. 7.12 1 kHz pure-tone signal and spectrum by using DFT (5.25 periods)

Table 7.1 Typical window functions (e.g. $n = 0, 1, 2, \dots, N - 1$)

Window	Function
Rectangular	
Triangular	$w(n) = \begin{cases} \frac{n}{N/2}; & n = 0, 1, \dots, N/2 \\ \frac{N-n}{N/2}; & n = \frac{N}{2}, \dots, N-1 \end{cases}$
Hanning	$w(n) = \sin^2\left(\frac{n}{N}\pi\right)$
Hamming	$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi}{N}n\right)$
Blackman-Harris	$w(n) = a_0 - a_1 \cos\left(\frac{2\pi}{N}n\right) + a_2 \cos\left(\frac{2\pi}{N}2n\right) - a_3 \cos\left(\frac{2\pi}{N}3n\right)$

7.8 Digital Filters

Digital filters are used for pre- and post-processing of signals. In measurements they serve as high-pass, low-pass or bandpass filters. In auralization and sound reproduction, they serve as a basis for filtering, convolution and final adjustment of audio effects including special cues like spatial attributes or equalizing sound reproduction equipment.

Digital filters are designed from combinations of addition, multiplication and delay components. Creating delay was always the biggest problem with analogue techniques. With digital tools, delay elements in particular are created much more easily (just using storage devices).

A discussion of digital filters is best illustrated with a plot of the complex transfer function, the modulus and phase response. Furthermore, in the pole-zero diagram, the order of the filter can be discussed. Figure 7.13 shows an example for a filter in both diagrams.

For a theoretical description of digital filters, the Hilbert transformation is applied. It is a general form of Fourier transformation that is also used for treatment of harmonic functions. By introducing the Laplace variable

$$z = e^{j\omega} \quad (7.31)$$

the frequency response is mapped to a complex two-dimensional function in the complex plane. The following rules can be applied to design filters: Poles and zeros must be either real, or they must appear as complex conjugates. For example, a pole at $z = 0$ leads to multiplication of the frequency response with $e^{-j\omega t}$, thus affecting the phase without changing amplitudes. A pole (or zero) on the unit circle corresponding to a filter response $H(j\omega)$ becomes infinite (or zero) at a certain

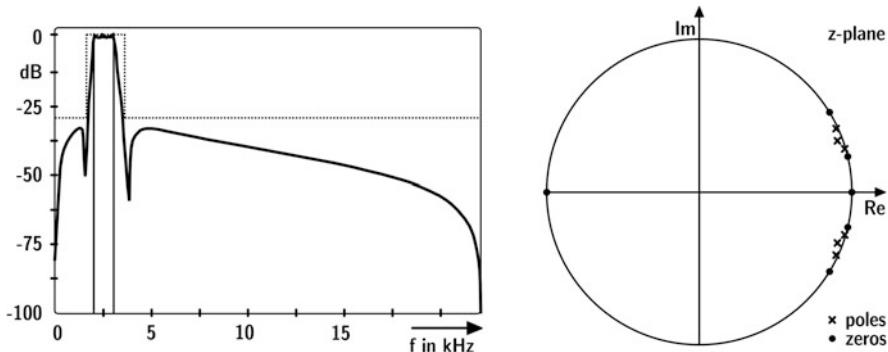


Fig. 7.13 Digital bandpass filter of sixth order; top, frequency response, and bottom, pole-zero diagram

frequency. A pole outside the unit circle creates an instability with increasing filter impulse response $h(t)$. Poles outside the real axis generally correspond to oscillations of the filter impulse response.

The frequency response can now be constructed from the pole-zero plot easily. The z plane is considered to represent a membrane. Poles are marked by vertical columns below the membrane, zeros by heavy stones put on the membrane. From the resulting landscape on the membrane, with hills and valleys, the modulus frequency response of the filter is the height along the unit circle, starting from 1 on the real axis. Digital filters can be divided into two groups: IIR and FIR filters.

7.8.1 IIR Filters (Infinite Impulse Response)

IIR filters make approximation of desired impulse response functions possible. Poles ($a(n)$) and zeros ($b(n)$) are placed in the complex plane. The filter transfer function is then

$$H(z) = \frac{\sum_{n=0}^N b(n)z^{-n}}{\sum_{n=0}^N a(n)z^{-n}}, \quad (7.32)$$

which should approximate the desired response with least possible order N . This can also be illustrated by using a block diagram with forward and feedback lines. z^{-1} means a shift by one sample, and the triangles mean multiplication (amplification) with factor a or b (Fig. 7.14).

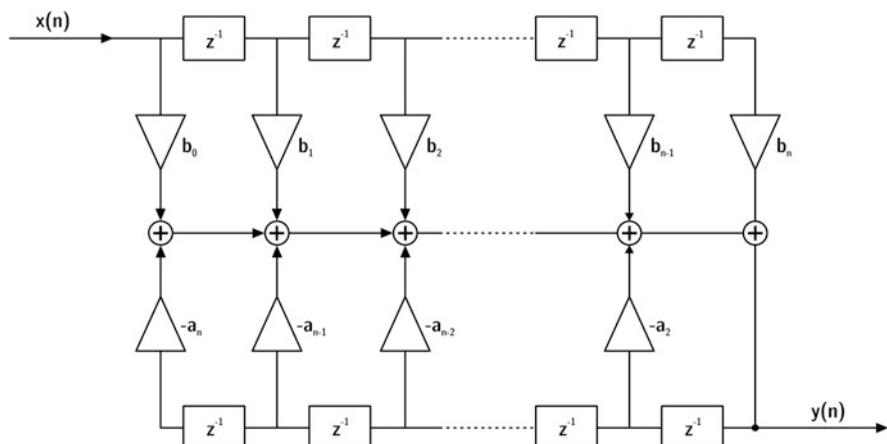


Fig. 7.14 Block diagram of an IIR filter; $x(n)$ and $y(n)$ are input and output signals, respectively

The output signal, $y(n)$, is created from amplifying and adding past samples. Due to the feedback loop, the filter impulse response can be infinitely long (infinite impulse response).

IIR filters can be optimized to produce a specific modulus response, although the phase response cannot be controlled independently. Due to feedback conditions, they may be unstable, unlike FIR filters. Furthermore, IIR filters require less effort and complexity than FIR filters and usually have a lower order.

7.8.2 FIR Filters (Finite Impulse Response)

FIR filters are created by approximating the desired function by placing zeros on the unit circle and choosing poles exclusively in the origin with

$$\sum_{n=0}^N a(n)z^{-n} = 1 \quad (7.33)$$

Thus the transfer function of the filter reads

$$H(z) = \sum_{n=0}^N b(n)z^{-n} \quad (7.34)$$

with the following block diagram (Fig. 7.15).

FIR filters are stable in each case. The output only depends on input data and not on feedback. The impulse response is identical with the coefficients, $b(n)$, and it is finite in length (finite impulse response).

In FIR filters the modulus and the phase can be controlled independently. For control at low frequencies, however, the filter order (length) must be quite high since one period of the corresponding spectral content must fit within the filter length.

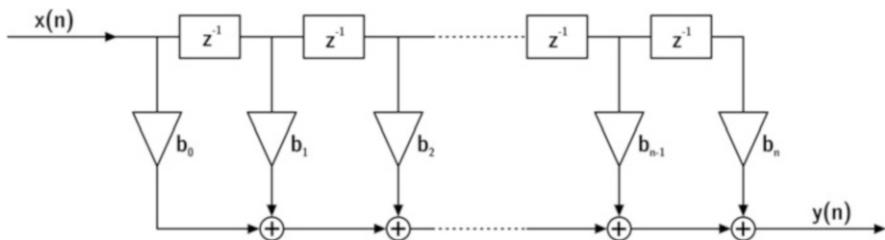


Fig. 7.15 Block diagram of an FIR filter

Filter concepts are useful and applicable for auralization in various kinds. There is no absolute preference for the one or the other approach. The optimum filter depends on the application and the software implementation. For more information see Papoulis (1981), Mourjopoulos (1994) and Kirkeby and Nelson (1999).

Chapter 8

Characterization of Sources



8.1 Introduction

Sound sources are the starting point when it comes to auralization. Their characterization includes the radiation properties such as power and directivity. And the sound signal, the temporal function of the generated sound, is the main input signal for auralization technology. Physical modelling, recording, and sound synthesis can be used for obtaining signals. Databases of frequency-dependent directivities, possibly more aspects with regard to the source mechanisms, their excitation levels and their near-field effects, may be relevant in practical cases, too. In this chapter we consider how sound generation and sound sources can be recorded, modelled and processed and how the main features of sound sources can be implemented in virtual acoustics.

The so-called “dry” source signal is generally defined as a source signal free of reverberation and of any other cues introduced by sound transmission. As soon as the relevant transfer functions are known by simulation or measurement, and the input signal is recorded (or simulated) properly, the signal transmission path is identified, and the output signal can be obtained (Fig. 8.1).

8.2 Airborne Sound Sources

Input signals for auralization, $s(t)$, are related to specific sources. Typical examples are musical instruments, the human voice or a noise-generating machine. At this point the question of the interface between the source and sound transmission system and corresponding feedback shall be briefly discussed, or rather the neglection of feedback will be justified. The radiated sound power of an instrument or voice should be independent of the environment. In airborne auralization problems, this prerequisite is mostly fulfilled easily, whereas in structure-borne sound auralization,

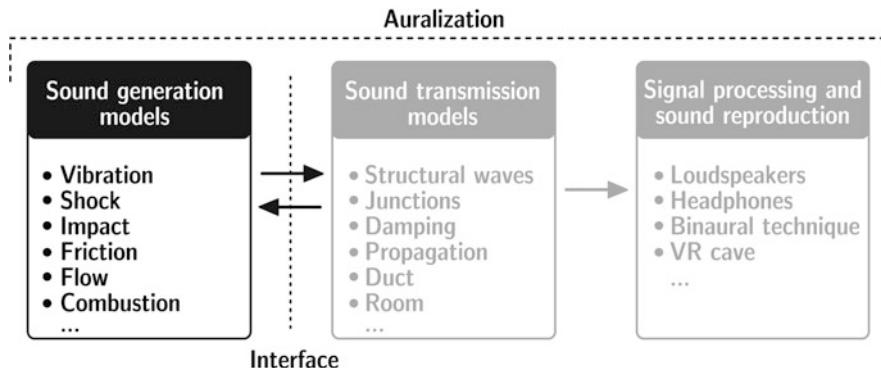


Fig. 8.1 Sound source characterization

feedback might occur as a severe problem. The reason is the relation between the inner impedance of the source and the radiation impedance (Sect. 2.4). As long as the radiation impedance is small compared with the inner impedance of the source mechanism, the radiated volume velocity is invariant to changes in the environment (room, duct, etc.).

8.2.1 Recording Sound

The straightforward approach is a recording of the sound source at a given distance and in a given direction in a free-field outdoors or in an anechoic chamber. The anechoic chamber must fulfil requirements similar to those for sound level measurements. Design standards prescribe the maximum wall reflection amplitude permitted for a given deviation from the free-field distance law ($1/r$). The direct sound level compared with the reflection level can then be evaluated in order to check if the comb filter effect and the corresponding change in timbre caused by the reflection could be audible.

Decreasing the distance to the source could help to improve the direct sound level. The spectral balance, however, must be ensured to be identical with the spectrum of the far-field radiation. This is not given in the near field. If near-field recordings are made, they must be calibrated and equalized to the far-field spectrum. Independently, the source directivity pattern must be taken into account; see below.

Source calibration is usually done with reference to the sound power or to the sound pressure level at a certain distance, at 1 m, for instance.

More details are discussed below in Sect. 8.2.4 on the example of recording musical instruments.

8.2.2 Synthesizing Sound

In some cases the sound sources are too complex to be placed in an anechoic environment. This applies to large engines for railbound vehicles or to aircrafts, for instance. The sound sources may also consist of many sound-radiating components such as combustion engine noise and intake flow noise. The problem is thus to separate the sound sources into their construction components and to use them for auralization purposes independently.

The sound radiated from those large distributed sources or the source components may be known from sound power measurements or predictions. If sound source recording is not feasible, a reconstruction technique based on measurement data is required. Measurement data can be (a) band-filtered sound pressure levels or sound power levels, (b) amplitudes or series of amplitudes of pure tones, (c) tone complexes with added noise floor or (d) any other signal representation without having the full complex spectrum or time signal available (Pieren 2018).

The basic idea consists of adding tonal and noise components. Let us discuss an example of noise inside trains, trams and subways, possibly also applicable to electric cars. Traction noise from the electric engine and the gearbox is highly tonal, and it is quite easy to calculate the main frequencies and their amplitudes from model parameters such as pole configuration in the electric engine and the gear teeth ratios. Rolling and ventilation of wind noise can be added by using filtered white noise. This concept is illustrated in Klemenz (2005) (Fig. 8.2).

In a more general approach, the sound source is separated into the main principles of impact, friction or harmonic vibration as primary physical process and a subsequent filter, which forms the sound spectrum due to elements such as tubes,

VEHICLE NAME		
1	Subway_X	
Rolling-Noise File		
2	Roll_X.wav	
3	Ratio	0.5
BIR File		
4	BIR_1.wav	
Calculation interval		
5	Start [s]	0.0
6	End [s]	10.0
Global		
7	Acceleration [m/s^2]	0.90
8	Pole-pair number	2
9	Wheel perimeter [m]	2.5
10	No. of harmonics	10
11	Motor level [dB]	70
Phase 1 - PWM		
12	End of phase a [s]	2.0
13	End of phase b [s]	6.0
14	f_s phase a [Hz]	400
15	f_s phase b [Hz]	600
16	Sweep for a (yes/no)	yes
17	Exponent harmonics	1.5
18	Ratio sidebands	0.05
19	Ratio AM	0.0
20	AM frequency [Hz]	0
21	Random (yes/no)	no
22	Noise bandwidth [Hz]	0
Gear coupling		
30	No. of sections	2
31	Exponent harmonics	1.3
32	Sec 1 Z small	28
33	Sec 1 Z big	67
34	Ratio for Sec 1	0.5
35	Sec 2 Z small	23
36	Sec 2 Z big	57
37	Ratio for Sec 2	0.3
Transfer function		
38	Lowpass cutoff [Hz]	2000
39	f [Hz]	100;200;400;800
40	Q	15;10;10;10
41	L [dB]	5;10;5;0
Output file		
42	Subway_X_1.wav	

Fig. 8.2 Examples of a synthesizer for auralization of electric vehicles

waveguides or resonators or any other modal system. This technique is very successfully developed and used in musical instrument modelling (Fletcher and Rossing 1998), but it can well be extended for noise sources, too, if the equations of the primary force excitation can be formulated and solved in the specific circumstances.

Coming back to the example of electric drives, this is illustrated in a modelling process for the electromagnetically excited noise of electrical machines. The synthesis process can be applied to several machine types and geometric configurations, including outer-rotor and transversal-flux machines as well as machines with rotor or stator skew. Electromagnetic air-gap force models for permanent magnet synchronous machines and for switched reluctance machines can be used to predict the forces and resulting vibrations of the rigid body of the machine components. Also switching frequencies and machine harmonics can be taken into account (van der Giet 2011; Bösing 2013). Sound radiation or transfer path analyses can later be added as well as sound radiation with directivity.

One of the most advanced general sound synthesis model for car and trains was presented by Pieren (2018). This group also studied synthesis of wind turbine noise. Calculation models for sound synthesis were developed from input parameters from measurements and data analysis. For wind turbine and road traffic noise, for example, a combination of additive and subtractive synthesis, denoted as spectral modelling synthesis, ensures that complex source superposition leads to realistic impressions at the sound emission point. This also includes a wind turbine synthesizer with different types of characteristic amplitude modulations. The synthesizer for road vehicles covers the specific complex spectral and directional characteristics of tire noise and engine noise. Tire noise models are based on interaction forces between the road surface and the tire profile. Complex models take into account the microtexture of the tire/road contact and the modal response of the tire material (rubber), shape and profile (Kropp et al. 2012).

For railway noise, a physically based synthesis approach has been developed that describes the mechanical excitation and the vibration of the dynamic wheel/rail system, depending on the microtexture of the wheel/rail contact as well as on structural resonances of the wheel/rail system; see Fig. 8.3.

8.2.3 *Directivity*

Modelling and auralization of a musical instrument, the human voice or loudspeaker is usually based on the sound signal recorded at a specific point (in the direction of “main” radiation or in a symmetry axis). The directional radiation pattern is accounted for by using a directivity database. For musical instruments, the human voice or loudspeakers, such data can be found indeed. For noise sources, however, such generally applicable directivity data are typically not available. The directivity must be modelled by assuming specific elementary radiation conditions, or it must be measured in anechoic situation or in the near field. Microphone arrays on a measurement surface such as arrays used in sound power measurements are an

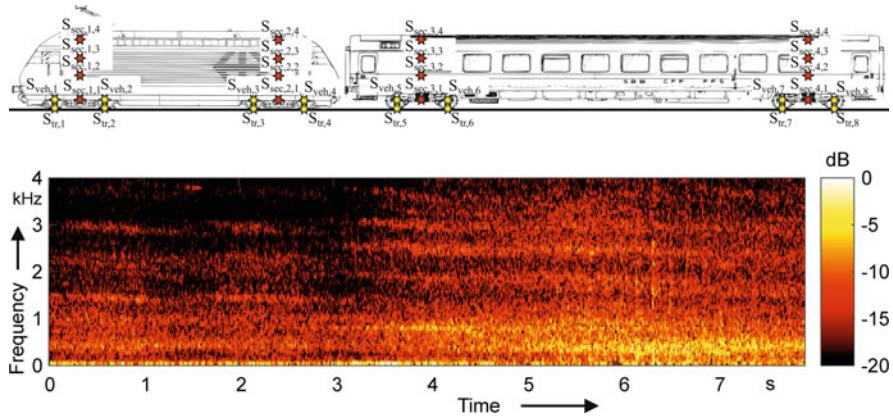


Fig. 8.3 Equivalent point source locations marked by stars along two Swiss rail vehicles of type SBB Re 460 and Bpm RIC (top), spectrogram of the sound pressure of a train pass-by (normalized to 0 dB), train speed 60 km/h (bottom) (Pieren et al. 2017)

option for collecting data for source directivity encoding. In this respect it is also of interest to reconstruct, at least in approximation, the surface vibration of the noise source. Methods of multipole analysis (Ochmann 1990) and acoustic holography (Williams 1999) may yield equivalent source parameters which represent the vibration pattern and the far-field radiation directivity.

Source encoding and reconstruction on the basis of several mathematical models is described in detail in Magalhães and Tenenbaum (2004). For spherical geometry or if approximations of point-like sources are possible, the concept of multipoles passes into the expansion of spherical harmonics. In this case the multipole sources coincide in the middle of a sphere surrounding the source. Whether multipole synthesis or spherical harmonics describe the actual source of interest better cannot be decided generally. In the next section, we take recording of musical instruments as an example for application of spherical harmonics decomposition.

8.2.4 Musical Instruments

Signals from musical instruments can be recorded in anechoic environment (anechoic chamber or at least a highly absorbing room like a recording studio) (Giron 1996). Recording must be made in the far field, and sufficiently many microphone positions must be used to cover the directional characteristics properly. It must further be ensured that the directional characteristics are constant (time invariant) for all signal components.

The prerequisite of time invariance of the directional characteristics seems to be no problem for brass instruments since the radiation is dominated from the horn opening (which remains unchanged while playing). In contrast, woodwind

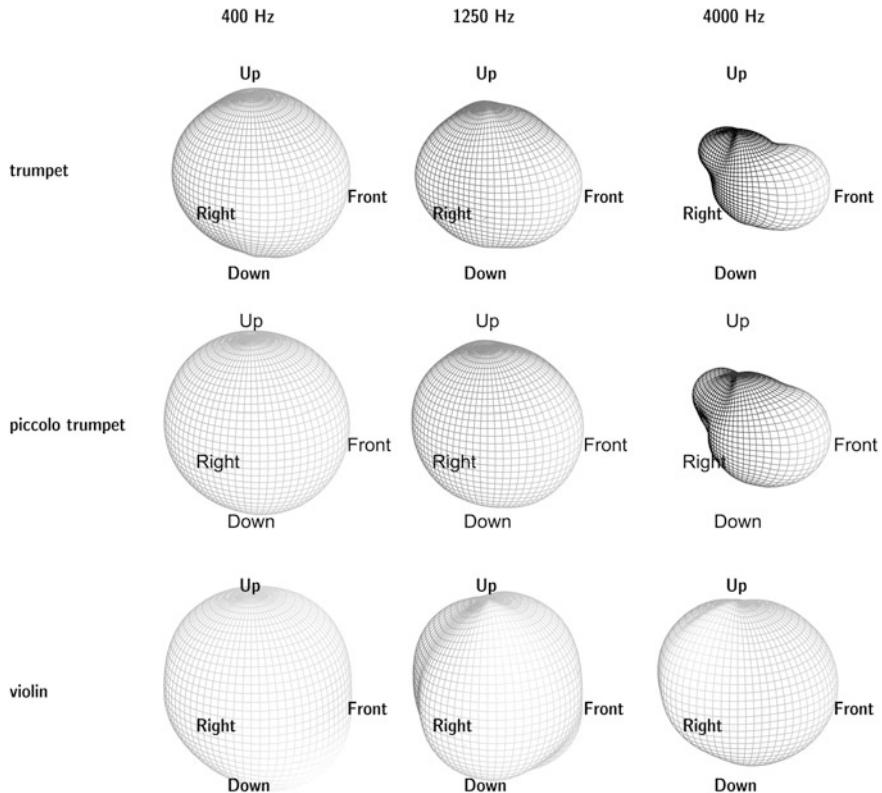


Fig. 8.4 Examples of directional characteristics (“balloons”) of musical instruments from the open-access database (Shabtai et al. 2017)

instruments have a fluctuating radiation pattern as the valves are opened and closed while playing the instrument.¹ Thus, we have to face the problem that not just one directional pattern is valid but the pattern depends on the signal frequencies (tones played). Multichannel recording and separate simulation of each channel is one way to overcome this problem (Otundo and Rindel 2005).

Another aspect is the floor reflection. In hemi-anechoic rooms the floor reflection will be included in the recording, like in the actual performance on a stage. In this case, the floor reflection must not be included in the simulation as well.

A comprehensive collection of directional characteristics of musical instruments was first published by Meyer (1995). New research on characterization of directional radiation from musical instruments became possible recently with the availability of inexpensive multichannel measurement technology. A large measurement campaign was launched in 2012 which extended the database to 41 musical instruments (Fig. 8.4). The directivity and the sound power were analysed and made available

¹Note that the valves and holes form a line array of volume sources.

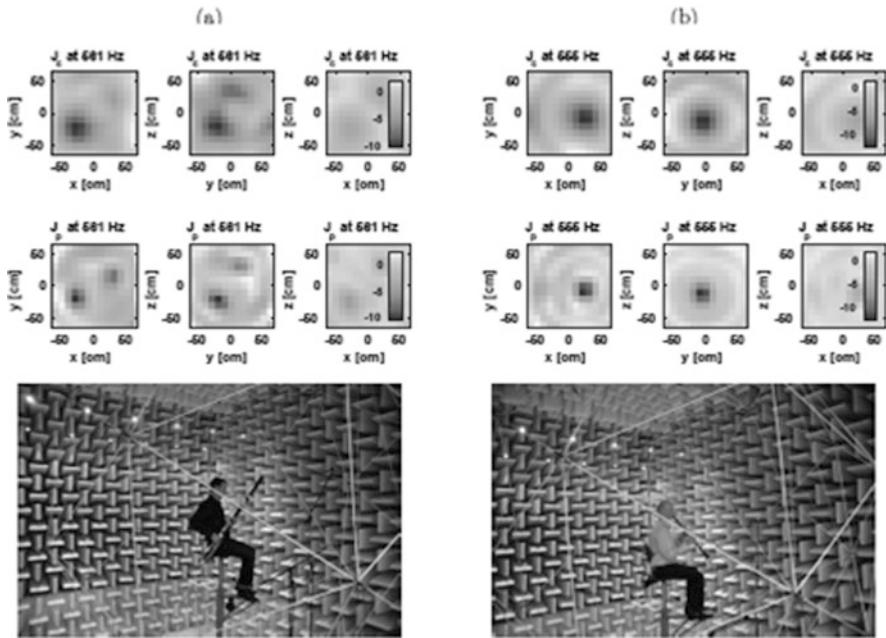


Fig. 8.5 Recording setup (bottom) and acoustic centring results (top) for bassoon (left) and clarinet (right), after Shabtai and Vorländer (2015)

for open access (Shabtai et al. 2017). In the measurement session, each instrument was played over the entire chromatic tone range in pianissimo and fortissimo in an anechoic chamber using a surrounding spherical microphone array (Behler et al. 2012). In post-processing, acoustic source centring according to Shabtai and Vorländer (2015) was applied in order to align the acoustic source centre with the centre of the recording sphere (Fig. 8.5). The acoustic radiation pattern is then generated in the spherical harmonics domain at each harmonic partial of each played tone. The directional data can be used for the implementation of radiation patterns in auralization.

8.2.5 Singing Voice

In the case of the human voice for talking or singing, the radiation pattern is not constant either since the mouth opening depends on the text spoken or sung. A method for recording of singers' directivities was developed by Kob (2002) who aimed at an artificial singer representing a human singer. For this purpose singers were recorded in anechoic environment with a two-channel technique. One channel was recording the signal sung (glissando over at least one octave) in the far field in 2 m distance, while the other channel was serving as reference near to the mouth.

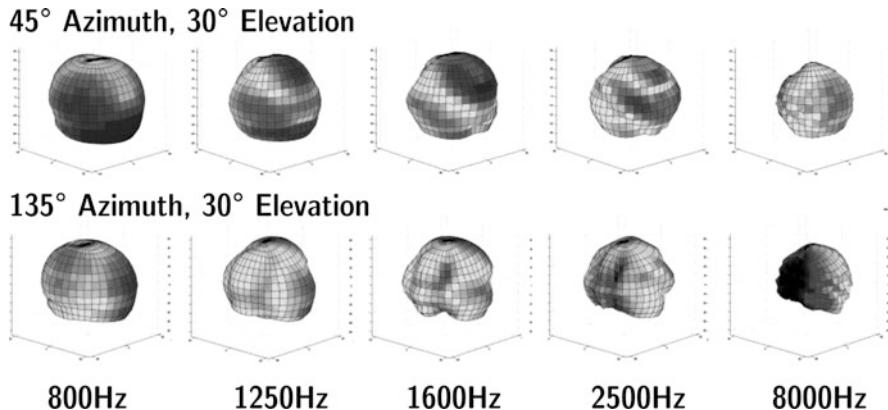


Fig. 8.6 Directivity of the singing voice at two view angles (after (Kob 2002))

With proper equalization and normalization to the frontal incidence, the directional pattern is obtained after placing the singer on a large turntable and repeating the procedure in angular steps (Fig. 8.6).

8.2.6 Speaking Voice

Speech sources have to be separated from singing voice sources since the mouth opening is different. For speech auralization (and speech excitation in measurements), data from artificial head can be used. These data are found in telecommunication standards (ITUp.58), and artificial “talking” heads are available (Fig. 8.7).

8.2.7 Anechoic Recordings

Single Instruments

Anechoic recordings as described above are made by using a set of microphones around the source. The reference condition is typically the frontal direction. More or less radiation in the specific directions is taken into account by the source directivity. This approach is appropriate when the source radiation pattern is independent of the signal (music) played.

Recording and directivity measurement can also be combined in one session. Here the player with the instrument is placed in a spherical microphone array (Pätynen and Lokki 2010; Zotter 2009, Behler et al. 2012). The multichannel recording is analysed, calibrated and rearranged in post-processing. From these

Fig. 8.7 Artificial head for the singing voice (after (Behler and Müller 2000; Kob 2002))



data the sound power, the directional pattern and the time signal with reference to the main direction (usually the front) are obtained.

For some instruments, however, like woodwind and string instruments, the directivity changes dynamically, which in fact created the necessity for multichannel recording of single instruments (Rindel et al. 2004). These recorded sounds contain the source signal as well as the directivity pattern, provided the channels are calibrated. The difference to the spherical recordings described above is that in the reproduction situation, the channels are treated separately and independently for the specific direction (Fig. 8.8). This is easy in case of room simulation (Chap. 11) where the simulated source can be driven in independent angular segments and the channel-specific results superposed at the receiver point.

For a general source characterization applicable for loudspeaker reproduction, the situation is more difficult. The channels now must be mapped to a kind of omnidirectional loudspeaker with adjustable directional characteristic. If the characteristic should not be constant but depend on the music played, like in the example of woodwind instruments, the directional pattern must be controlled by an adaptive process. This process requires a set of parameters. One option in this respect is a set of multipole coefficients or spherical harmonics.

Orchestra Recordings

At present, some recordings of anechoic music and speech are available (Denon 1995; Hansen and Munch 1991; Freiheit 2005; Vigeant et al. 2007). These

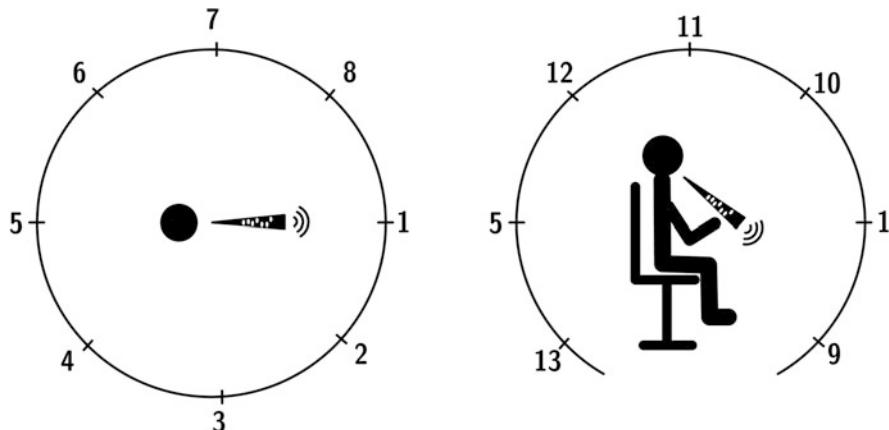


Fig. 8.8 Multichannel source recording (After Otondo and Rindel 2005)

Table 8.1 Denon “Orchestral Music Recording”, Osaka Symphonic Orchestra (Denon 1995), 1987

Excerpts from	
Mozart	Overture Le nozze di Figaro
Mendelssohn	Symphony no. 3 “Scottish”
Bizet	Menuet L’Arlésienne
J. and J. Strauss	Pizzicato-Polka
Glinka	Overture Ruslan and Lyudmila
Verdi	Prelude La Traviata
Bruckner	Symphony no. 4 “Romantic”
Debussy	Prélude à l’Après-Midi d’un Faune

recordings were made with orchestras and choirs in anechoic chambers or in near-field condition with least possible crosstalk. Other signals must be created in each case specifically. The most “clean” solution (but the most tedious one) is surely to record an ensemble by the instruments one by one while replaying the other voices by insert earphone (Vigeant et al. 2007).

The recordings, mostly available on CD, contain the music material listed in the tables below. The first project to be mentioned here goes back to the 1970s. These first anechoic recordings were used for studies on psychoacoustic evaluation of room acoustic field in several European concert halls (Gottlob 1973; Siebrasse 1973).²

²For these studies, Mozart’s Jupiter Symphony was played by the BBC orchestra in an anechoic chamber. Originally it was not intended for use in computer auralization. The recording was broadcast in European concert halls, recorded binaurally by using dummy heads and then replayed for listening tests on the psychoacoustic dimensions of hearing in rooms.

Table 8.2 B&O “Music for Archimedes” (Hansen and Munch 1991)

Samples	
Speech	English and Danish
Guitar	Tárrega, Bach, Sor, Villa Lobos
Violoncello	Weber, Martini
Drums and percussion	Various
Brass	Haydn, Nielsen, Mason, Purcell

Table 8.3 Wenger Anechoic Choral Recordings, October 2003 (Freiheit 2005)

Song list	
C.V. Stanford	Beati Quorum Via
R. Thompson	Alleluja
J. Ferguson	Who is this
E. Aguiar	Psalm 150
H. Willam	Kyrie from Missa Brevis no. 4

Later, about middle of the 1980s, it was aimed at anechoic recordings for more general use, such as those produced in a concert hall in Osaka, Japan. The concert hall stage was modified to obtain approximate anechoic conditions. Due to relatively near microphone positions to pick up the instruments and instrument groups (clearly within reverberation distance), the direct field could be recorded with only very little influence of the hall’s reverberation at a very low level (Table 8.1).

In the years 1987–1992 in Denmark, the “Archimedes” project was focused on subjective effects of loudspeakers in room acoustics. Listening rooms were also simulated so that studies could be concentrated on the balance between direct loudspeaker sound and room reflections. For this purpose, anechoic recordings were produced (Table 8.2):

Since choral source material was rarely available until the year 2000, Wenger Corporation initiated a project for recording the St. Olaf College Choir in Northfield, Minnesota, in an anechoic chamber (Table 8.3).

Finally, in the study by Vigeant et al. (2007), it was examined which recording technique is superior in realism and source width. Multichannel auralizations and single-channel auralizations for both solo instruments and a full orchestra were compared in listening tests. The recordings made were based on the multichannel technique proposed by Otundo and Rindel (2005) but extended for larger ensembles (each instrument playing one by one). This set of anechoic sound examples representing music played by solo instruments and orchestra is currently the most advanced approach with respect to channel separation between instruments and directions or radiation (Fig. 8.9, Table 8.4).



Fig. 8.9 Multichannel source recording session (after (Vigeant et al. 2007), courtesy of J.H. Rindel, DTU Lyngby)

Table 8.4 Five-channel anechoic recordings of an orchestra of solo instruments using the technique developed in the DoReMi project (Otundo and Rindel 2005), June 2005

First bars (about 1 min 30') from

Mozart	Symphony No. 40 in G minor, 1st movement
Brahms	Symphony No. 4, 3rd movement

8.3 Structure-Borne Sound Sources

Recordings of structure-borne sources must be made in each specific case under specific conditions. The variety of sources is very large. Numerous recording details have to be considered. At present there is no standard set of structure-borne sources known. The variety is given by:

- Dimensionality of motion (velocity, force vectors)
- Dependence of force and velocity output on parameters of the transmitting system
- Point, line or area contact

However, one standard force source is well-known: the tapping machine in building acoustics. The tapping machine should represent the force injected into a floor by a walking person. This force will be transmitted completely into the contact surface, provided the input impedance is very high compared with the impedance of the hammer. This kind of force source is representing the ideal case with contact condition:

- Point force in normal (vertical) direction
- High contact impedance of the transmitting system

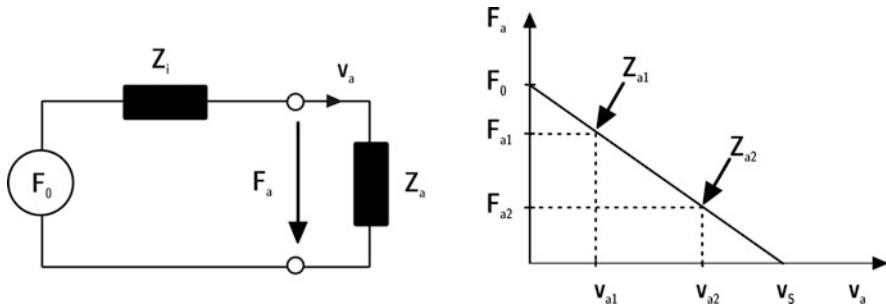


Fig. 8.10 Characterization of real force sources

8.3.1 General Approach

For a more general treatment of sources in combination with structures, the kind of contact and the impedances involved must be considered. In first (one-dimensional) approach, the open-circuit force and inner impedance of the source must be determined. One solution is to substitute this mechanical problem by electrical equivalent. The force source, thus, is given by an ideal source with open-circuit (blocked) force, F_0 , and a serial impedance, Z_i , or mobility Y_i . Real force sources can be characterized in first approximation with this simple model. Any source can be described by an open-circuit force, F_0 ; a short-circuit (free) velocity, v_s ; and a mechanical impedance, Z_i , or mobility Y_i . The crucial point is that the source parameters can be determined rather easy by using different (known) load impedances, Z_{a1} and Z_{a2} , for determination of force and velocity (Fig. 8.10).

For a point contact, the interaction of the source with the structure is then obtained, and the actual force and velocity generated by the source acting on any structure can be calculated. The problem, however, is that point contacts are not a sufficient model in the majority of cases in structure-borne vibration sources; see Sect. 8.3.2.

$$Z_i = \frac{F_{a2} - F_{a1}}{v_{a2} - v_{a1}} \quad (8.1)$$

$$F_0 = F_{a1} - v_{a1}Z_i \text{ or } F_0 = F_{a2} - v_{a2}Z_i \quad (8.2)$$

For discussing the interface problem further, we choose the model of power flow. The power of a source injected into the structure is

$$P = \frac{1}{2} |v_s|^2 \frac{\operatorname{Re}\{Y\}}{|Y_i + Y|^2} \quad (8.3)$$

The two extreme cases are easily defined. As mentioned above and valid for the tapping machine used on heavy constructions holds (force source)

$$P \approx \frac{1}{2} |F_0|^2 \operatorname{Re} \{Y\} \quad (8.4)$$

and for heavy sources working on lightweight constructions

$$P \approx \frac{1}{2} |\mathbf{v}_s|^2 \operatorname{Re} \left\{ \frac{1}{Y^*} \right\}. \quad (8.5)$$

If the source is acting on a plate, their mobility is of interest. Plate mobilities depend strongly on the modal distribution. In case of a very large area, the point mobility of a plate with bending stiffness B and mass surface m'' can be approximated by

$$Y_{\text{plate}} = \frac{1}{8\sqrt{B'm''}}. \quad (8.6)$$

For finite-size plates, the modal pattern must be calculated first (Cremer et al. 1973) to obtain the point mobility (see also Sect. 5.3.1). One important aspect is that the modal pattern is not independent of the point of contact.

8.3.2 3D Force Sources

The principle described in the section above can also be extended to more complex structures. The complexity, however, may become very much greater in case of dynamic interaction of distributed sources or line area contacts. Furthermore, various wave types will interact in a complicated way. Structure-borne energy will be transmitted through various paths and via various contacts. These contacts may involve several degrees of freedom which illustrates the complexity of interaction. For these problems of multiple contact points, the formulation of the total injected power is still similar to Eq. (8.5), but the force/mobility coupling is expressed in a matrix. Petersson and Plunt (1982) and Gibbs et al. (2007) extend this approach towards practical cases in building structures. The main extension is that a multiple contact is considered (see also (Petersson and Gibbs 2000)).

$$P = \frac{1}{2} \left((\mathbf{v}_s)^T [(\mathbf{Y}_i^T)]^{-1} (\mathbf{v}_s^*) \right) \quad (8.7)$$

For multiple-point contact, the relative force amplitude, F_j/F_i , and the coupling transfer mobility, \mathbf{Y}_{ij} , between the contacts i and j , yield the effective point mobility of the i th contact point (Gibbs et al. 2007).

$$Y_i^\Sigma = Y_i + \sum_j \frac{F_j}{F_i} Y_{ij} \quad (8.8)$$

Simplified equations may be used for forces of same magnitude and random phases between the point contacts, for example, on plate fields with statistically many modes. In the latter case holds

$$|Y_i^\Sigma|^2 = |Y_i|^2 + \sum |Y_{ij}|^2 . \quad (8.9)$$

Chapter 9

Convolution and Binaural Sound Synthesis



9.1 Introduction

Auralization requires signal processing for merging the sound source signal with the sound propagation filter which was created by measurement or computer simulation. By adding the effects of the environment to the dry sound, we imprint the specific sound character the specific environment into the source. The classical mathematical tool to achieve this is convolution. It can be implemented in brute-force algorithms or in block processing based on FFT algorithms. Furthermore, binaural hearing of human receivers can be implemented along the filtering in a straightforward way. The main concepts are introduced in this chapter.

When the source signal and the system's transfer function or impulse response are obtained separately, the resulting output signal can be calculated by convolution (Eq. (7.3)). The convolution can be processed in various ways, either directly in time domain by using FIR filters or by using FFT convolution. In the latter case, however, it should be kept in mind that FFT requires fixed block lengths and is related to periodic signals. Time windows might be required for reducing artefacts from discontinuities. The same holds true for frame-wise convolution in the case of slow time-variant systems. Also, the technique of convolution or “filtering” (IIR, FIR) is valid for LTI systems exclusively. In case of time-varying systems, the excitation signal must be processed in frames representing pieces of approximate time invariance. In this case, filters might be adapted while processing and fading must be used to move from frame to frame (Fig. 9.1).

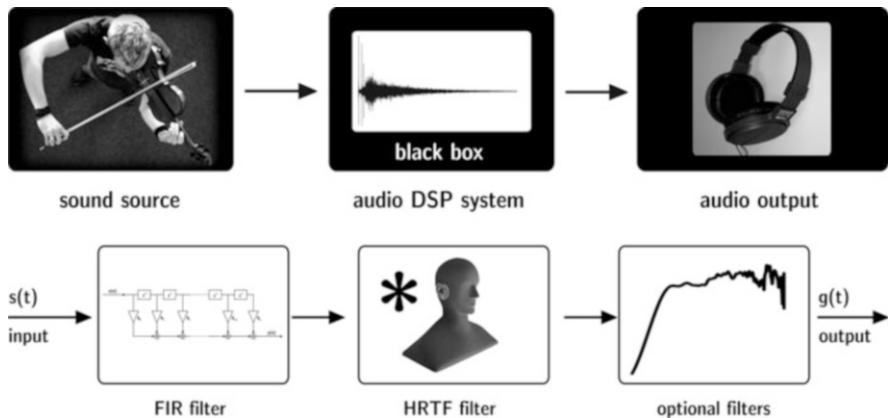


Fig. 9.1 Block signal processing (top) and signal filters (bottom) for auralization

9.2 Discrete Convolution

Signal processing requires a certain number of calculations steps. Floating point multiplications are the most time-consuming operations, while adding and storing is normally negligible in single-processor programming. Higher computational cost may take place when parallel processing and network communication is involved (see also Chap. 16).

Another special aspect is given by the usage of DSP processors in integer data formats. In this case the signals must be normalized before processing in order to avoid full-scale overflow. Although being faster in convolution, they might show disadvantages due to their smaller dynamic range available.

Here we use the convolution integral (Eq. (7.3)) in discrete form. The input signal, $s(n)$, and the impulse response filter, $h(k)$ are stored in vectors as temporal sequences. Thus the discrete convolution for calculation of the output signal, $g(n)$ reads (Fig. 9.2):

$$s(n) * h(n) = \sum_{k=0}^{N-1} s(k)h(n-k) = g(n). \quad (9.1)$$

MATLAB® Code for Discrete Convolution

Explicit code:

```
function g = convolution(s,h)

%Initialize signal vectors
s = s(:); %input signal
h = h(:); %impulse response
```

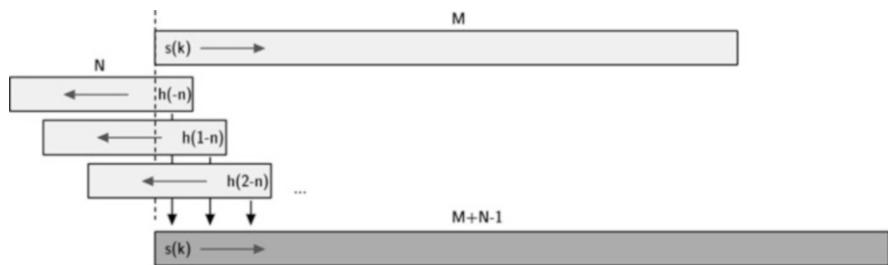


Fig. 9.2 Discrete convolution

```
%Determine length of vector
N = length(s);
L = length(h);
M = N + L - 1;

%Generate matrix for convolution
S = zeros(M,L);
for idx = 1:L
    %copy shifted s vector in matrix
    S( (1:N)+(idx-1), idx ) = s;
end

%Process multiplication for convolution
g = S * h; %output signal
```

or in short, using the built-in Matlab function

```
g = conv(s,h)
```

This process requires $N \cdot L$ floating-point multiplications, with N denoting the length of the input signal and L the length of the impulse response (FIR filter coefficients).

9.3 FFT Convolution

For this we use the convolution integral in the frequency domain. The input data is stored in a vector as a temporal sequence, and the impulse response and the output signal are considered vectors of same length. Then the FFT convolution reads as follows:

MATLAB® Code for Cyclic (FFT) Convolution

```
function g = FFTconvolution(s,h)

%Initialize signal vectors
s = s(:); %input signal
h = h(:); %impulse response

%Determine length of vector
N = length(s);
L = length(h);

%append zeros in the end for same length
s = [s; zeros(L-1,1)];
h = [h; zeros(N-1,1)];

%Fourier Transformations
s_spk = fft(s);
h_spk = fft(h);

%element-wise multiplication
g_spk = s_spk .* h_spk;

%Inverse Fourier Transformation
g = ifft(g_spk);
```

9.3.1 Segmented Convolution

For continuous signals, however, processing in one block is not possible, unless we have extremely large memory space available. But even if it was possible to store all the data, we could not wait for the complete input signal to be passed to and stored in the output vector¹. Furthermore, with adaptive filters, the latency would become unacceptably large. Therefore we have to cut the signal into temporal segments,

¹This is obviously not possible in real-time applications, where the output signal must be replayed instantaneously.

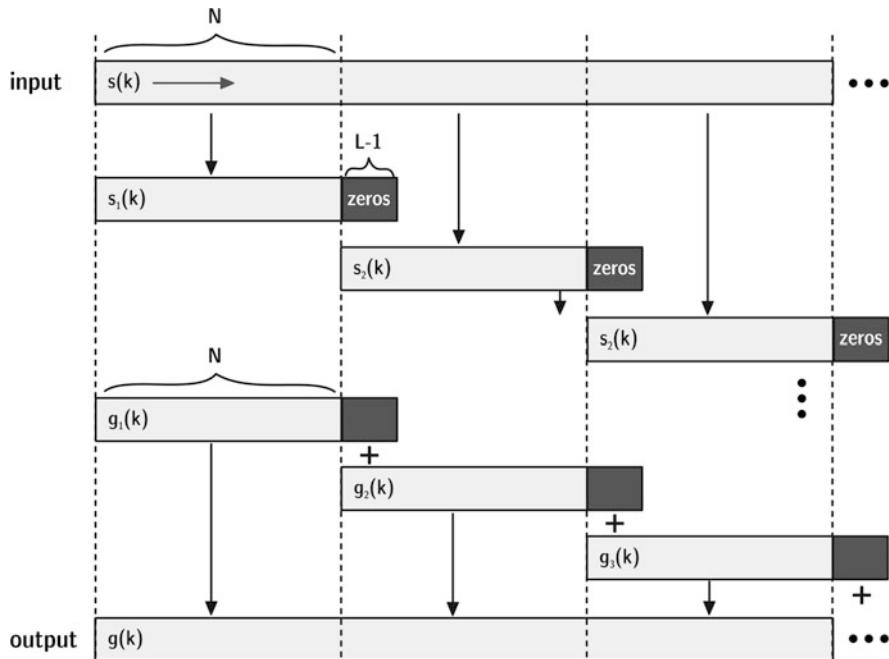


Fig. 9.3 Overlap-add convolution

process them frame by frame and transfer the results to the output unit sequentially. From frame to frame, the filter can also be changed. Fast segmented convolution typically is used for convolving short impulse responses with quasi infinite signals. The overlap-add algorithm is one example out of various alternatives (Papoulis 1981).

In temporal sequences of Δt , parts of the signal are copied with a length $\Delta t + t_{\text{fade}}$. Δt denotes the reciprocal update rate of the filter (or any chosen filter length) and t_{fade} the duration of a fading interval from frame to frame (Fig. 9.3).

A segment of the input signal is now discussed using one segment between t_1 and $t_2 + t_{\text{fade}} = t_1 + \Delta t + t_{\text{fade}}$. This segment is extended to the filter length (zero padding). Then the signal segment and the (actual) filter are transformed by FFT. The resulting spectra are multiplied and the product transformed back to time domain by IFFT. The convolution product is then multiplied by a fading function with exponentially weighted slopes to fade in and out.

Finally, the signal is fed into the data buffer of the output signal at the corresponding time index. The next frame is added into the buffer at time t_2 , just overlapping with the previous segment in the fading zone. It must be ensured that the signal power in the overlap zone remains constant (Fig. 9.4).

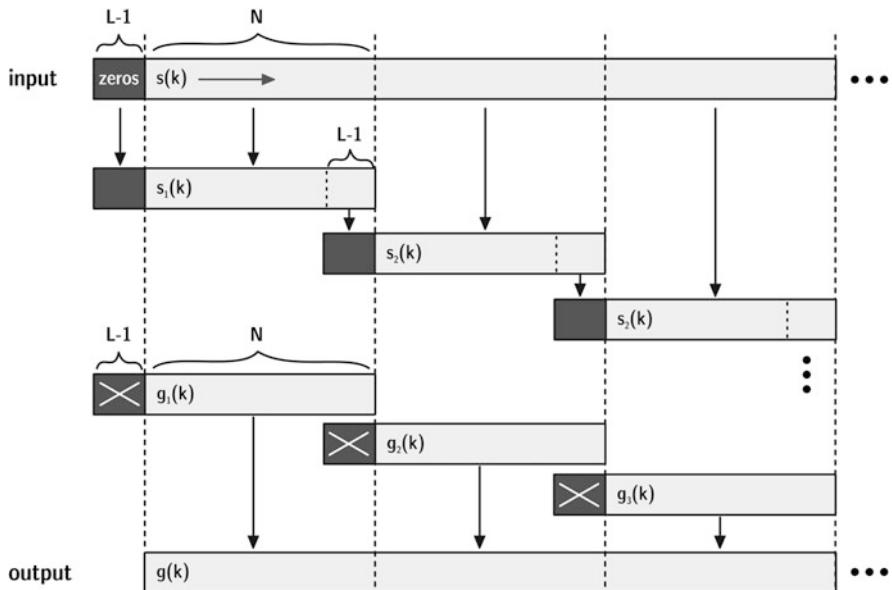


Fig. 9.4 Overlap-save convolution

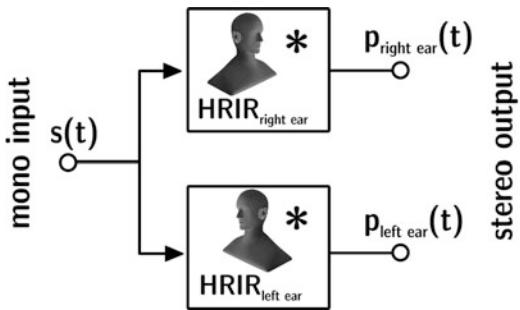
9.4 Binaural Synthesis

With the tools described in the previous chapter, it is a very small step towards the most important component of virtual reality systems: Filtering is used to connect a sound signal to a spatial cue. In order to achieve user immersion in VR systems, it is indispensable to create spatial sounds which match the visual spatial impression and other multimodal dimensions. Interesting applications of acoustic virtual environments are found already in free-field situations. Examples are outdoor or near-field scenes, indoor scenes with small relevance of reverberation or artificial scenes of augmented reality, where an additional sound is mixed into an existing acoustic stimulus.

The basic task in creating an auralization is to place a sound source into 3D space. Any mono source signal properly characterized and calibrated according to Chap. 8 can be processed in such a way that its perceptual cues are amended by a spatial component. A stereo or surround setup is capable of creating an effect of phantom sources (see 17.3.1) which can produce an appropriate spatial effect. A binaural mixing console can be used for processing headphone signals by using HRTF.

As mentioned in Sect. 6.4, sound localization and spatial hearing can be understood as an effect of the transfer function of the outer ear, the head-related transfer function, HRTF. With a database of HRTF², any direction of sound incidence can be

²Databases of HRTF are available in the Internet; see 4 on page 86.

Fig. 9.5 Binaural synthesis

simulated, when a mono source $s(t)$ is convolved with a pair of head-related impulse response, HRIRs.

$$\begin{aligned} p_{right\ ear}(t) &= s(t) * HRIR_{right\ ear} \\ p_{left\ ear}(t) &= s(t) * HRIR_{left\ ear} \end{aligned} \quad (9.2)$$

Instead of convolution (FIR filter, segmented convolution, etc.), other filter techniques can be used, provided the binaural cues (ILD, ITD) of HRTF are represented properly (Fig. 9.5). Also IIR filters, for instance, were created to simulate the poles and nodes of HRTF with success (Kistler and Wightman 1992; Huopaniemi et al. 1997; Hammershøi and Møller 2002); see also the excellent overview in (Hammershøi and Møller 2005).

The method explained in this single page seems to be a minor detail, but it is not. Binaural synthesis is the key to many techniques for auralization and virtual reality systems. More information will be given later; see Chaps. 16 and 17.

9.5 Binaural Rendering

Binaural rendering is the technique for creating multi-source binaural synthesis or complex scenes. This tool is ideal for the auralization of free-field environments when a small number of sources or reflections are to be modelled³. The modelling of free-field propagation requires a source model or recording and analytic calculation of the complex amplitude of the sound pressure signal at the receiving point.

The methods introduced in Chaps. 2 and 3 are well suitable to obtain the propagation path. In a most elementary case, this is solved by recording sound in a certain distance and correcting it for the propagation law (spherical wave, for instance) to account for the distance of the receiver. Binaural synthesis will take

³This will be no problem up to 32 or 64 sources (channels), but it is currently not applicable in case of simulating binaural reverberation in room acoustics.

care of the direction of the sound incidence. Of course, multiple sources can be simulated and auralized at the same time.

For spherical waves, the free-field binaural filter corresponding to one source reads:

$$\underline{H}|_{\text{left,right}} = \frac{e^{-j\omega t}}{ct} \cdot \underline{H}_{\text{source}}(\theta, \phi) \cdot \underline{H}_{\text{air}} \cdot \text{HRTF}(\vartheta, \varphi)|_{\text{left,right}} \quad (9.3)$$

with \underline{H} denoting the filter spectrum normalized to 1 m distance, t its delay, $j\omega t$ the phase lag due to retardation, $1/(ct)$ the distance law of spherical waves, $\underline{H}_{\text{source}}$ the source directivity (Sect. 2.5) in source coordinates (θ, ϕ) , $\underline{H}_{\text{air}}$ the low pass of air attenuation (Sect. 3.7), and HRTF the head-related transfer function (Sect. 6.4.1) of the sound incidence the listener coordinates at a specified orientation (ϑ, φ) .

The total filter is used as the convolution filter for processing the source signals $s_i(t)$. In the case of N sources⁴, the resulting signals are superposed. In time domain, this procedure leads to

$$h(t)|_{\text{left,right}} = \sum_{i=1}^N s_i(t) * \mathcal{F}^{-1}(\underline{H}_i)|_{\text{left,right}} \quad (9.4)$$

with $\mathcal{F}^{-1}(\underline{H}_i)$ denoting the inverse Fourier transform of the contribution of source i .

Sound transmission or diffraction is modelled by adding attenuation filters (see Sects. 13.4 and 3.5) into Eq. (9.3), the parameters of which are related to geometric data or material properties. In the same way, transfer paths of BTPS can be integrated easily, as long as they do not affect spatial cues.

With the possibility to place and move 3D sounds in space, the most important components of virtual acoustics are available. Typical block lengths of HRTF filters are in the range of 128 or 256 samples⁵. FIR or IIR filters (see Sect. 7.8) can be used, either on DSP platforms or by using PC environments with standardized audio interfaces. Investigations regarding effects of reduced filter lengths on localization are available (Kulkarni and Colburn 1995, 1998). According to (Hammershøi and Møller 2005), filter lengths can be reduced to 72 taps without audible effects. Filter processing can be performed in the frequency domain as well, by using overlap-add or overlap-save FFT and multiplication with 64 complex data (Fig. 9.6).

⁴Which may also include “image sources”, see Sect. 11.4.

⁵The sampling rate for audio processing is usually 44.1 kHz.

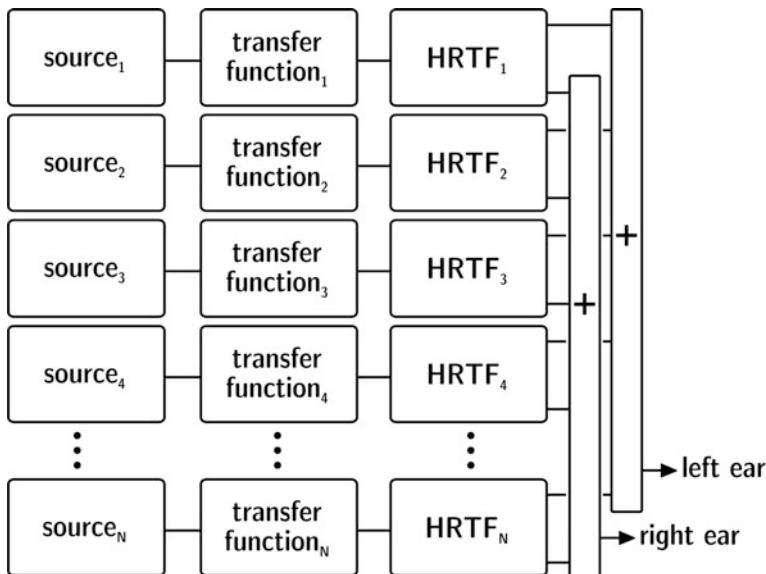


Fig. 9.6 Block diagram of free-field auralization

9.6 Spatial Resolution of HRTF

The HRTF database should cover all psychoacoustic effects of localization and dynamic changes in localization. Subjective tests of localization showed that humans are able to discriminate between differences of 1 degree in azimuth in frontal direction in the horizontal plane (see Sect. 6.4). In other directions the angle intervals can be chosen larger. By using interpolation between the HRTF pairs available or by encoding HRTF in a suitable functional basis (Torres et al. 2004), the size of the database can further be reduced (Minaar et al. 2005)

In applications of auralization with small distances between source and head, standard HRTF are not sufficient (Brungart et al. 1996). Standard HRTF are defined for plane waves from specified directions. The prerequisite of independence of the distance is valid for distances > 1 m. Most obviously, distances of sources can be modelled by changing levels. But in distance perception not only the relative level is evaluated. Spectral changes, and thus changes of ITD and ILD are also caused by sources approaching the head at constant angle, but distances closer than 1 m. This kind of dataset must be determined with special care and unambiguous definition. A reasonable definition for a reference sound field is the spherical wave. Accordingly, near-field HRTF are defined in the classical way, but they correspond to a point source at near distances⁶ (Fig. 9.7).

⁶Proper point sources must be constructed for this kind of measurement. Alternatively, near-field HRTF can be calculated by using BEM or FDTD.

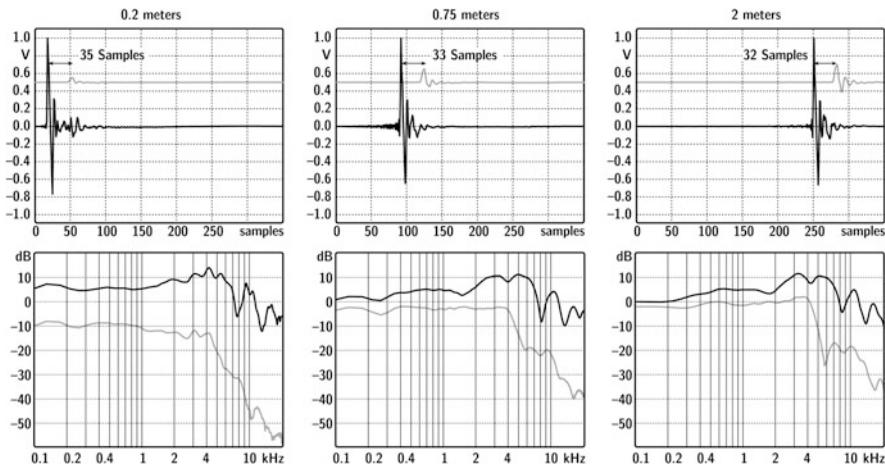


Fig. 9.7 Right to left: From standard to near-field HRTF (after Lentz 2007)

One can easily calculate the memory space required for such a high spatial resolution of HRTF for various angles and distances. Several approaches for data compression are possible, too. The complexity can be reduced either in space and distance or in FIR filter length (which corresponds to spectral resolution). Another possibility is to choose a rough amplitude scale (low bit scaling).

With the technique of binaural mixing and the opportunity to create any kind of spectral, temporal and spatial sound event, numerous applications which use sound as the carrier of information come into play. One tool in this respect is the so-called auditory display.

Chapter 10

Simulation Methods



10.1 Introduction

After getting familiar with the principle of auralization, i.e. the separation of the acoustic problem in a model of signal propagation and binaural synthesis or other spatial audio formats, and after having characterized the sources, we will now focus on the second key component of auralization. We consider the excitation signal as known and ready for convolution. Now, the propagation functions for sound and vibration must be measured or modelled. The task is to define and apply a theoretical approach to the propagation problem, either in free propagation or in a problem with boundary conditions. The methods are defined for frequency-domain or time-domain formulations, such as FEM, BEM and FDTD, respectively. Other methods are adapted from network or energy models in electrical and mechanical engineering. It is clear that not all methods listed in this chapter can be used for virtual reality applications. The computation time involved in the methods is to be discussed separately. In the future, however, more and more simulation methods will be applicable. The focus, therefore, is put on the physical background of the simulation methods and not on computational constraints. In Chaps. 11, 12, 13, 14 and 15, we will discuss the up-to-date simulation methods applicable for virtual reality systems in detail.

10.2 Simulation Methods for Sound and Vibration Fields

The modelling of sound and vibration propagation is one of the main problems in theoretical and numerical acoustics (Fig. 10.1). All basic features of sound radiation (Chap. 2) and of sound fields (Chap. 3) come into play now. Nevertheless, specific methods particularly for non-analytic approaches must be discussed in this chapter. Boundary conditions and field geometries mostly do not match the elementary

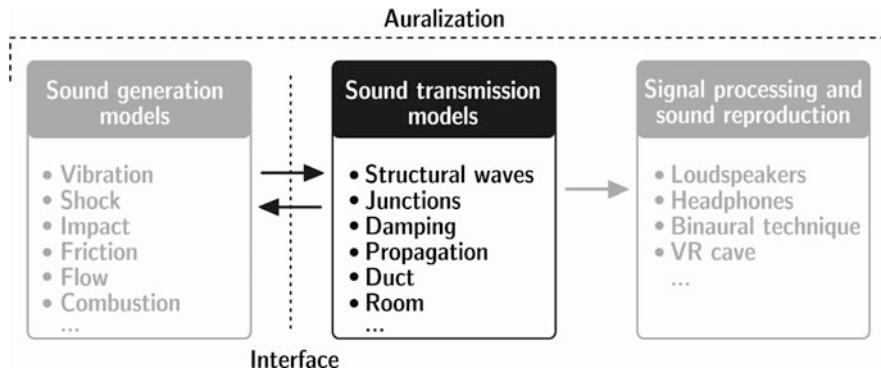


Fig. 10.1 Simulation models in acoustics

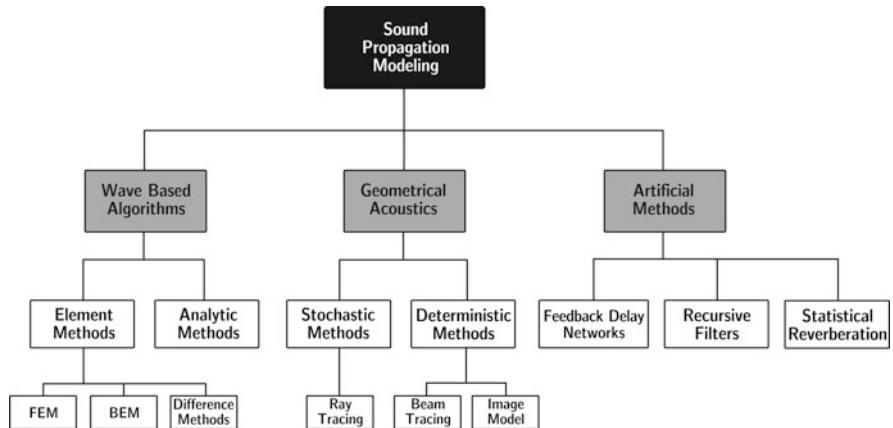


Fig. 10.2 Classification of models for simulation of sound propagation

conditions of standard coordinate systems like Cartesian, spherical or cylindrical geometry. The basic solutions we found in Chaps. 2 and 3 are still interesting as they show the basic features of sound sources and propagation. The details and the fine structure in the results, however, can only be obtained when the real geometry and the conditions of the propagation space are taken into account with sufficient accuracy.

The accuracy of the models can be discussed on a physical basis and on a psychoacoustic basis. The discussion on the physical basis is related to the size of objects in relation to the wavelength (diffraction), to the possibility of neglecting of phase effects (high modal density) to the variety of wave types contributing to the transfer function and to elementary features of the signals simulated concerning density of samples in time and frequency domain (Fig. 10.2).

A sound propagation or transmission problem can be described by Green's functions. They result from a formulation of the wave equation by using the potential

function, $g(r|r_0)$. It corresponds to the sound field quantities by the derivations in space and time (Skudrzyk 1971; Mechel et al. 2002):

$$\begin{aligned} p &= \rho_0 \dot{g}, \\ v &= -\nabla g. \end{aligned} \quad (10.1)$$

The benefit from Green's function is easily understood when it is applied for sources and spatial propagation paths from a point r_0 to a point r . For example, in the case of a point source with volume flow Q (see 2.2), Q can be described as

$$Q = \lim_{a \rightarrow 0} \left(4\pi a^2 \frac{\partial g}{\partial r} \right). \quad (10.2)$$

Its Green's function in free space is

$$g(r|r_0) = \frac{e^{-jk|\vec{r} - \vec{r}_0|}}{4\pi |\vec{r} - \vec{r}_0|}. \quad (10.3)$$

Coming back to the acoustic radiation problem and to the Helmholtz equation with source term ($f = jkZ_0q$, see Eq. 4.5)

$$\Delta p(r) + k^2 p(r) = -f(r_0), \quad (10.4)$$

the Green's formulation leads to the Helmholtz-Huygens integral¹.

$$\begin{aligned} p(r) &= \iiint f(r_0) g(r|r_0) dV_0 \\ &\quad + \iint \left(g(r|r_0) \frac{\partial p(r_0)}{\partial n} - p(r_0) \frac{\partial g(r|r_0)}{\partial n} \right) dS_0 \end{aligned} \quad (10.5)$$

With this integral, the resulting sound pressure of various kinds of source distributions $f(r_0)$, in a volume including any kind of reflections from boundaries on a surface surrounding, the sources can be calculated. The integration surface can also represent a virtual surface where the sound field is expanded into elementary (secondary) sources (Huygens principle). It is interesting that the surface source arrangement consists of monopoles (second term in the surface integral) and dipoles (first term with the pressure gradient $\partial p/\partial n$).

The discussion of Green's functions was, so far, related to harmonic signals. If we expand the radiation in time domain by assuming an impulse excitation², the

¹The derivations in the surface integral defined in the direction normal to the surface elements (Mechel et al. 2002).

²With impulse excitation we consider a constant harmonic spectrum with zero phase; see 7.3.2.

tremendous importance of Green's functions becomes clear. They are *filters* transporting signals, $f(r_0, t_0)$, to the sound pressure at the receiver, expressed in the convolution of source functions with Green's functions (note the temporal relationship between source and receiver point):

$$p(r, t) = \int_{-\infty}^t f(r_0, t_0) g(r|r_0, t - t_0) dt_0 \quad (10.6)$$

By Fourier transformation we obtain adequate solutions in the frequency domain. The kernel of the convolution can then be expressed in terms of the source volume velocity, Q , which leads to the formulation of a transfer impedance

$$Z = \frac{p(\omega, r)}{Q(\omega, r_0)}. \quad (10.7)$$

The concept of a transfer function between the sound pressure at one point to the sound pressure of another point can also be chosen, if the source characterization is based on a near-field pressure signal:

$$H(\omega) = \frac{p(\omega, r)}{p(\omega, r_0)} \quad (10.8)$$

Similarly, transfer functions and Green's functions for structural acoustics can thus be defined.

Which approach is preferable (Green's function, transfer impedance or transfer function) depends on the kind of source. Force or pressure sources are more straightforward to be coupled to Green's functions; velocity sources can be coupled to transfer functions. Most easy to remember is the fact that the multiplication of the source signal with the transmission function should yield a sound pressure signal in the end.

10.2.1 Reciprocity

Green's functions are reciprocal (Lyamshev 1959). Reciprocity is one of the most powerful tools in the determination of acoustic transfer functions. The problem of sound and vibration transmission in a passive linear time-invariant system can be solved in both directions. Transfer functions, transfer impedances and Green's functions are identical when source and receiver points are interchanged. For an accurate description we have to distinguish between sound propagation from an airborne source to a receiver point and vibro-acoustic transmission from a force source exciting a structure which radiates to a receiver point (Fahy 1995).

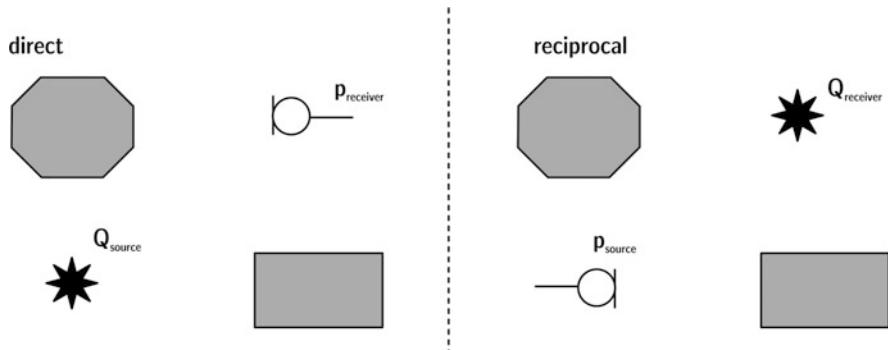


Fig. 10.3 Reciprocity of airborne sound propagation from a monopole source. (After Fahy 1995)

Airborne Sound Reciprocity

The direct formulation consists of a real volume source emitting sound which is received in the field space (Fig. 10.3). The ratio between the sound pressure at the receiver and the volume flow of the source³ is $p_{\text{receiver}}/Q_{\text{source}}$. In the reciprocal arrangement, this ratio is identical with the ratio of the pressure at the source point and the volume flow at the receiver, $p_{\text{source}}/Q_{\text{receiver}}$. However, specific reference conditions must be defined. The reference condition concerns the determination of the sound pressure, which must be obtained in a mechanically blocked state ($Q = 0$)⁴. Thus

$$\left. \frac{p_{\text{receiver}}}{Q_{\text{source}}} \right|_{Q_{\text{receiver}}=0} = \left. \frac{p_{\text{source}}}{Q_{\text{receiver}}} \right|_{Q_{\text{source}}=0}. \quad (10.9)$$

Vibro-acoustic Transfer Function Reciprocity

In a problem of a force source exciting a structure which radiates sound, the reciprocity relationship is given by

$$\left. \frac{p_{\text{receiver}}}{F_{\text{source}}} \right|_{Q_{\text{receiver}}=0} = \left. \frac{v_{\text{source}}}{Q_{\text{receiver}}} \right|_{F_{\text{source}}=0}. \quad (10.10)$$

We consider the sound pressure received from the direct formulation as dependent on the force injected on the sound-radiating structure. This ratio is identical with the

³= acoustic transfer impedance

⁴Equivalent to open-circuit situation for obtaining the voltage in terminated electric circuits

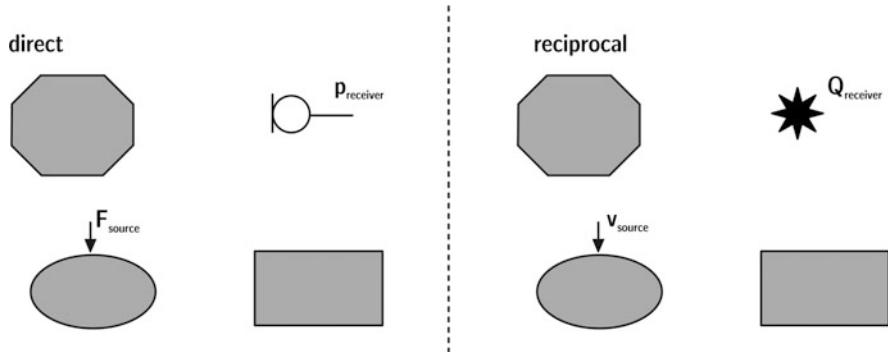


Fig. 10.4 Vibro-acoustic transfer function reciprocity. (After Fahy 1995)

ratio in reciprocal approach between the volume flow at the receiver point in the fluid and the free ($F = 0$) normal velocity on the structure (Fig. 10.4).

Vibro-acoustic Green's Function Reciprocity

When not the force but the velocity of a vibrating surface element is the input at the source side, the reciprocity relationship is given by

$$\frac{p_{\text{receiver}}}{v_{\text{source}} dS} \Big|_{Q_{\text{receiver}}=0} = \frac{p_{\text{source}}}{Q_{\text{receiver}}} \Big|_{v_{\text{source}}=0}. \quad (10.11)$$

Here, we consider the sound pressure received from the direct formulation as dependent on the local normal velocity of a surface element dS of the sound-radiating structure. This ratio is identical with the ratio in reciprocal approach between the sound pressure at the receiver point in the fluid and the blocked ($v = 0$) sound pressure on the structure. This relationship is valid for all surface elements of the structure.

After having the principles of sound propagation discussed between points in a fluid and between structures and fluid on analytic examples, in the next section we will focus on numerical methods in the frequency and time domains. These methods are applicable in a more general sense and for any kind of geometric conditions of source, receiver and environment (Fig. 10.5).

10.2.2 Frequency Domain Models

In frequency domain calculations, a constant frequency is considered. Hence the discussion of transfer functions is based on harmonic signals.

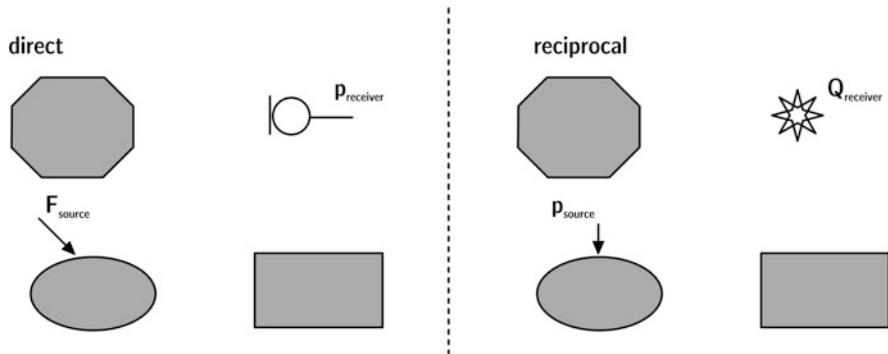


Fig. 10.5 Vibro-acoustic Green's function reciprocity. (After Fahy 1995, with $dQ_{\text{source}} = v_{\text{source}} dS$)

$$s(t) = A \cdot e^{j\omega t}, \quad (10.12)$$

with $s(t)$ denoting any complex-amplitude harmonic signal of vibration, source volume velocity or sound pressure in the field point. In this case the wave equation is reduced to the homogeneous Helmholtz equation (compare Eq. 4.5), for instance, for the sound pressure in the field volume, $x \in V$:

$$\Delta p(x) + k^2 p = 0. \quad (10.13)$$

For obtaining the free (modal) response, we consider a source-free-field domain, with the right-hand side of the equation being zero. This equation enables us to calculate the sound pressure relationships between field points, and it is the perfect approach for calculating transfer functions (Eq. 10.8).

Depending on the geometry, as mentioned above, we mostly cannot solve the Helmholtz equation straightforward, since the problem geometry and the boundary conditions do not match elementary coordinate systems. Instead, the problem can be solved by numerical methods. The most prominent approach is the spatial discretization into small elements. In the discrete formulation, the Helmholtz equation can be transformed into a linear system of equations in the field space. The approach to solve the problem of wave physics is (a) the usage of the Helmholtz equation in integral formulation or (b) the principle of energy conservation (energy minimum) in the Lagrange formulation as a variational problem. The first concept is used in the boundary element method and the latter in the formulation of the finite element method.

Meshing

Numerical models require spatial discretization by introducing a “mesh” (Fig. 10.6). A mesh is a discretized grid of surface and/or field points (nodes) and corresponding elements of groups of nodes. The elements can be rectangular, triangular and tetrahedral, just to give a few examples. The volume or the boundary is meshed into elements depending on the method used. The degree of discretization depends on the local waviness of the sound or vibration field. At high frequencies (small wavelengths) the discretization must be sufficiently large to allow interpolation between field points without too much loss of precision. The final limit, of course, is similar to the sampling theorem. The practical limit is roughly six nodes per wavelength.

Meshes must be strictly designed with respect to the numerical method used. For finite time differences, the mesh must be geometrically regular, whereas meshes for finite or boundary elements can be more flexible in shape and size. However in the latter case, mesh elements are crucially coupled to the specific formulation of the numerical wave model.

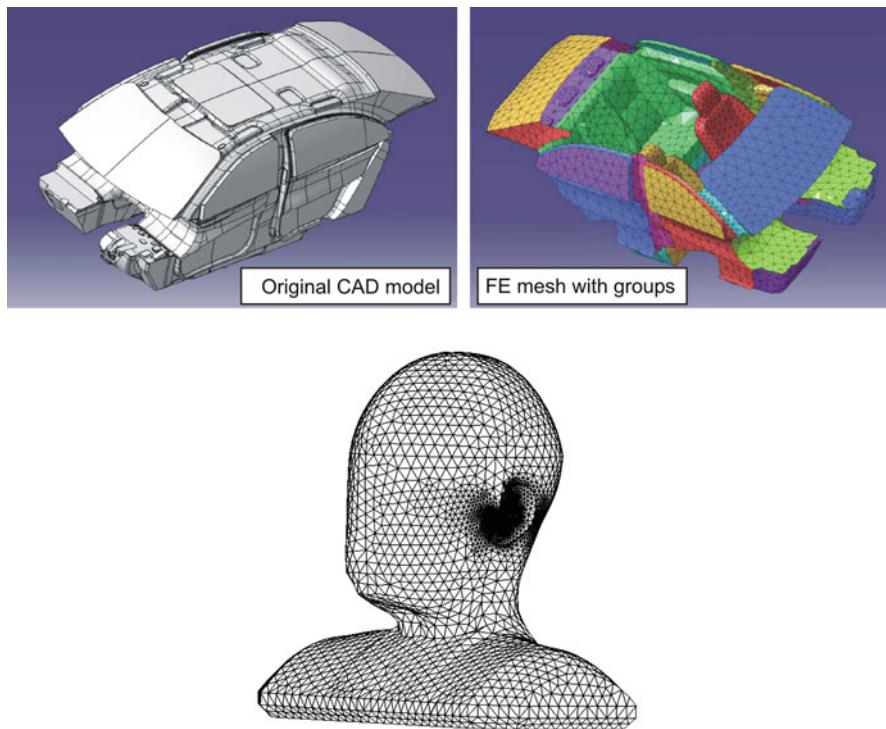


Fig. 10.6 Examples for meshes for numerical wave propagation analysis. A car compartment (top) and a customized dummy head (bottom)

Boundary Element Method BEM

The boundary element method is explicitly related to the Green's function, $G(r|r_0)$, where r_0 denotes a source position and r a set of field points. The radiation problem, thus, is re-arranged into the Helmholtz-Kirchhoff integral equation and discretized. The Helmholtz-Kirchhoff integral is the source-free Helmholtz-Huygens integral (see Eq. 10.5):

$$p(r) = \iint \left(g(r|r_0) \frac{\partial p(r_0)}{\partial n} - p(r_0) \frac{\partial g(r|r_0)}{\partial n} \right) dS_0 \quad (10.14)$$

with the Green's functions into 3D free space

$$g(r|r_0) = \frac{e^{-jk|\vec{r} - \vec{r}_0|}}{4\pi|\vec{r} - \vec{r}_0|} \quad (10.15)$$

which fulfil the far field radiation (Sommerfeld) condition of vanishing sound pressure for $r \rightarrow \infty$.

The kernel of the integral thus contains monopole and dipole sources. The main application for BEM is radiation or equivalent radiation problems such as scattering. Radiation problems are characterized by boundary conditions (local impedances or admittances), including a vibration velocity as the driving source region. This integral is formulated in discretized form on a surface mesh and solved numerically with matrix algebra. The crucial point of the BEM formulation is the numerical non-uniqueness. It is worth to be mentioned that in contrast to FEM matrices, BEM matrices are full. In the famous Burton/Miller approach (Burton and Miller 1971), these problems are discussed in all details. Another strategy for avoiding numerical problems is the so-called CHIEF point (combined integral equation formulation) method (chapter by Ochmann in (Mechel et al. 2002)). BEM matrix solvers are available, some codes even in free software.

The complexity of BEM can be roughly summarized as follows: A simulation which must be calculated up to a frequency of f requires a mesh element size of at most $c/6f$. The resulting model size of a surface S then is

$$N = \frac{36 S f^2}{c^2} \quad (10.16)$$

The BEM matrix then contains N^2 entries. Solvers of PC software today are capable of inverting a matrix of 8000 nodes in about 30 s⁵. This result holds for one frequency. In a problem of a required frequency resolution of 1 Hz, the calculation

⁵Example: iterative solver on a machine with 3.4 GHz i7-6700 (4 core) processor and sufficiently large RAM to keep the matrix problem in-core.

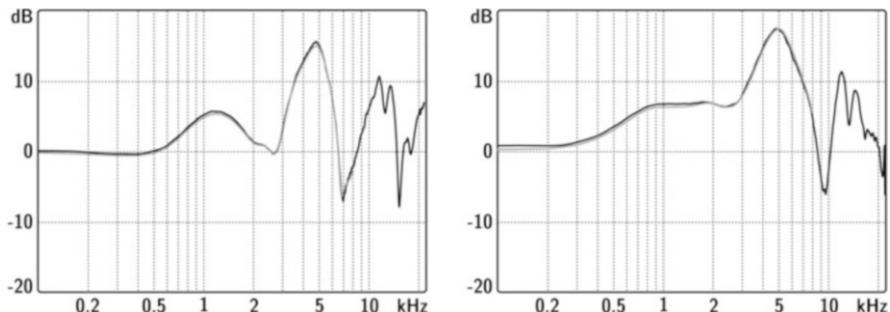


Fig. 10.7 BEM-calculated HRTF of a customized dummy head; left, frontal incidence; right, 45° in the horizontal plane for the ipsilateral ear. Head model corresponding to Fig. 10.6 (bottom). Example following the method described in Fels et al. (2004); see also 6.4.1

time multiplies by f , which yields some $f/60$ h for numerical generation of a complete set of transfer functions to all field points.

Advanced techniques like fast multipole BEM (Sakuma and Yasuda 2002; Yasuda and Sakuma 2003; Marburg and Schneider 2003) are developments which allow meshes to be separated into regions of high discretization and other regions, with the effect of transfer propagation. The complex linking between mesh elements is thus re-arranged in a hierarchical way (Fig. 10.7).

Finite Element Method FEM

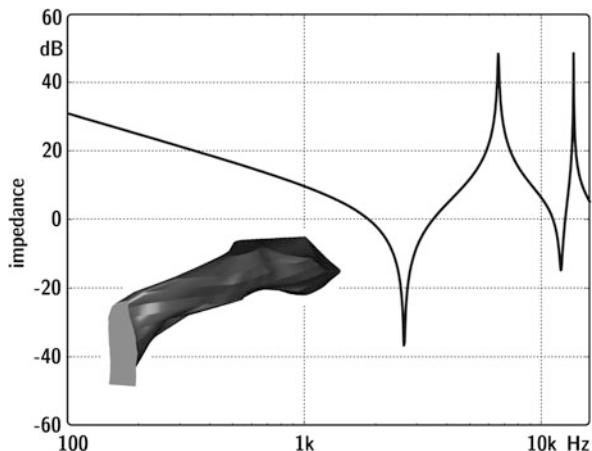
Finite elements are created by the discretization of a field volume into volume elements. In these elements the energy formulation of the harmonic field equations is used. This is generally known as Hamilton's principle of minimum energy. Any disturbance of the system equilibrium⁶ leads back to a stable and a minimum energy state. Due to its general energetic formulation, this principle is used for mechanical problems of static load and deformation (also for crash test simulation), for fluid dynamics, heat conduction, electromagnetic or acoustic field problems, and it is also the basis for FEM (Zienkiewicz 1977).

The field space for the acoustic problem must be discretized into suitable volume elements. For each element the relation between the forces and the displacements is introduced by using the variational approach. Thus the variational approach is used to identify the field quantities for minimum energy, element by element. The total energy is thus the sum of all element energies.

In every element the so-called shape functions, ψ , are defined to represent the sound pressures within the elements. At the nodes between the elements, the shape functions must fit continuously.

⁶For example, by applying virtual displacement

Fig. 10.8 Example of an ear canal impedance $20\log(\underline{Z}/Z_0)$ (for the CAD model shown, ear canal entrance and reference plane on the left) calculated using FEM



All elements' entries are combined into a so-called “stiffness” matrix, S ⁷, a mass matrix, M , and a damping matrix, C . Furthermore, source contributions and boundary conditions are formulated and integrated into a matrix equation including S , M and C , which is to be solved to obtain the sound pressures in the field space from the shape functions, read at their nodes.

In FEM solvers the direct solution to determine the matrix' eigenvalues can be used, by using the matrix equation without further subspace conditions. In the indirect method, the problem is projected on a modal basis onto an equivalent eigenvalue problem of orthogonal modes. The latter method has the great advantage that sources and boundary conditions can be studied in a second step. The numerical complexity is then given by the size of the modal basis and not by the FE mesh size.

FE used for sound pressure calculation in small- or mid-size rooms for frequencies up to some kHz is currently state of the art. Using PC software, typical mesh sizes are in a range of 100,000 nodes, and typical calculation times are in the order of a few minutes per frequency (Fig. 10.8).

Modal Approach and Modal Superposition

As in the finite element formulation, modes may serve as an orthogonal functional basis for expansion of broadband results into a series of modes. The modes as such can be described by using elementary second-order resonators (see also Sect. 4.2) with midband frequency, half-width or quality. This way of modal analysis is well known as powerful tool in the measurement of complex systems. Also, in simulation problems, a modal basis gives very important information about the system. Modal density and modal overlap can be studied, and it can be decided to which extent the

⁷Although other forces may be used here, the historic name is related to problems of static deformation by Hooke's forces.

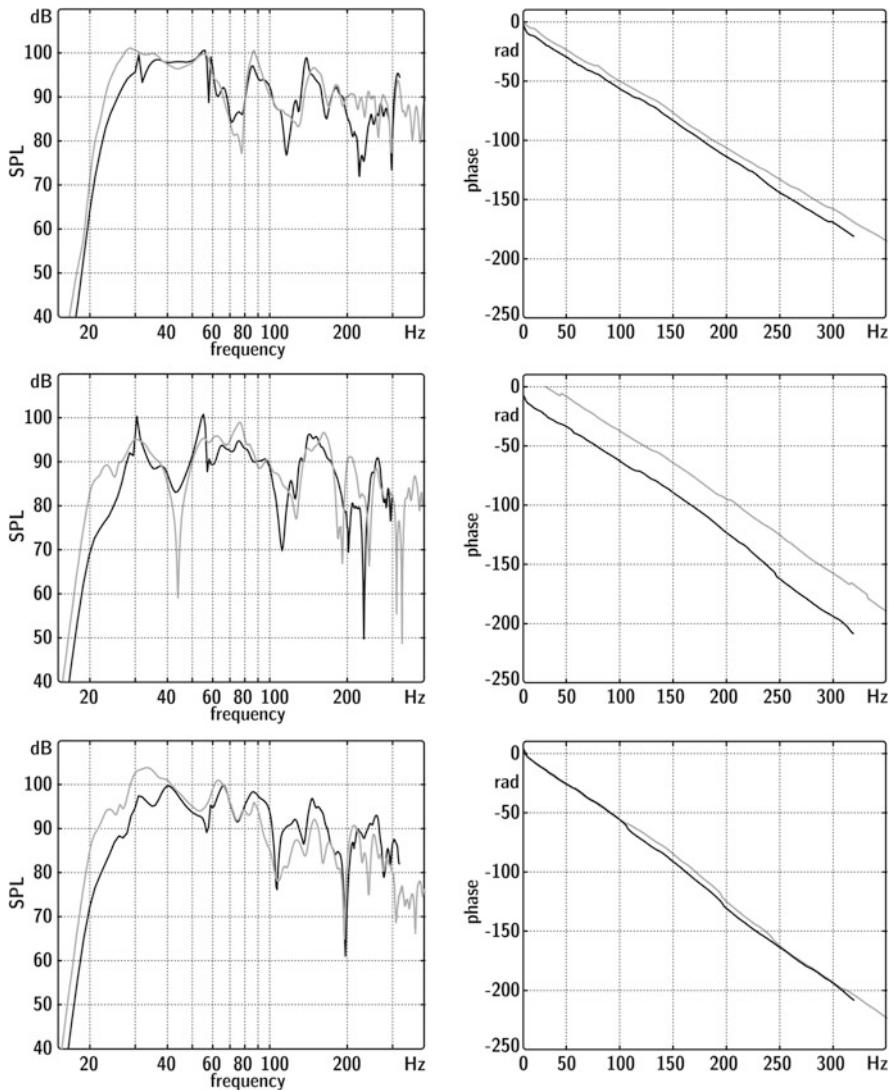


Fig. 10.9 Modulus and phase of results from FEM calculations and measurements in a recording studio. (example after (Aretz 2007))

exact complex modal response is relevant. The transition point from separated modes to highly overlapping modes is of crucial importance. This was first discussed in Sect. 4.3.1, but it is generally interesting for all problems of acoustics and vibration (Fig. 10.9).

Frequency Spacing

The resolution of numerically determined spectra must be sufficiently high to identify all relevant modal details. Too high a resolution, on the other hand, is useless and just contains redundant information. The scale for defining reasonable frequency spacing is given by the width of modes. Each mode of width

$$\Delta_{f,\text{mode}} = \frac{\delta}{\pi} \quad (10.17)$$

must be covered by at least two frequency lines. $\Delta_{f,\text{mode}}$ denotes the full width at half maximum of a typical resonance. This kind of line spectrum can well approximate a second-order bandpass function with a damping of δ . The decay time corresponding to 60 dB level drop of this mode is approximately

$$T = \frac{6.9}{\delta}. \quad (10.18)$$

The required frequency spacing to model the system is thus

$$\Delta f = \frac{2\delta}{\pi} = \frac{4.4}{T}, \quad (10.19)$$

with T denoting the average decay time derived from the system impulse response.

Statistical Energy Analysis SEA

Statistical energy analysis, SEA, was introduced by Lyon and Maidanik in the early 1960s (Lyon 1975). The model of coupled resonators serves well to understand the basic principles. On this basis, any complex system of coupled resonances, in problems of airborne sound or structural vibration, or both, can be described by using energy balance and energy flow. At this stage of abstraction, SEA offers a powerful technique to calculate sound pressure levels and thus power flow between the subsystems of a complex structure.

The general approach can be compared with the situation of water reservoirs and a connecting tube system. Water pumps are “sources”; losses from damping and radiation can be modelled as water loss in porous ground or vapourization, respectively. Coupling is given by connecting tubes with a certain cross section and capacity.

The amount of water represents energy in a subsystem, water flow produced by a pump represents sound power injected into a system; connecting tubes are similar to sound power transfer between systems. By using SEA the amount of water in each basin can be calculated for steady-state conditions (Fig. 10.10).

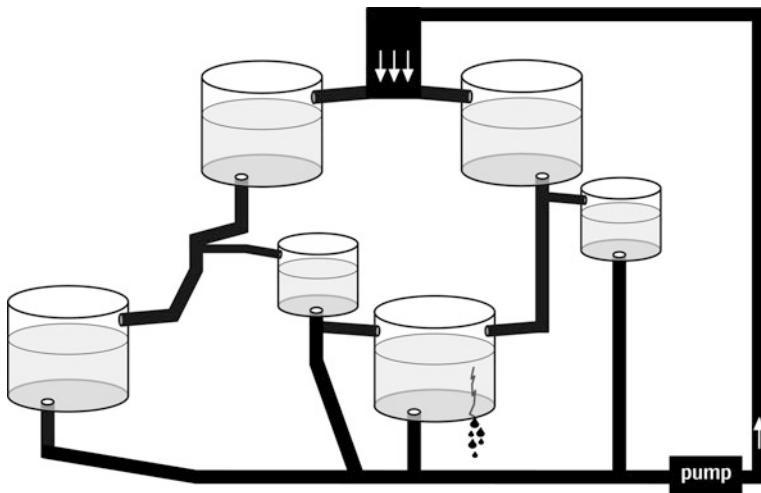


Fig. 10.10 Water flow in a system of basins and tubes

The difficult task, however, must still be discussed. Because calculation of sound energy in a subsystem is a more complex problem than dealing with water, we must define proper conditions of energy analysis. The crucial point of SEA is energy stored in resonators (modes) and statistical modal overlap in certain frequency bands. The subsystems contain modes of all kind, compressional waves, flexural waves on plates, etc. Typically subsystems are defined separately for each medium, material and shape separately. They should be clearly separated so that the energy exchange by coupling is small.

All data used in SEA calculations are frequency-dependent since energy, energy flow and boundary conditions are defined for frequency bands. The more modes are present in the frequency bands considered, the more precise the calculation.

The power extracted from a subsystem i is given by

$$\Pi_i = \omega \eta_i E_i, \quad (10.20)$$

where η_i is the damping loss factor and E_i the steady-state energy stored in modes of spectral density ν_i . When two subsystems are coupled, conservation of energy and balance leads to equilibrium when

$$\Pi_{ik} = \omega \nu_i \eta_{ik} \left(\frac{E_i}{\nu_i} - \frac{E_k}{\nu_k} \right). \quad (10.21)$$

The effective energy transport is, thus, expressed by the ratio of absolute energy in the frequency band and the modal density. This is appropriate in a statistical sense since the probability of energy exchange in modal coupling depends on several factors including the interaction of normal and tangential modes, fluid-structure

coupling and geometric factors. It is essential that subsystems are independent, weakly coupled and statistically contain many independent modes.

Furthermore, reciprocity can be found in SEA, and this is a powerful tool to describe coupling loss factors:

$$\nu_i \eta_{ik} = \nu_k \eta_{ki} \quad (10.22)$$

With this concept we can formulate the energy injected into a subsystem by accounting for a source and all power input from other systems.

$$\Pi_{i,\text{input}} = \omega \eta_i E_i + \sum_k \omega \nu_i \eta_{ik} \left(\frac{E_i}{\nu_i} - \frac{E_k}{\nu_k} \right) \quad (10.23)$$

The equation of modal energies of all subsystems, coupling and power flow can thus be expressed in matrix form, such as

$$\mathbf{\Pi} = \mathbf{C}\mathbf{E}, \quad (10.24)$$

which for three subsystems is expanded in the form

$$\mathbf{C} = \begin{pmatrix} \omega \nu_1 \left(\eta_1 + \sum_{k \neq 1} \eta_{1k} \right) & -\omega \nu_1 \eta_{12} & -\omega \nu_1 \eta_{13} \\ -\omega \nu_2 \eta_{21} & \omega \nu_2 \left(\eta_2 + \sum_{k \neq 2} \eta_{2k} \right) & -\omega \nu_2 \eta_{23} \\ -\omega \nu_3 \eta_{31} & -\omega \nu_3 \eta_{32} & \omega \nu_3 \left(\eta_3 + \sum_{k \neq 3} \eta_{3k} \right) \end{pmatrix} \quad (10.25)$$

Typical applications in acoustic engineering involve hundreds or thousands of subsystems.

With the energy density or the sound intensity known, the sound pressure can be estimated on the basis of a specific sound field type. In diffuse field conditions in a room, for example, the squared sound pressure is related to the energy density by using Eq. (1.31).

10.2.3 Time Domain Models

Wave propagation can be simulated in the time domain as well. We do not start the discussion with harmonic signals (Dirac pulse in frequency domain) but with its “opposite”, the temporal Dirac pulse. Pulse propagation can now either be studied in a mesh from node to node (finite difference model) or on a large scale assuming

special wave types (ray tracing). The results are wavefronts propagating in time. Reflections, diffraction and other propagation effects can be calculated. At chosen field points, impulse responses can be obtained.

Finite Difference Time Domain (FDTD) Method

For studies of the propagation paths which have no need for exactly predicted propagation spectra, wavefront synthesis offers an insight into the field of travelling waves by using finite time differences and the finite difference time domain (FDTD) method. In such mesh-based time domain models, there is no a priori assumption of wave types. The waves developed on the basis of the mesh itself (waveguide mesh) provided the degrees of freedom of motion, and forces are integrated in the model equations. The mesh must be uniform.

The equations are the discretized acoustic field equations (Eqs. 1.11 and 1.12). This yields (here in one-dimensional form)

$$-\frac{\Delta p}{\Delta x} = \rho_0 \frac{\Delta v}{\Delta t}, \quad (10.26)$$

$$-\rho_0 \frac{\Delta v}{\Delta x} = \frac{1}{c^2} \frac{\Delta p}{\Delta t} + \rho_0 q. \quad (10.27)$$

The problem, however, is that the wave propagation, its speed and its interaction with boundaries can only be solved exactly only for 1D cases. Examples for 2D problems were presented as approximations, but FDTD in 3D domains suffers from severe artefacts like dispersion, unless corrections are introduced (Savioja and Välimäki 2000). The reason for this is the geometrically impossible condition of creating a perfectly uniform (isotropic) mesh. Furthermore, specific impedance boundary conditions other than ideally hard (or soft) surfaces cannot be implemented easily. The finite time difference method, however, is very useful for the illustration and animation of wavefront propagation.

Lam and Hargreaves (2012) discuss promising ways to further extend FDTD towards simulation in 3D including appropriate source models and impedance boundary conditions. They also compare the results with time-domain boundary elements models (see also Langer and Schanz 2008). FDTD was also optimized to an extent that real-time performance is on the horizon, at least for small model sizes. This is possible due to parallel processing and evaluating the audibility of dispersion errors (Saarelma and Savioja 2019).

Waveguides

The imagination of wave propagation in one dimension is represented perfectly by waveguides. They introduce propagation delay and attenuation due to divergence and damping. A spherical wave (omnidirectional) can also be described by

waveguides. Points of reflections due to interfaces between different impedances are also easy to be implemented by connecting waveguides (delay lines) with transfer functions of reflection and transmission factors. At this point it is obvious that waveguides contain waves travelling back and forth, thus being bidirectional.

In the case of 2D or 3D wave propagation, the waveguide model is mapped to a corresponding CAD model, like a room. The delay lines then are geometrically fixed at some point or patches at the walls. Otherwise the combinations of geometric paths would increase exponentially. With the geometric concentration of the waveguide nodes, a finite number of node connections results (Krämer 1995).

Waveguides are well described and studied in application to physical modelling of musical instruments and vocal tracts (Välimäki et al. 1993; Välimäki 1995; Fant 1970). In these cases the transmission system is separated into adjacent tubes of varying cross sections. The famous “Kelly-Lochbaum” model of the vocal tract explains the formation of formants in vowels.

Frequency-dependent losses are included by adding digital low-pass filters into the delay loop. FIR and IIR filter networks are used, partly involving sophisticated phase models and sub-sample shifts for better adjusting the actual geometric relations between nodes (Karjalainen 2005). The geometric conditions and the corresponding wave divergence must be specifically included, except for 1D applications like wind instruments and vocal tracts (Figs. 10.11, 10.12 and 10.13).

Fig. 10.11 Mapping room geometry on a set of coupled delay lines. (After Krämer 1994)

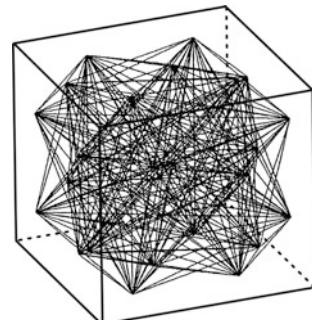


Fig. 10.12 Nodes coupling delay lines. (After Karjalainen et al. 2005)

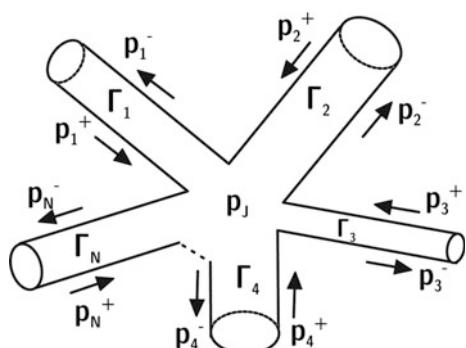
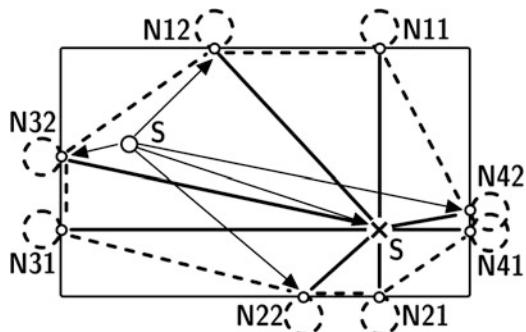


Fig. 10.13 Room model with waveguide network.
(After Karjalainen et al.
2005)



Geometrical Acoustics

This model, already introduced for calculating basic features of room acoustics in Sect. 4.4, can be applied for various other purposes. Geometrical acoustics is easily understood by using the analogy of geometrical optics. A laser beam representing a straight line carrying light energy is well known. We now interpret sound propagating as rays, too. Rays are reflected, refracted or diffracted (at least in first order). Rays carry sound energy, and the quantity hitting a receiver area or volume determines the received sound energy. Due to the fact that rays are not describing near-field wave effects, they are related to long distance approximation of quasi-plane waves. In this way, any field geometry, spherical or cylindrical wave fields, can be modelled by sending rays in an appropriate arrangement. Fields of application of geometrical acoustics are:

- Room acoustics
- Outdoor sound propagation
- Underwater acoustics
- Ultrasound

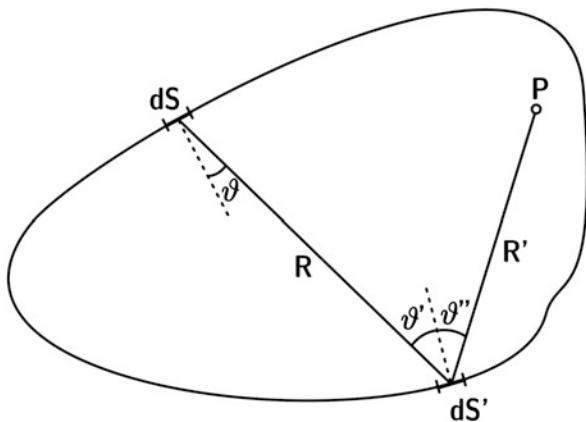
For auralization, room acoustics and noise emission prognosis are the most interesting fields. Accordingly, geometrical acoustics is well developed in these areas.

Ray construction is the key of geometrical acoustics. In elementary definition, rays will travel along the path with the shortest travel time (Fermat's principle). Depending on the medium, this might lead to straight lines and specular reflections. In layered media like in the atmosphere with height-dependent temperature or wind profile, the rays are bent by refraction. But as in geometrical optics and discussion of lenses, refraction can well be included into the model. The strategy to construct rays is twofold:

- Forward geometric construction from the source to the receiver
- Reverse construction from the receiver to the source.

Basically these two approaches are equivalent, which is inherently given by the law of reciprocity (source and receiver may be interchanged). But to some extent the

Fig. 10.14 Radiosity based on Kuttruff's integral equation. (After Kuttruff 1971)



approaches also show extremely diverging advantages and disadvantages, which are discussed in Chap 11.

When the ray propagation is known, impulse responses containing Dirac pulses of delays and energies can be constructed. Impulse responses may also contain specific impulse responses of edge diffraction or boundary reflection factors.

Radiosity

A geometric model, too, consisting of elements of energy radiation over distances and via observation angles is “radiosity”. The concept is composed of irradiation and re-radiation of energy from surface elements⁸. It is essential that the energy is diffusely scattered. With these prerequisites, the integral equation between the total irradiation strength received from all other surface elements can be solved. This method is also known in simulation of illumination on diffusely reflecting surfaces (Fig. 10.14).

According to (Kuttruff 1971), the energy portion irradiating the surface element dS' at R' from the other surface elements located at R is given by

$$B(R, t) = \frac{1}{\pi} \iint_S (1 - \alpha(R')) B\left(R', t - \frac{R}{c}\right) \frac{\cos \theta \cos \theta'}{R^2} dS'. \quad (10.28)$$

In this integral equation, Lambert's law of diffuse scattering is included. Note that exact timing is also included due to retardation by $t-R/c$.

The pure calculation of energy irradiation on the walls is interesting, but more relevant is the energy at field point R in the interior field space. This is achieved by

⁸Often called “patches”

$$w(R, t) = \frac{1}{\pi c} \iint_S (1 - \alpha(R')) B\left(R', t - \frac{R'}{c}\right) \frac{\cos \theta'}{R'^2} dS' + w_d(R, t) \quad (10.29)$$

w_d denotes the direct sound (see Eq. 4.40). In its discretized form, the integral (10.29) is the basis for acoustic radiosity. It is used for the calculation of the total energy density in the receiving point, R , via a direct path and reverberant paths under conditions of diffuse reflections. If not only the boundary (in patches) is discretized, but also the time, the temporal process of energy transition from wall to wall and to the receiver can be calculated (room impulse response).

10.3 Two-Port Models

Many acoustical/mechanical and electrical equations are formally similar, for example

$$U = L \frac{dI}{dt} \quad F = m \frac{dv}{dt} \quad I = C \frac{dU}{dt}. \quad (10.30)$$

Here, the notations are used for the voltage, U , the inductance, L , the current, I , the force, F , the mass, m , the velocity, v , and the capacitance, C . Hence it is worthwhile to use these analogies for solving mechanical or acoustical problems with techniques known in the analysis of electrical circuits. The power of the methods in analysing electrical circuits can be seen in the coupling of current and voltage in complex circuits and in the identification of major and minor current flow paths. The model, therefore, is excellently qualified to analyse force and velocity ratios over mechanical systems, force feedback and resonance systems. When comparing the equations including the concentrated elements of mechanical masses, springs and resistors and the corresponding equations of voltage, current and the typical electronic elements, two possibilities are found (Table 10.1):

Analogy I shows impedance conserving, and analogy II represents conserving the circuit plan. This means that an equivalent mechanical circuit developed from an electrical circuit will have the same general structure. In Fig. 10.15 some examples

Table 10.1 Electromechanical analogies

	Analogy I		Analogy II	
Voltage U	\leftrightarrow	Force F	\leftrightarrow	Current I
Current I	\leftrightarrow	Velocity v	\leftrightarrow	Voltage U
Electr. impedance Z_{el}	\leftrightarrow	Mech. impedance Z_m	\leftrightarrow	Electr. conductance Y_{el}
Resistance R	\leftrightarrow	Friction losses w	\leftrightarrow	Conductivity $1/R$
Inductance L	\leftrightarrow	Mass m	\leftrightarrow	Capacitance C
Capacitance C	\leftrightarrow	Spring n	\leftrightarrow	Inductance L

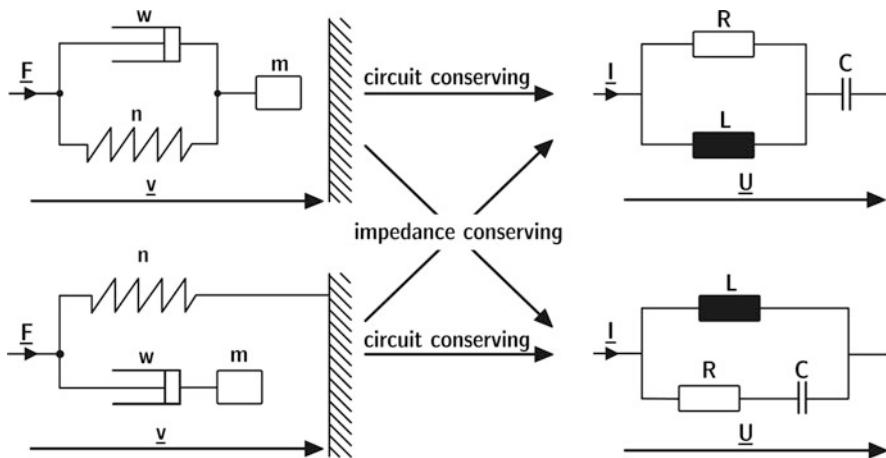


Fig. 10.15 Equivalent electromechanical circuits

are shown. It is worthy to be mentioned that electroacoustic transducers and, thus, coupled electrical-mechanical devices can be modelled by this as well.

Concentrated elements are valid in case of one-dimensional or multidimensional orthogonal signal transmission. In particular, masses, springs and losses should act as clearly separated elements. A mass, thus, should not show any internal spring or waveguide behaviour, for instance.

Waveguide elements are delay lines corresponding to a specific kind of wave propagation, mostly the plane wave.

From the definition of circuit elements, the step towards system modelling is quite easy. A so-called “two-port” is defined as a system with two input terminals and two output terminals. Between these terminals a voltage is defined as different in potential of the electric field. In language of mechanics and acoustics, this difference is to be interpreted as a force or a velocity, depending on the analogy used. The inside circuit hidden in the two-port is a priori unknown. Its circuit plan and the concentrated elements are not even necessary for modelling purposes. Instead, the transfer function and the transfer impedance serve as descriptors. Matrix formulations help, too, in integrating two ports in complex networks. Transfer functions, transfer impedances, matrix elements, etc. are spectral data which can be used in the coupling of two ports and in the simulation of networks. If the network represents a system acting between a source and a receiver, it can perfectly be used for calculation of auralization filters (Fig. 10.16).

Passive circuits are reciprocal. If they are fed by excitation signals from the left or from the right side, the ratio of the open-circuit forces (infinite mechanical load) and the velocities on opposite side

Fig. 10.16 Mechanical two port, analogy I (force/voltage)

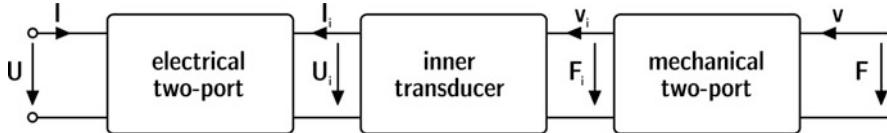


Fig. 10.17 Network representing an electromechanical transducer (analogy I)

$$\frac{F_{2,o}}{v_1} = \frac{F_{1,o}}{v_2}, \quad (10.31)$$

remains invariant. Infinite mechanical loads are achieved by blocking the motion (Fig. 10.17).

Similarly, electromechanical transducers can be modelled. They require circuits on the electrical and on the mechanical/acoustical side. The inner transducer effect (electrodynamic, electrostatic, piezoelectric, etc.) is modelled by a transformer or by a gyrator.

Here, reciprocity holds true as well, for instance, in the form of

$$\begin{aligned} \frac{F_{v=0}}{I} &= \pm \frac{U_{i=0}}{v} \\ \frac{v_{F=0}}{I} &= \frac{U_{i=0}}{F} = \mu \end{aligned} \quad (10.32)$$

with the upper sign denoting the rule for the electric field transducer (force/voltage transduction) and the lower for magnetic field transducers (force/current transduction). The left side represents an actuator (or loudspeaker) and the right side a sensor (microphone) with sensitivity μ .

Electroacoustic transducers like loudspeakers and microphones are modelled as mechanical transducers with division of the force by their membrane area, S . The force is thus turned into a pressure and the velocity into volume flow. This is exactly represented by a transformer with the ratio 1: S . The port with volume flow and pressure can finally be coupled with radiation impedances (Eq. 2.18), for instance (Fig. 10.18).

It is worth mentioning that the two-port model can be extended into multiports, if the paths are clearly separated. One very efficient way of separation in terms of linear combination is a modal basis. This way, if the fields are clearly defined in their modal contribution factors, even distributed fields can be used as input and output data.

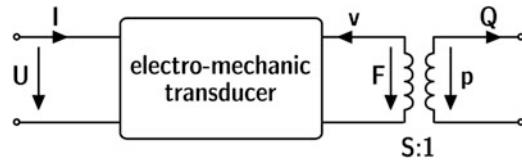


Fig. 10.18 Electroacoustical two port (analogy I)

10.3.1 Transfer Path Models

On the basis of the general two-port theory, complex networks can be established. They can be used for airborne sound propagation, vibration propagation, structure-borne sound radiation and auralization. The energy transmitted in the network is separated into paths, as illustrated in Fig. 10.19. Ideal or real force sources or velocity sources are connected directly to the two-port network. The difference between ideal and real sources is the impedance coupling between the source and the transmission system.

For the airborne sound paths, the source signals are coupled with transfer functions to the receiver (monaural or binaural). Modelling feedback is generally not required. Structure-borne sound paths should be modelled with feedback, and this requires network analysis in a first step and signal flow calculation in a second step.

As illustrated in Fig. 10.20, in the first step, the two-port network must be analysed with regard to the total impedance at the interface to the source. Only when this load impedance is known, the power output of the source can be calculated, in dependence on the inner impedance of the source. This general principle can be applied to simple one-dimensional networks or to complex matrix formulations.

The matrix \underline{A} (chain matrix) contains the mechanical impedances, admittances and other (dimensionless) parameters. The exact meaning of the matrix elements depends on the analogy used (see Sect. 10.3). The matrix can also be re-arranged into a pure impedance matrix, \underline{Z} :

$$\begin{pmatrix} F_1 \\ F_2 \end{pmatrix} = \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \underline{Z} \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \quad (10.33)$$

The two-port parameters Z_{11} and Z_{22} are called input and output impedance, respectively, while Z_{12} and Z_{21} are called transfer impedances. \underline{A} can be easily transformed into \underline{Z} . For reciprocal networks holds true $\det \underline{A} = 1$. The task in transfer path analysis is the determination of the matrix elements; see also Chap. 15 (Table 10.2).

The prerequisite of concentrated elements should be discussed a little more in detail. In case of large construction elements like dampers or springs, the elastic and viscous effects cannot always be separated into concentrated elements. Also at small wavelengths, they act as a waveguide of propagating or standing waves. As in

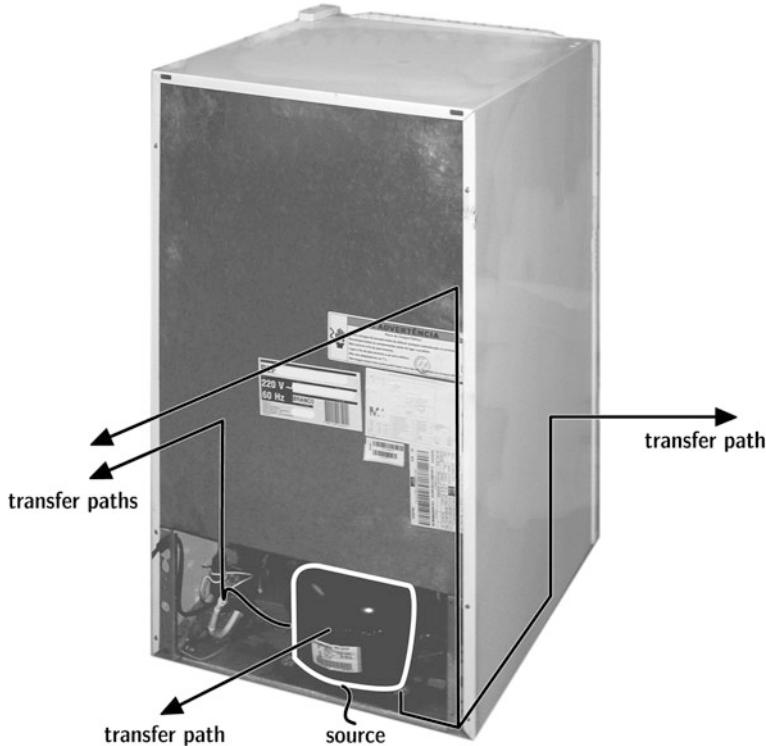


Fig. 10.19 Transfer path separation. Illustration with the example of a refrigerator

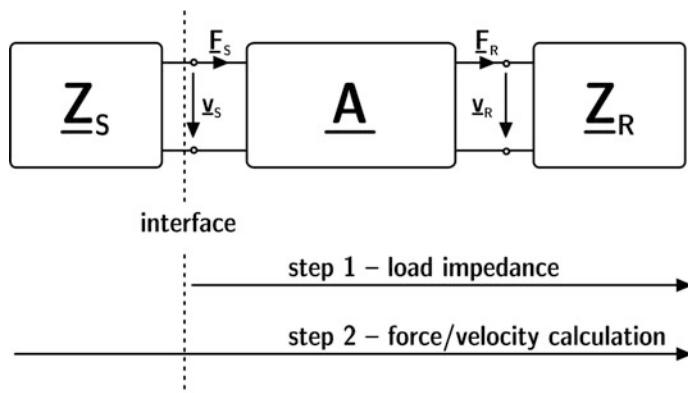
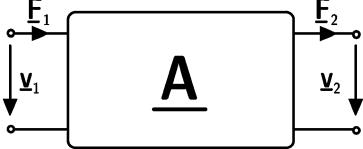
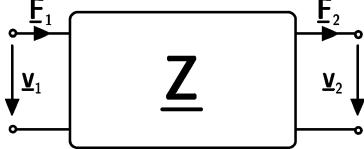


Fig. 10.20 Two-port network of source impedance \underline{Z}_S , transfer matrix \underline{A} and receiver impedance, \underline{Z}_R in analogy II (force/current)

Table 10.2 Two-port matrix conversion

Chain form	Impedance form
	
$\begin{pmatrix} \underline{E}_1 \\ \underline{v}_1 \end{pmatrix} = \begin{pmatrix} \underline{A}_{11} & \underline{A}_{12} \\ \underline{A}_{21} & \underline{A}\underline{Z}_{22} \end{pmatrix} \begin{pmatrix} \underline{E}_2 \\ \underline{v}_2 \end{pmatrix} = \underline{\mathbf{A}} \cdot \begin{pmatrix} \underline{E}_2 \\ \underline{v}_2 \end{pmatrix}$	$\begin{pmatrix} \underline{E}_1 \\ \underline{F}_2 \end{pmatrix} = \begin{pmatrix} \underline{Z}_{11} & \underline{Z}_{12} \\ \underline{Z}_{21} & \underline{Z}_{22} \end{pmatrix} \begin{pmatrix} \underline{v}_1 \\ \underline{v}_2 \end{pmatrix} = \underline{\mathbf{Z}} \cdot \begin{pmatrix} \underline{v}_1 \\ \underline{v}_2 \end{pmatrix}$
$\begin{pmatrix} \underline{A}_{11} & \underline{A}_{12} \\ \underline{A}_{21} & \underline{A}_{22} \end{pmatrix}$	$\underline{\mathbf{A}} = \begin{pmatrix} \frac{\underline{Z}_{11}}{\underline{Z}_{21}} & \frac{\det \underline{\mathbf{Z}}}{\underline{Z}_{21}} \\ \frac{1}{\underline{Z}_{21}} & \frac{\underline{Z}_{22}}{\underline{Z}_{21}} \end{pmatrix}$
$\underline{\mathbf{Z}} = \begin{pmatrix} \frac{\underline{A}_{11}}{\underline{A}_{21}} & \frac{\det \underline{\mathbf{A}}}{\underline{A}_{21}} \\ \frac{1}{\underline{A}_{21}} & \frac{\underline{A}_{22}}{\underline{A}_{21}} \end{pmatrix}$	$\begin{pmatrix} \underline{Z}_{11} & \underline{Z}_{12} \\ \underline{Z}_{21} & \underline{Z}_{22} \end{pmatrix}$

After Dohm (2004)

modelling of electromagnetic waveguides, they can be modelled as a quasi-continuous network of parallel circuits, each representing small interval Δx which is appropriate for one-dimensional wave propagation; see also 10.2.3, waveguides.

10.4 Other Models

Not all models for simulating sound and vibration can be described in detail here. If the sources and the sound propagation need not to be separated or if stimuli are to be created on the basis of an experimental approach mixed with a technical parameter model, sound synthesis is a possible tool.

Sound synthesis can be an adequate model to create mixtures of tones and noises with specific harmonic, stochastic, and temporal content for subjective testing. The approach is similar to transfer path models with the difference that not transfer paths but signal content is separated (in the analysis) and recombined in the synthesis. This model can be applied in noise control as well as in the modelling of musical sounds.

Chapter 11

Simulation of Sound in Rooms



11.1 Introduction

In this chapter, we will discuss the fundamental algorithms of geometrical acoustics, their variations of implementation in software and their efficiency. It will be shown that pure specular models (image models) are not capable of simulating room sound fields sufficiently accurate. But combinations of models of image sources and statistical models such as Ray Tracing and radiosity will allow a simulation with acceptable plausibility. Hybrid models which can handle specular and diffuse reflections for the estimation of the late reverberation spectrum are the solution to obtain impulse responses very near to measurement results. For inclusion of wave effects, hybrid models combining numerical methods (FEM, BEM, etc.) with geometrical acoustics have been introduced.

11.2 General

Computer modelling of room acoustics was first proposed during the 1960s by Schroeder et al. (1962) and used in practice by Krokstad et al. (1968) and Schroeder (1973). Although room-acoustic scale model experiments are a powerful tool still today, computer simulations are increasingly taking over the part of scale models in consulting. Commercial software became more user-friendly, more accurate and, last but not least, cheaper than scale models. As soon as the architectural plan is transferred into a computer file and the wall data, and the source and receiver locations are defined, the sound propagation in the room can be simulated quite fast and modifications can be tested without large effort.

The algorithms of typical programmes are based on geometrical acoustics (Sect. 10.2.3). In geometrical acoustics, the description of the sound field is reduced to energy, transition time and direction of rays. This approach is correct as long as the

Table 11.1 Basic algorithms of room acoustics computer simulations

Algorithm	Category	Energy spreading by distance	Energy detectors
Ray tracing	Stochastic	Stochastic by counting	Volumes
Image sources	Deterministic	Deterministic by distance	Points

dimensions of the room are large compared to the wavelengths and as long as broadband signals are considered. These approximations are valid with sufficient accuracy in large rooms above cut-off frequency f_c , see Eq. (4.13).

In intercomparison tests (so-called round robin tests) (Vorländer 1995; Bork 2000a, b), the efficiency and the limits of room acoustics computer simulation (see 11.6) were evaluated. Finally, auralization has been included in room acoustics simulation since the beginning of the 1990s (see Sect. 11.7).

The methods introduced in this chapter are, at first, used to calculate the room acoustic criteria (T , EDT, D , C , T_s , LF, IACC...), see Sect. 6.5, to get a more specific result than given by the estimation (Sect. 6.5.6).

Two techniques of geometrical acoustics have to be distinguished: “ray tracing” and “image sources”. Independent of their software implementation, they represent different physical approaches. In order to achieve a clear understanding, this point is to be stressed, since ray tracing algorithms can well be used to calculate and handle image sources. The main difference between ray tracing (in classical definition) and image sources is the way how energy detection and the internal nature of the physical energy propagation are implemented (Table 11.1).

11.2.1 CAD Room Model

For implementation of the algorithm described in the software, the room geometry, the sound sources and the receivers must be defined as mathematical objects. We start the discussion with the question of how to create a CAD model of the test room. The complexity of the model may be very high, particularly when it was exported from software for architectural design. These models include small details like staircase steps or even smaller geometric features.

From an acoustic point of view, however, this kind of model would not only be too large in memory size and polygon complexity, it would also be wrong. As will be shown below, the acoustic characterization of surfaces is based on absorption and scattering. And the physical principles of wave reflection and scattering are clearly defined (see 3.4). Objects or surface corrugations which are not large compared to the wavelengths have to be taken out the CAD model and replaced by a flat surface with adequate acoustic properties. This holds true also for chairs and audience seats, for lamps and for doorknobs. For the purpose of visualization, these elements are essential for a realistic impression. For the “acoustic view” they are invisible or at most visible with some diffuse halo (Fig. 11.1).

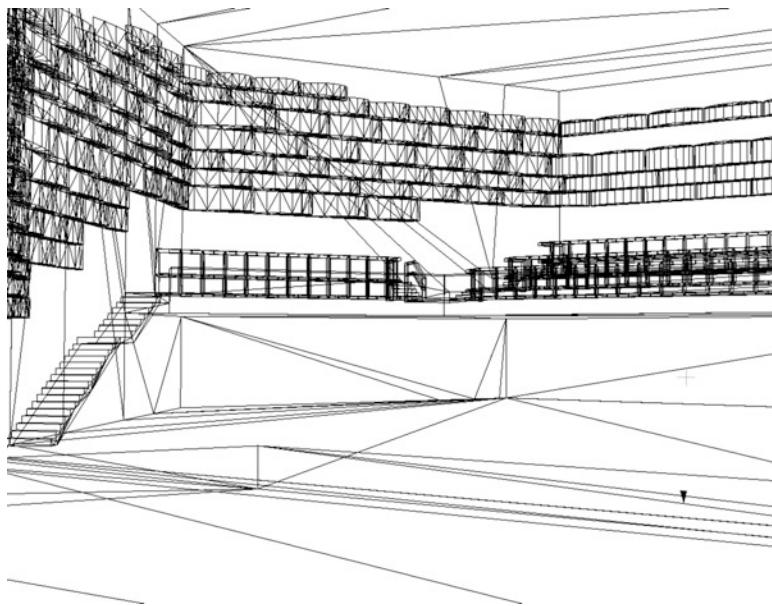


Fig. 11.1 Visual wireframe model of a multipurpose hall

As a rule of thumb, the guideline might be used to draw the room surfaces and the interior with a resolution of 0.5 m, representing a wavelength scale corresponding to a frequency of about 700 Hz. Below that frequency, the surfaces should be modelled as flat specularly reflecting polygons, above as partly scattering surfaces or objects (Fig. 11.2). Sound scattering dominates at high frequencies anyway, except for very large walls or the ceiling (>10 m). The only reason for including a small resolution of details (<0.5 m) would be a study of a frequency range with small wavelengths compared to 0.5 m, say, for $\lambda = 5$ cm. Then, frequencies around 7 kHz and above are discussed. For several reasons this would only be of interest in very special cases. The main arguments for choosing a range of lower frequencies are:

1. The spectral content excited by natural sources (voice or instruments) is small above 7 kHz.
2. In broadband signal situations, masking will not allow humans to identify details of low levels at high frequencies.
3. The accuracy of the simulation model is not sufficiently high to guarantee results within an acceptable confidence limit.

Not much research has yet been done on automatic simplification of CAD models from detailed resolution towards a specific acoustically relevant resolution. Polygon smoothing by using spatial low-pass filters with a cut-off wave number of 125 m^{-1} (equivalent to a minimum length resolution of 50 cm) could be an interesting option for future automated CAD user interfaces.

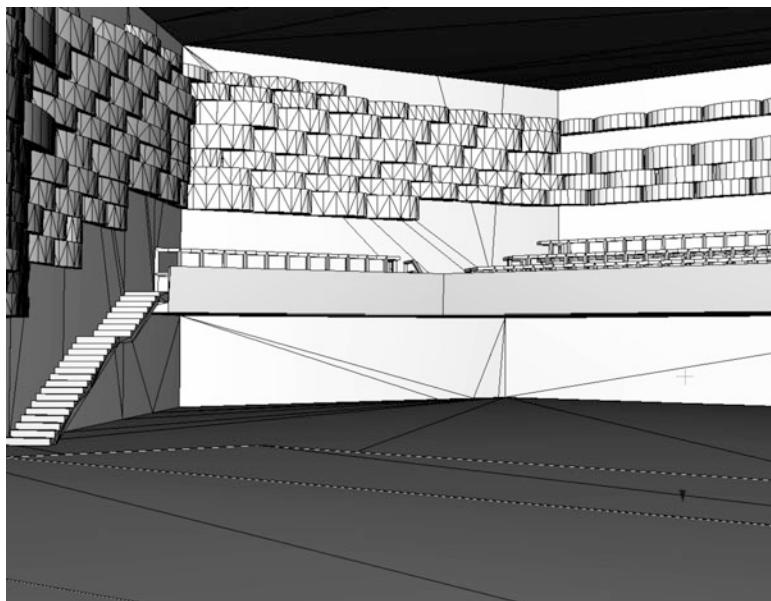


Fig. 11.2 Visual surface model of a multipurpose hall

In the end, the room must be approximated by analytic surfaces, typically by planes. Surfaces of higher order (cylindrical, spherical, parabolic, etc.) can be used in principle (see below). All surfaces lying in a common plane are forming a “wall”. A wall, thus, needs not to be simply connected. Walls can also be made up of more than one surface material. One very effective simplification, however, is the requirement for convex wall shapes, as it simplifies and thus accelerates time-consuming tests, such as the point-in-polygon test which is a frequently called function in all kinds of room acoustics simulation algorithms. This rule is no practical constraint in modeling a room shape at all (Fig. 11.3).

All plane surfaces are defined by three points in a previously fixed coordinate system. Preferably, points (called “vertices¹”) of the wall polygon are used. To define the polygon completely, the other vertices of the polygon must be added, too. In practice, geometric uncertainties in adding the fourth, fifth, etc. vertices in the polygon’s plane can be allowed, if the new corners are automatically adjusted to match the plane (at least numerically) exactly (Fig. 11.4).

Figure 11.5 illustrates the strategy to build a 3D room from 2D polygons. The polygon-to-vertex notation, the polygon orientation with reference to the normal vector and their address management must be specified with great care in order to avoid errors caused by open faces or gaps.

¹From “vertex”.

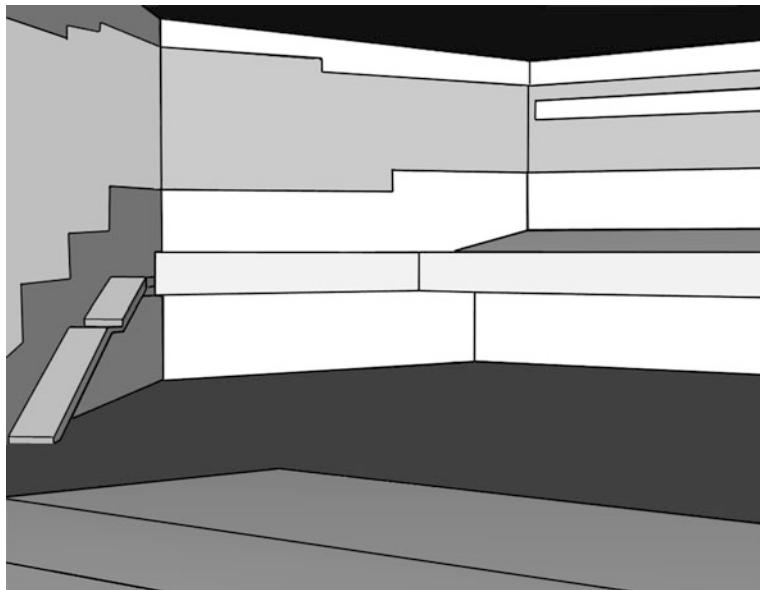


Fig. 11.3 From architectural to acoustic CAD models: Acoustic surface model of a multipurpose hall. Geometric details are replaced by smooth wall polygons

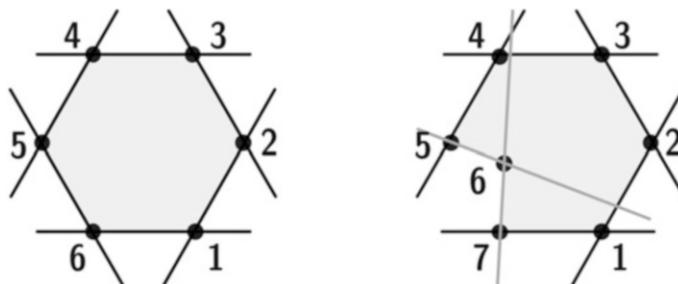


Fig. 11.4 Left: convex polygons, right: concave polygons

11.2.2 Absorption Coefficients

Absorption coefficients can be found in tables (see Annex) or they can be taken from specific test results. Most absorption coefficients correspond to diffuse sound incidence, as they were obtained from measurements in reverberation chambers (ISO354). In ray tracing, this data represents the average absorption at random incidence. Thus, in rooms which provide an approximate diffuse field, they serve well, particularly in the late response where in the region of high-order reflections the average absorption coefficient is an appropriate quantity. The energy losses of the first reflections at their specific angles of incidence, however, are not modelled

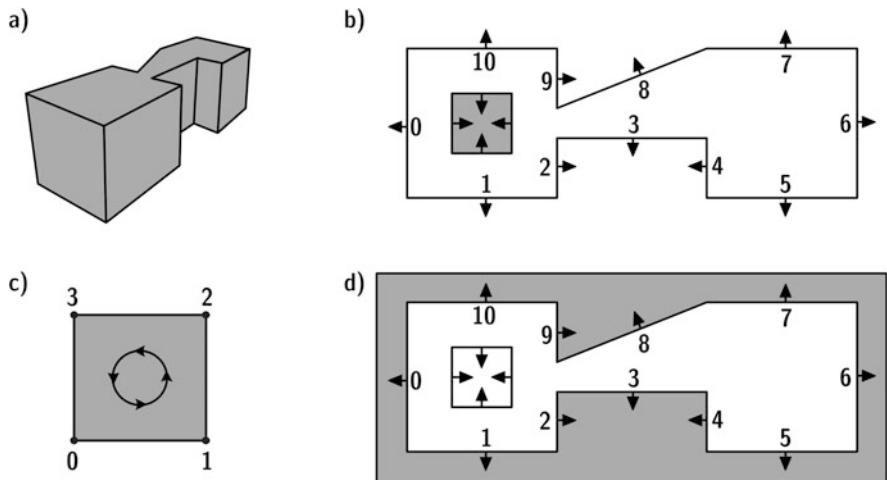


Fig. 11.5 (a) Example: a simple room constructed by a set of convex polygons. (b) The room’s 2D projection: the green area marks a solid object inside the room. (c) Definition of the spanning direction for polygons. (d) The room’s 2D projection: the green area marks the outside world. (Example after Schröder 2011; Schröder and Lenz 2006)

precisely. Nevertheless the errors are small. It should be kept in mind that the difference between energy loss at a specific angle of incidence and random incidence may be up to 40%, until the error of the reflection level exceeds 2 dB. Only for grazing incidences, the errors may become larger. But in grazing incidences other influences come into play anyway (see Sect. 11.6).

If the complex wall impedance is known, the angle-dependent absorption coefficient can be calculated. This is easy for the case of real-impedance locally reacting walls (see Eqs. 3.3, 3.4, 3.5 and 3.6). In each case of room simulation, it should be checked if random-incidence absorption coefficients are applicable or if angle-dependent data are possibly required. The latter may, for example, apply to long or flat rooms, tunnels, corridors, and coupled rooms.

11.2.3 Scattering Coefficients

If absorption is treated in a simplified way (assuming random incidence), this simplification is even more justified for surface scattering. Nevertheless angle-dependent scattering is interesting for discussing low-order reflections. The total field in front of a diffuser, depending on the signal’s spectrum and coherence length, is very complicated (Sect. 3.4). Data from free-field measurements (part 2 of (ISO17497)) or calculations can be obtained by using standardized methods, but these are valid only for harmonic signals.

Therefore, with the same argument as used above, we assume random-incidence appropriate, and we accept the unavoidable uncertainties which are small in the case of high-order reflections. But we must be aware of the problems which can arise at first- or second-order reflections.

Random-incidence scattering coefficients are not available to the same extent as absorption data. However, with the measurement method recently standardized in (ISO17497), more information may be published in the future (see Annex).

11.3 Stochastic Ray Tracing

In this model sound is radiated as a bunch of particles – many particles (= rays). The number of particles is one of the crucial features of ray tracing. It is considered to be in the category of Monte-Carlo methods, to express the influence of probabilistic effects, as, for example, in gambling (Fig. 11.6).

Stochastic Ray Tracing in Short

The sound source radiates an impulse. To simulate this, the particles are started at the same initial time in various directions. Each ray carries certain energy, propagates at speed of sound and once in a while hits the room boundary, for simplicity called “wall” throughout. From the wall it is reflected, hits another wall and so on. Each wall absorption reduces the particle’s energy (Fig. 11.7).

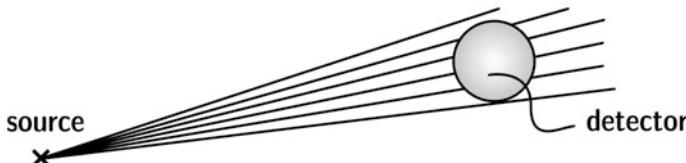


Fig. 11.6 Free-field propagation and distance law “by counting”

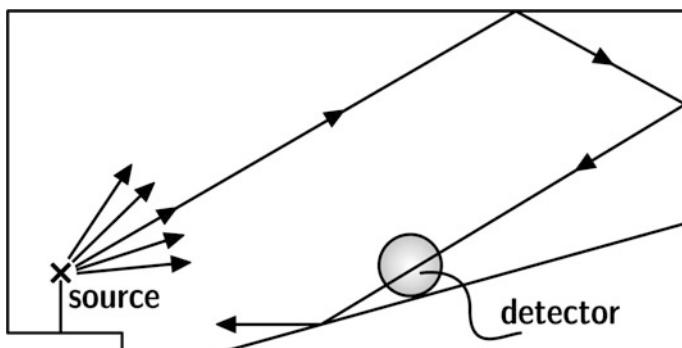


Fig. 11.7 Tracing a ray from the source to the detector

Fig. 11.8 Creating an impulse response by counting events

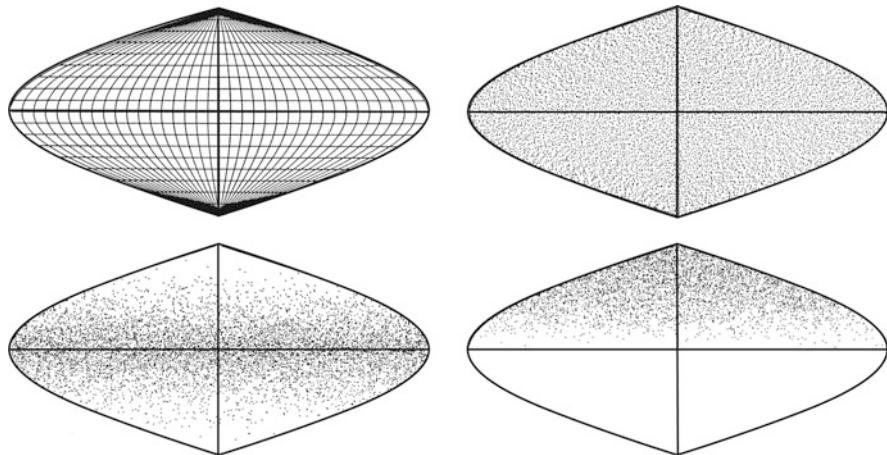
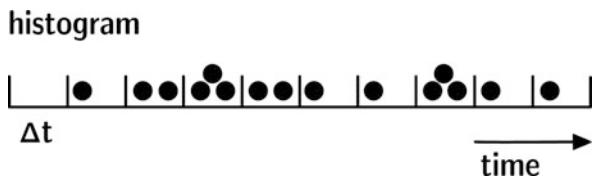


Fig. 11.9 Variants of ray sources, spherical coordinate system and regular grid (top left), uniform random distribution (top right), concentration of rays at the equator (line-arry-like) (bottom left) and concentration of rays at the north pole (baffled-piston-like) (bottom right)

By means of particle detectors, the particle's energy and the time elapsed since radiation from the source is registered (see Fig. 11.8). The number of counts represents the energy detected at the receiver point.

Source Modelling

Sound sources are characterized by sound power and directivity (Sect. 2.5). Both depend on frequency. For ray tracing we just need the source position and the reference (on-axis) direction. The directivity can be modelled by choosing direction-specific start energies of the particles, or by variation of the density of the particles, as illustrated in Fig. 11.9 in an area-accurate plot of the unit sphere.

To create a spherically uniform radiation, we proceed as follows: The azimuth angles are distributed evenly around the perimeter. The distribution of polar angles is to be chosen in a way that the ray density is constant. This is achieved by using a random number $z \in (0,1)$ (or by dividing the interval $(0,1)$ into fixed steps) and taking the polar angle of radiation as

$$\theta = \begin{cases} \arccos z \\ \arccos z + \pi/2 \end{cases}, \quad (11.1)$$

the upper choice for the northern and the lower for the southern hemisphere. $\theta = 0$ at the north pole. Other weighting functions create polar distributions with concentration of rays at the north pole (reference axis) or at the equator.

Any measured or calculated directional pattern can also be used as a spherical weighting function of the ray energy or of the ray density.

Flow Diagram

Due to the frequency dependence of absorption, ray tracing must be repeated for the frequency bands of interest, usually octave bands or one-third octave bands.² Air attenuation can also be modelled by reducing the ray's energy according to the flight distance and the attenuation coefficient.

We will now focus on ray tracing algorithms in more detail. There are two options for modelling absorption. It is obvious the energy can be reduced by multiplying the incident energy with a factor $(1-\alpha)$. The alternative is to apply stochastic annihilation of particles. Both algorithms yield identical results in the limit of large numbers. But they differ in computation time and in accuracy.

In absorption by multiplication, the particle starts with energy e_0 and is traced until a maximum travel time t_{\max} , or until a minimum energy, e_{\min} , is reached.

In the method of absorption by annihilation, a random number $z \in (0,1)$ is compared with the absorption coefficient, α . If $z < \alpha$, the particle is annihilated, and the next particle is traced; see flow diagram in Fig. 11.10. As long as there is absorption somewhere in the room, we do not need to fix another truncation criterion, like t_{\max} or e_{\min} .

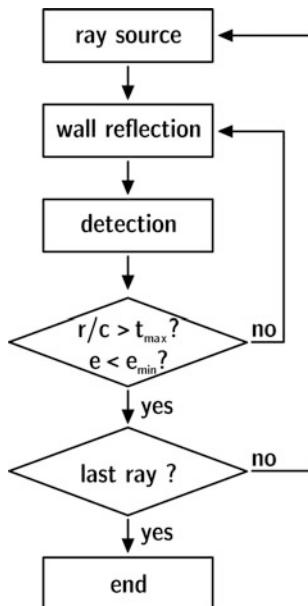
11.3.1 Point-in-Polygon Test

The innermost loop of ray tracing algorithms is a test for (a) whether a ray (particle), represented as line segment, is hitting a plane and (b) whether the intersection point is inside or outside a wall polygon. At this step, the actual straight line and the direction of propagation, thus a vector, is known. Also known is the set of polygons.

The polygons qualified (those located in the forward direction of the ray) are transformed in 2D coordinate systems of the polygon planes and tested by calculating vector products of the vectors from the plane intersection point to vertices of the polygon. For all neighbouring vertices, the vector product must have the same

²Under special circumstances parallel processing of frequency bands is possible; see below.

Fig. 11.10 Flow diagram of ray tracing



direction. Otherwise the intersection point is outside the polygon. An advantage of this algorithm is that it can be cancelled if the vector is oriented in opposite direction.

11.3.2 Reflection

In the case of the intersection point being inside the polygon, the new travel line will be the input variable for the next plane hit. Before reflection takes place, wall materials affect sound by absorption, scattering and diffraction. If a ray/particle hits a wall, it will lose energy or be annihilated,³ and it will change the travel direction.

Specular Reflection

At specular reflection, a new direction at a specular angle with reference to normal incidence will be calculated. The new ray vector is in the plane spanned by the incident ray and the normal vector of the polygon.

Classical Surface Scattering

When scattering occurs, which applies to the case that a random number exceeds the scattering coefficient, the new direction is obtained from two more random numbers.

³Depending on the way absorption is implemented.

Mostly, a directional distribution according to Lambert's law is assumed. On average, the polar angles of the new flight direction (with reference to the wall normal vector) must be distributed according to

$$w(\theta)d\Omega = \frac{1}{\pi} \cos \theta d\Omega. \quad (11.2)$$

The azimuth angle is evenly distributed in $(0, 2\pi)$. Hence, two random numbers z_1 and z_2 in $(0,1)$ and transformations

$$\theta = \arccos \sqrt{z_1} \quad (11.3)$$

and

$$\phi = 2\pi z_2 \quad (11.4)$$

yield a polar scattering distribution according to Lambert's law. Superposition of the specular direction vector and the scattering direction vector is also a possible solution to define the new direction.

11.3.3 Detectors

Surface or volume detectors are possible options. While surface detectors have an angle-dependent cross section, spherical detectors are independent of the angle of incidence. The detection cross section of spheres is

$$M_{\text{sphere}} = \pi r_d^2 \quad (11.5)$$

with r_d denoting the detector radius. Thus, the detector position (centre point) is a sufficient information. Each line segment representing the ray has to be checked whether it "hits" the detector. This test requires a calculation of the distance between the detector centre point and the ray vector. It is solved by finding the perpendicular line between the ray and the detector and by checking if this distance is smaller than the detector radius, r_d .

This way of ray detection represents an omnidirectional microphone. It does not represent a live audience. All physical effects of audience areas known in room acoustics are neglected. The most prominent effects are the seat-dip effect and the forward scattering at the audience's heads. It was tried to include both (wave) effects in models of detectors, but all attempts suffer from inconsistencies with geometrical acoustics. Anyway, audience effects can be modelled as specific feature of audience areas, independent of detector modelling. Complex impedance and diffraction

models are options for including audience effects in future simulation models, however, not in stochastic ray tracing models, but in deterministic models (see below).

For inclusion of spatial information, the receiver can be modelled as a binaural receiver (see 6.4.1) or as an equivalent to a microphone array (see 2.6). Depending on the choice of the spatial encoding scheme, the incoming sound field represented by the rays can be transformed into spherical harmonics or filtered in two channels for the left and the right ear by using HRTF. With the array approach, the binaural format can also be introduced later in post-processing, which is more flexible as concerns usage of individual HRTF. For this, the spherical harmonic representation of the rays can be filtered with the HRTF of interest, or the sound field can be decomposed into plane waves and then filtered with HRTF.

11.3.4 Linking Scattering and Detectors: “Diffuse Rain”

In stochastic ray tracing, the simulation results suffer from the discrete representation of sound by using a finite number of rays, as discussed in 11.3.7. The concept of the “diffuse rain” is beneficial for reducing the stochastic uncertainties by replacing the discrete ray tracing between a diffusely scattering wall and the receiver. The energy reflected at this point is calculated analytically by using the wall’s absorption and scattering coefficients, the distribution of scattered energy, following a Lambert distribution, for example, and the detector’s effective cross section, a purely geometrical quantity.

The effective cross section is determined by integrating the respective probability density function, which is formed by the contour of the receiver projected onto a sphere around the intersection point. γ is the opening angle, and r the distance from the ray’s intersection point to the centre of the detector (Fig. 11.11).

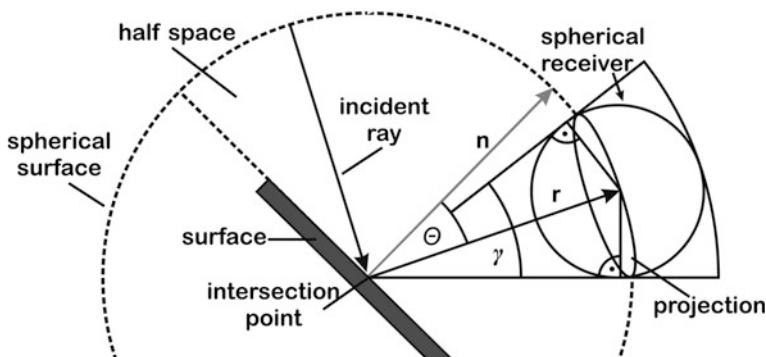


Fig. 11.11 Diffuse rain. (After Schröder 2011)

The crucial problem is to find a sufficiently precise estimate for arbitrary solid angles. A good approximation is to evaluate this distribution only for a fixed angle, θ , which is the angle between the surface normal and the connection vector between the intersection point and the receiver. The effective cross section is then combined from the (Lambert) energy radiation and the geometric cross section.

$$w_{\text{total}} = w_{\text{Lambert}} \cdot w_{\text{geo}} \quad (11.6)$$

For spherical detectors, the scattered energy is after Schröder (2011) referring to Heinz (1993)

$$e_s = e_{\text{ray}} \cdot s(1 - \alpha) \left(1 - \cos \frac{\gamma}{2}\right) 2 \cos \theta \quad (11.7)$$

e_{ray} denotes the energy of the incident ray, and α and s the absorption and scattering coefficients of the wall, respectively. Similarly, the diffuse rain processing can be performed for plane or cylindrical receivers.

11.3.5 Presentation of Results

Up to this point ray tracing was described in an abstract way. By discussing typical results, we can better interpret the quality, the efficiency and the benefit of ray tracing.

At each detection, the ray's energy, arrival time and direction of incidence are stored in an array called histogram of the energy impulse response (Fig. 11.12). The array must have a sampling rate which allows a sufficiently high temporal resolution, on the one hand, and sufficiently large integration intervals, on the other (see below, 11.3.7). A good compromise is a resolution in the order of milliseconds psychoacoustic post processing (Sect. 6.5) can be taken into account properly (Fig. 11.13), but binaural attributes like ITD are uncertain⁴. At the same time it is clear that we cannot meet the requirement for a sampling rate adequate for audio signal processing and auralization.⁵

One could argue that auralization with stochastic ray tracing could be possible with faster computers. This, however, is not a reasonable goal since ray tracing is a rough approximation of wave propagation. Any increase of accuracy and temporal resolution would only mimic a real gain of physical accuracy. Ray tracing still remains an approximation of wave propagation.

⁴It will be shown in Sect. 11.3.7 that the required number of rays would be unacceptably high.

⁵The temporal resolution for a sampling frequency of 44.1 kHz is 22.7 μ s.

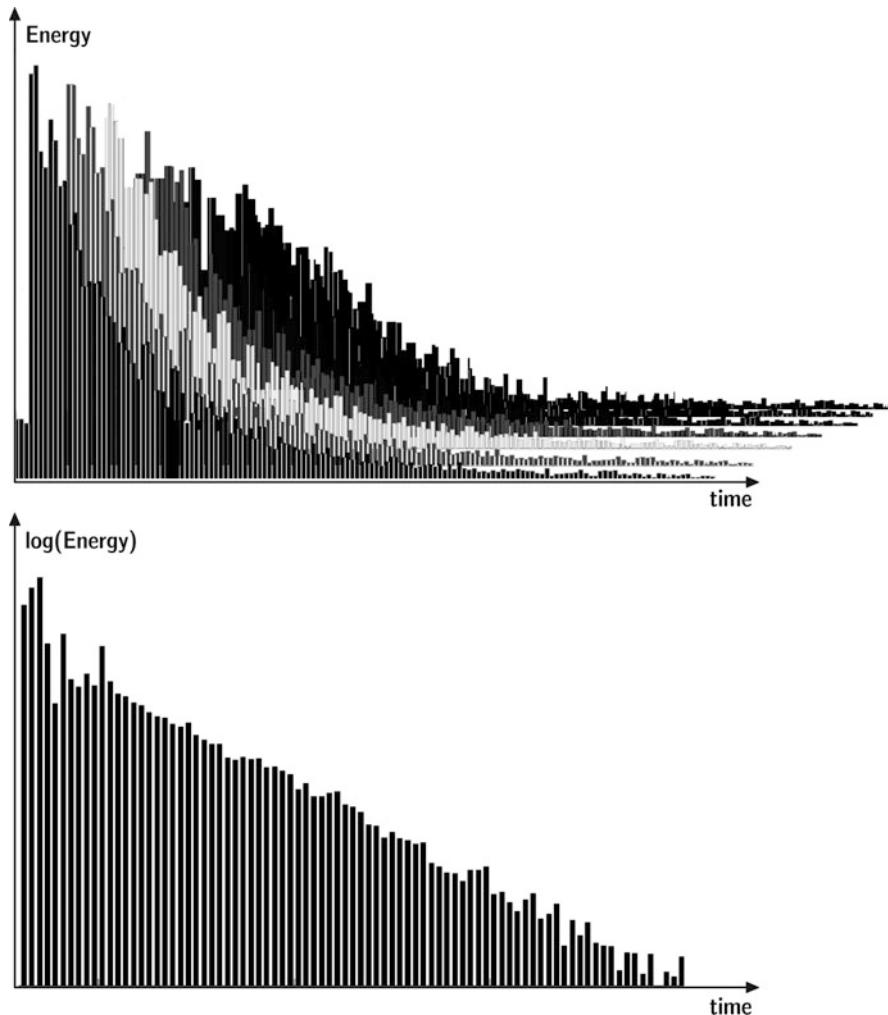


Fig. 11.12 Ray tracing energy impulse response (top, linear plot for several frequency bands; bottom, logarithmic plot for one frequency band)

11.3.6 Curved Surfaces

Inappropriate architectural room designs may lead to severe acoustic problems. Curved walls belong to the category of high-risk potential. Strong spatial and temporal concentrations of sound (= “echoes”) must be avoided, since sound focusing is against all goals of room acoustic design to achieve the best possible

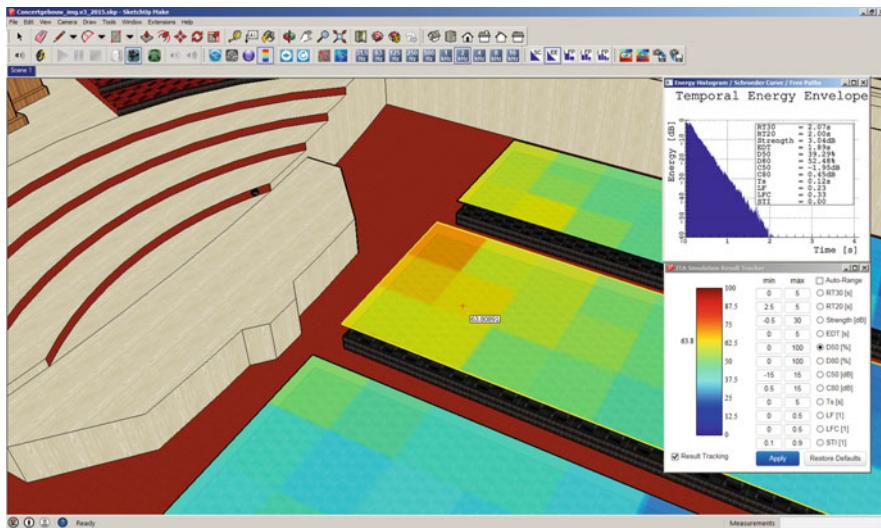


Fig. 11.13 Examples for room acoustic parameters plotted in the CAD model (audience area); see Sect. 6.5 and particularly 6.5.6

uniform distribution of good acoustic quality. In rooms with cylindrical side walls or with dome-shape ceilings, we have to expect focusing. If a room acoustic computer simulation is to be a tool for design, it must be ensured that occurrences of focusing and echoes are clearly visible in the results (Fig. 11.14).

An exact model of the room geometry by using curved surfaces may be even faster than the approximation by planes. To obtain accurate results with ray tracing, the detectors must be small and the number of rays large. It was shown by Kuttruff (1993) and later by Vercammen (2010) that deterministic approaches with coherent wave methods can be used. This is obvious since the sound pressure in the focal region cannot be obtained by energetic models. Kuttruff pointed out that the approximation of a cylinder with radius a by planes of width b is sufficient if

$$f b^2 = \frac{1}{2} c a \quad (11.8)$$

is fulfilled. It is interesting that this equation is frequency dependent. Thus, the condition must be checked for the highest frequency band involved, or the shape must be modelled to be frequency-dependent. For example, a cylindrical shape with radius 10 m must be modelled by panels of a width of 50 cm, if the frequency range should reach up to 8 kHz, which means that the cylinder should be subdivided into 136 plane elements.

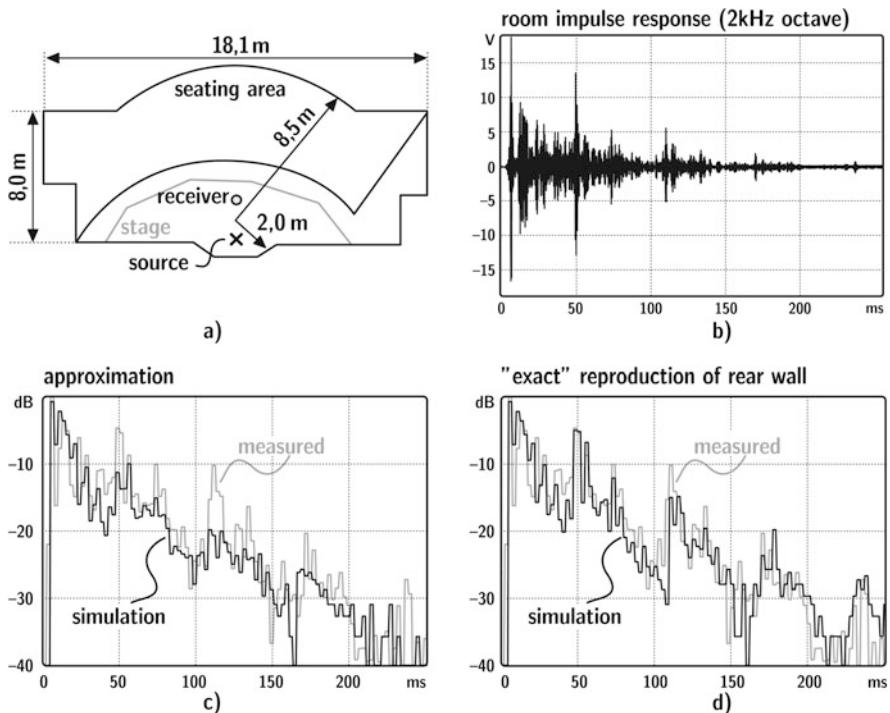


Fig. 11.14 (a) Ground plan of a small theatre (900 m^3) with source and receiver positions, (b) measured impulse response (2 kHz octave) and comparison with simulated impulse responses, (c) approximation of 5 plane surfaces, (d) mathematically exact model of the cylindrical back wall. (After Mommertz 1996)

11.3.7 Reproducibility in Stochastic Ray Tracing

In all physical measurement and simulation methods, systematic uncertainties are present. The approximations inherent in the input data (room geometry, absorption and scattering coefficients) and in the model itself (approximation of wave physics by a geometrical model) are causes for systematic errors. These errors cannot be reduced by an increased computational effort.

Independently, in Monte-Carlo methods like ray tracing, we must consider stochastic fluctuations in the results. By increasing the number of rays or by averaging repeated simulations, these errors can be reduced. The number of launched rays, N , is of crucial importance, because it affects both the reproducibility and the computation time t_{calc} .

The goal of this chapter is the prediction of the standard deviation in the simulation results in dependence on N and t_{calc} . For obtaining an expectation value

of the error, we need to remember that the energy impulse response was created by counting events of particle detection. The number of hits, k , in a detector sphere with radius r_d , in a time interval, Δt , is of particular interest, as well as the detection rate, r , of hits per second.

The probability of a ray hitting the detector within Δt is small but constant in time, if the number of rays in the room throughout the simulation is constant. The expectation value $\langle k \rangle$ is thus constant and follows a Poisson distribution:

$$P(k) = \frac{\langle k \rangle}{k!} e^{-\langle k \rangle} \quad (11.9)$$

The variance of the Poisson distribution, σ^2 , describes the fluctuations around the expectation value. It is

$$\sigma_k^2 = k = r\Delta t \quad (11.10)$$

The number of hits is thus related to the following expected error:

$$\frac{\sigma_k}{\langle k \rangle} = \frac{1}{\sqrt{\langle k \rangle}}. \quad (11.11)$$

The expectation value of the count rate $\langle r \rangle$ is calculated from the number of wall hits per second, $\bar{n}N$, and the ratio of the detector surface, $4\pi r_d^2$, and the room surface, S :

$$\langle r \rangle = \bar{n}N \frac{4\pi r_d^2}{S} \quad (11.12)$$

Herein, \bar{n} is the mean reflection rate according to Eq. (4.22). The mean count rate is thus

$$\langle r \rangle = N \frac{\pi r_d^2 c}{V}, \quad (11.13)$$

and the mean number of hits is

$$k = N \frac{\pi r_d^2 c \Delta t}{V}. \quad (11.14)$$

Now we can calculate the expectation value of the energy, $\langle E \rangle$, per time interval Δt . This can be obtained only when some features of the sound field are assumed. The diffuse sound field is chosen, as the best guess for a room sound field. This approach is, by the way, exactly the same as in statistical reverberation

theory (Sect. 4.5). Of course, an exponential decaying $\langle E \rangle$ will be found, which actually better matches the late decay in a real room. Thus, early reflections are not discussed here. The theory is identical with that discussed in Sect. 4.5. Therefore,

$$E(t) = e_0 N (1 - \alpha)^{\bar{n}t} = e_0 N e^{[\bar{n} \ln(1-\alpha)]t} = e_0 N e^{-\frac{13.8}{T}t}, \quad (11.15)$$

with e_0 denoting the start energy of rays. Depending on the absorption model used, the energy-time curve is represented by the count rate and the number of hits.

Energy Multiplication

In this model the number of rays (particles) in the room and the average hits of the detector remain constant (Eq. 11.14). The mean energy therefore is

$$\langle E \rangle = \langle k \rangle e_0 e^{-\frac{13.8}{T}t} \quad (11.16)$$

and

$$\frac{\sigma_E}{E} = \frac{\sigma_k}{k} = \sqrt{\frac{V}{N\pi r_d^2 c \Delta t}}. \quad (11.17)$$

The relative error $\sigma_E/\langle E \rangle$ is easier to be interpreted when expressed in decibels. Now, we can use the fact that the time-dependent sound pressure level is

$$L_{ETC} = 10 \log (E/E_0) \quad (11.18)$$

with an arbitrary reference E_0 , and

$$\sigma_{ETC} \approx \frac{dL}{dE} \sigma_E = \frac{10}{\ln 10} \frac{\sigma_E}{\langle E \rangle}. \quad (11.19)$$

Thus,

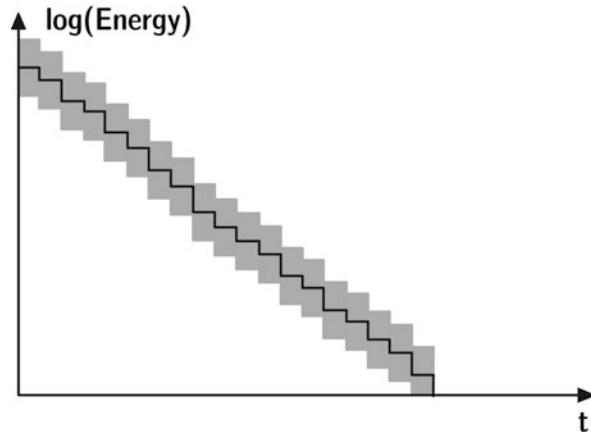
$$\sigma_{ETC} = 4.34 \sqrt{\frac{V}{N\pi r_d^2 c \Delta t}}. \quad (11.20)$$

This error is constant with respect to the time variable in the impulse response. It depends on the ratio of the room volume and the number of rays, the detector size and the width of the time intervals, $V/N\pi r_d^2 c \Delta t$ (Fig. 11.15).

Furthermore on the basis of E , integral parameters, like clarity or strength, can be obtained. The strength, for instance (see Eq. 6.17), is derived from the integral:⁶

⁶Accordingly, the sound power of the source, P , is equivalent to $\pi c r_d^2 e_0 N$.

Fig. 11.15 Histogram result in logarithmic scale of stochastic ray tracing (absorption by energy multiplication), expectation values, and uncertainty range according to Eq. (11.20) (example)



$$\begin{aligned} w &= \sum_{\Delta t} E \approx e_0 N \frac{\pi r_d^2 c}{V} e_0 \int_0^{\infty} e^{-\frac{13.8}{t}} dt \\ &= \frac{4\pi r_d^2 e_0 N}{A} \end{aligned} \quad (11.21)$$

The expectation value of this energy integral, $\langle w \rangle$, also is affected by stochastic deviations. Due to the statistical independence of the hits in the Δt intervals, the total variance of the integral is

$$\sigma_w^2 = \sum_{\Delta t} \sigma_E^2. \quad (11.22)$$

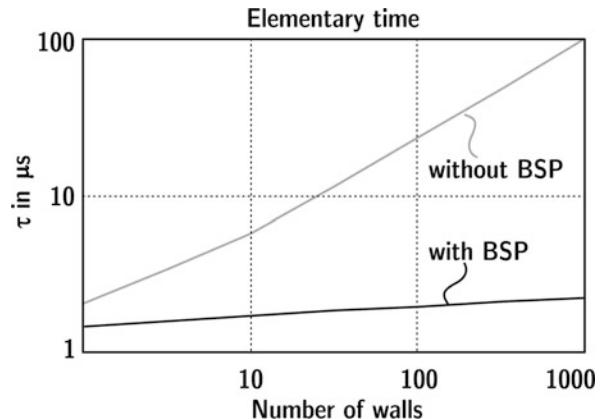
and, thus, again expressed in decibels (applying the principle of Eq. (11.19) again)

$$\sigma_L = 4.34 \sqrt{\frac{A}{8\pi N r_d^2}}. \quad (11.23)$$

The most important fact we observe is that the level errors depend on the square root of A/N . Thus, in large and highly absorbing rooms (large A), a larger number of rays (larger “sound power”) is required, like in impulse response measurements, to overcome the limit of the background noise floor. The noise, in this case, is represented by the numerical noise given by the discrete ray formulation.

Finally, the computation time needs to be estimated. To obtain the number of operations, we have a look into the flow diagram (Fig. 11.10). The inner loop is the tracing and reflection procedure (vector/plane intersection). The number of walls to be checked is n_w and the number of detectors n_d . The computation time required for the point-in-polygon test is t_w , and the time for checking detectors is t_d . Without spatial data structures such as voxel technique or binary space partitioning, BSP (see

Fig. 11.16 Order of magnitude of elementary computation time per reflection as function of the number of wall polygons, without and with binary space partitioning (example for PC with 2 GHz CPU in 2007) (From 2007 to 2020, computation times decreased by a factor of roughly 5)



11.4.5), the complete set of walls and detectors must be tested. The computation time for one reflection is considered as elementary time, τ

$$\tau \approx n_w t_w + n_d t_d + t_c \quad (11.24)$$

with t_c denoting a constant offset for data management independent of the number of walls and detectors.

With BSP or similar spatial data structures, the elementary time is reduced to

$$\tau \approx \log_2(n_w t_w) + \log_2(n_d t_d) + t_c. \quad (11.25)$$

The elementary time is to be multiplied with the events of reflections, $N\bar{n}t_{\max}$:

$$t_{\text{calc}} = N\bar{n}t_{\max}\tau \quad (11.26)$$

As a rough estimate the following elementary computation times can be expected (see Fig. 11.16 and Fig. 11.17). These plots illustrate the dependence of the computation time (order of magnitude) on the number of walls and on the number of rays. The number of receivers is set to $n_d = 1$.

Annihilation

When using this method, more and more rays vanish. The decision to leave the inner loop is made according to a random number. In the case of absorption, the next ray is started (Fig. 11.18).

The expectation value of the energy decreases with the running time in the impulse response since the number of hits decreases exponentially

$$k = N \frac{\pi r_d^2 c \Delta t}{V} e^{-\frac{13.8 \cdot t}{T}}, \quad (11.27)$$

and because of

Fig. 11.17 Order of magnitude of the total computation time (order of magnitude) as function of the number of rays, N (example for PC with 2 GHz CPU). The upper two broken lines represent classical ray tracing, the lower curves show ray tracing with room subdivision strategy (example BSP). The room volume is the same in all four cases

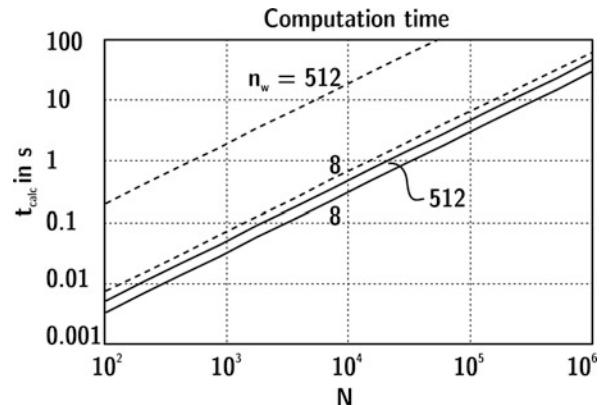
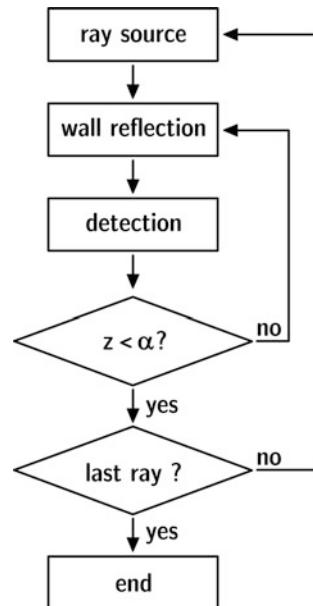


Fig. 11.18 Flow diagram of ray tracing (annihilation)



$$\langle E \rangle = \langle k \rangle e_0 \quad (11.28)$$

results in an increasing relative variation of the energy impulse response of

$$\frac{\sigma_E}{E} = \frac{\sigma_k}{k} = \sqrt{\frac{V \cdot e^{\frac{13.8}{T}t}}{N\pi r_d^2 c \Delta t}} = \sqrt{\frac{V}{N\pi r_d^2 c \Delta t}} e^{\frac{6.9}{T}t}. \quad (11.29)$$

The increase of the uncertainty in the late reverberation tail is due to the loss of countable rays caused by absorption. The statistical error of the integral energy level (in decibels) is

$$\sigma_L = 4.34 \sqrt{\frac{A}{4\pi N r_d^2}}. \quad (11.30)$$

When comparing the equation with Eq. (11.23), it can be observed that the difference in the integral energy error has a factor of 0.7 in favour of the energy multiplication. This is mainly due to the fact of disappearing rays and a smaller statistical basis.

The computation time can be estimated from the flow chart, too. But we have to determine the average number of loops for each ray. The probability that it survives the ν -first reflection and is annihilated at the ν -th reflection is

$$w(\nu) = (1 - \bar{\alpha})^{\nu-1} \bar{\alpha}, \quad (11.31)$$

with $\bar{\alpha}$ denoting the mean absorption coefficient (Eq. 4.30). Accordingly,

$$\langle \nu \rangle = \sum_{\nu=0}^{\infty} \nu w(\nu) = \bar{\alpha} \sum_{\nu=0}^{\infty} (1 - \bar{\alpha})^{\nu-1} = \frac{1}{\bar{\alpha}} \quad (11.32)$$

and

$$t_{\text{calc}} = \frac{N\tau}{\bar{\alpha}}. \quad (11.33)$$

11.3.8 Computation Times Versus Uncertainties: Case Studies

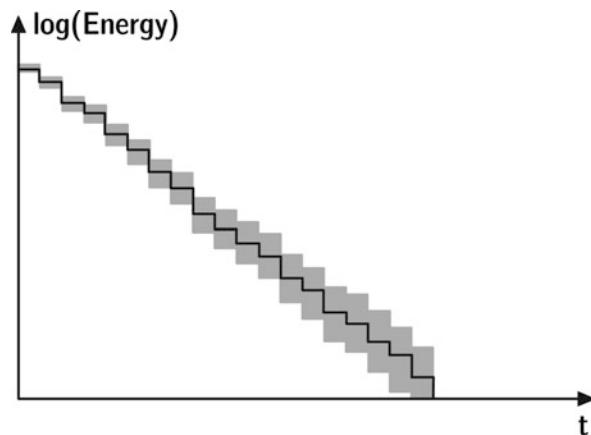
In the following we consider four example rooms as typical cases for categories of rooms used for different purposes: A living room ($V = 100 \text{ m}^3$), a lecture room (classroom) ($V = 1000 \text{ m}^3$), a concert hall ($V = 10,000 \text{ m}^3$) and a large church ($V = 100,000 \text{ m}^3$). The rooms are modelled with a constant average absorption of $\bar{\alpha} = 0.15$. The increasing number of walls, n_w , is due to the larger complexity of the room as its volume is increased. Together with the constant average absorption, the reverberation time also increases (Table 11.2).

In these categories of rooms, we then compare the computation times which are required to achieve a certain quality of the results (one receiver, one frequency

Table 11.2 Examples of room situations

	Room 1 “living”	Room 2 “lecture”	Room 3 “concert”	Room 4 “church”
Volume in m^3	100	1000	10,000	100,000
Surface area in m^2	136	906	4725	21,733
Mean reflection rate in s^{-1}	116	77	40	18
Reverberation time in s	0,8	1,2	2,3	5,0
Number of walls	8	20	50	100

Fig. 11.19 Histogram result of stochastic ray tracing (absorption by annihilation), expectation values and uncertainty range according to Eq. (11.29) (example)



band). At first, impulse responses are discussed. They may be intended either for visual inspection (see Fig. 11.12) of the quality of the energy impulse response or as a basis for further processing with the aim of auralization (see 11.9). The time interval in the histogram is set to 10 ms. Now the type of algorithm for absorption must be chosen. According to the increasing error rate illustrated in Fig. 11.19, the energy multiplication method is clearly preferable. The truncation time is set to $t_{\max} = T$. With the definition of the maximum error at each time interval as 1 dB,⁷ the required number of rays, N , can be defined and the computation times, t_{calc} , in the following table can be expected. The data, of course, only represents an order of magnitude, since they strongly depend on the hardware used and on the software implementation. A standard PC CPU from the year 2019 serves as an example for a rough estimate. The programming language is C++ (Table 11.3).

Next, an integral parameter such as the sound level (strength G) is the goal of the simulation.⁸ Again, the limit is set to 1 dB, but in this example it is related to the total energy in the impulse response. The time limit of ray tracing is $T/2$, which is

⁷Inspired by the limit given by the jnd for sound level of roughly one decibel.

⁸It is assumed that the specific impulse is not relevant in detail. Instead the goal of the simulation is a quick estimation of the level, clarity, definition or other integral parameters.

Table 11.3 Order of magnitude of computation times in s for an expected error of 1 dB in each time interval the energy impulse response (energy multiplication method), $t_{\max} = T$

Computation time in s (Eq. 11.26)	Room 1 “living”	Room 2 “lecture”	Room 3 “concert”	Room 4 “church”
N	200	2000	20,000	200,000
t_{calc}	0.006	0.1	2	36
$t_{\text{calc}} (\text{BSP})$	0.004	0.04	0.5	6

Table 11.4 Order of magnitude of computation times in ms for an expected error of 1 dB in the parameter strength, G. Energy multiplication method, $t_{\max} = T/2$

Computation time in ms (Eq. 11.26)	Room 1 “living”	Room 2 “lecture”	Room 3 “concert”	Room 4 “church”
N	16	100	500	2500
t_{calc}	0.4	3	30	260
$t_{\text{calc}} (\text{BSP})$	0.25	0.8	10	40

Table 11.5 Order of magnitude of computation times in ms for an expected error of 1 dB in the parameter strength, G. Annihilation method

Computation time in ms (Eq. 11.33)	Room 1 “living”	Room 2 “lecture”	Room 3 “concert”	Room 4 “church”
N	31	200	1000	5000
t_{calc}	0.001	0.1	1	10
$t_{\text{calc}} (\text{BSP})$	0.01	0.06	0.34	1.7

absolutely sufficient for the determination of the total level (Eq. 11.41). The number of rays is calculated according to Eq. (11.23). Note that computation times are now given in milliseconds (Table 11.4).

Now, the sound level (strength G) is once again the goal of the simulation and the limit is set to 1 dB. This time, the annihilation method is used. The required number of rays is calculated according to Eq. (11.30). No time limit such as t_{\max} applies since the rays are traced to their annihilation (Table 11.5).

Note that the computation times for a detailed analysis in the impulse responses are much greater than those required for integral results. In other words, integral results can be estimated very quickly. For auralization purposes without reference to the specific room acoustic field, therefore, it might be of interest to quickly estimate the room acoustic parameters and to adjust artificial room impulse processors of early reflections and exponential late reverberation with reference to the integral simulation results (see 16.5.3). Note that this type of quick parameter estimation is done best by stochastic ray tracing with the annihilation method, since it offers real-time capability (time limit about 50 ms; see Chap 16) even for large volumes such as in the case of the church (room 4).

11.4 Image Source Model

On the basis of the image source principle (Sect. 4.4) and its extension towards geometric phase superposition, the total sound pressure of direct sound and various reflections can be modelled by adding (complex) spherical wave amplitudes. However, the model is exact in the case of $|R| = 1$. For absorbing walls it is a good approximation, as long as the angle of sound incidence, ϑ_0 , is small and far from a grazing incidence (see Sect. 3.3).

11.4.1 Classical Model

The reflection factor is assumed to be angle-independent, corresponding to a quasi-plane wave:⁹

$$p = \frac{j\omega\rho_0\hat{Q}e^{-jkr_0}}{4\pi r_0} + \frac{j\omega\rho_0\hat{Q}e^{-jkr_1}}{4\pi r_1} R(\vartheta_0). \quad (11.34)$$

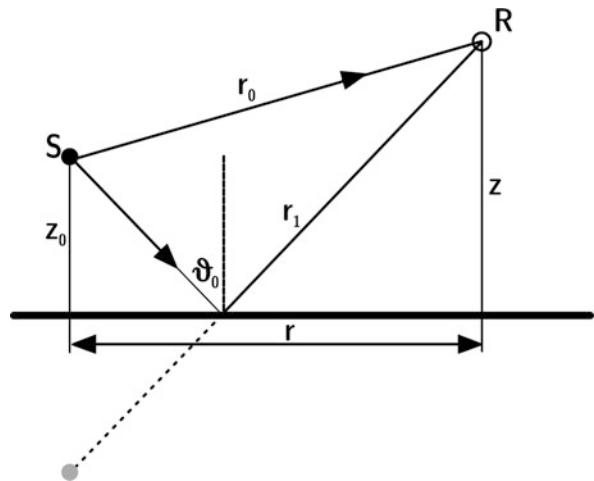
At the same time it is clear that the wall must be smooth and specularly reflecting. From the definition of the single reflection, the image source model can be described as can be seen below. For specific geometries (rectangular, triangular), it has served as a model for analysing the basic features of room impulse responses since the middle of the last century (Cremer 1948). Then, the so-called “Allen-Berkley/Borish” (Allen and Berkley 1979) model was first implemented by Borish (1984) in arbitrary polyhedra and later used in numerous versions, not only in acoustics but also in radio wave physics and in computer graphics in similar ways (Fig. 11.20).

If the room reflections are purely specular, the sound paths (rays) can be traced back from the receiver to the source. This is achieved by using virtual (image) sources. At first, they must be constructed for the room of interest. The original source is mirrored at the wall planes. Each image source is then mirrored at wall planes, to create image sources of higher order. All permutations of the walls must be considered, except a constellation involving the same wall subsequently. Under specific circumstances, walls can be excluded due to geometrically inconsistent ray paths (Mechel 2002).

With \vec{S} denoting the source position, \vec{S}_n the position of the image source, \vec{n} the wall normal vector ($|\vec{n}| = 1$) and \vec{r} the vector between the foot point, \vec{A} , of the wall normal and source, \vec{S} , the scalar product of and yields the distance between the source and the wall, d .

⁹The plane-wave assumption is equivalent to the prerequisite of non-grazing sound incidence.

Fig. 11.20 Image source model



$$d = \vec{r} \cdot \vec{n} = |\vec{r}| \cos \alpha \quad (11.35)$$

With this distance we get the position of the image source:

$$\vec{S}_n = \vec{S} - 2 d \vec{n}. \quad (11.36)$$

With this procedure applied to all walls, we can construct all image sources of first order. Image sources of higher order are constructed in the same way by considering the first-order image sources as “mother” sources and so on. The process of mirroring is continued until a certain maximum order of image sources is reached. The truncation of the process is similar to the truncation of ray tracing at a maximum time, t_{\max} . Image sources of n th order correspond to rays hitting n walls (Fig. 11.21).

11.4.2 Audibility Test

With the set of image sources created, a so-called audibility test must be performed. It is necessary to check the relevance of each image source for the specific receiver position. Receivers, by the way, are points.

Each image source is interpreted as the last element of a chain of sources. The indices denote the series of walls hit on the path of the corresponding ray, while the number of indices denotes the order of the image source. A chain of i th order is

$$S \rightarrow S_{n_1} \rightarrow S_{n_1 n_2} \rightarrow S_{n_1 n_2 \dots n_{i-1}} \rightarrow S_{n_1 n_2 \dots n_i} \quad (11.37)$$

Fig. 11.21 Construction of image sources

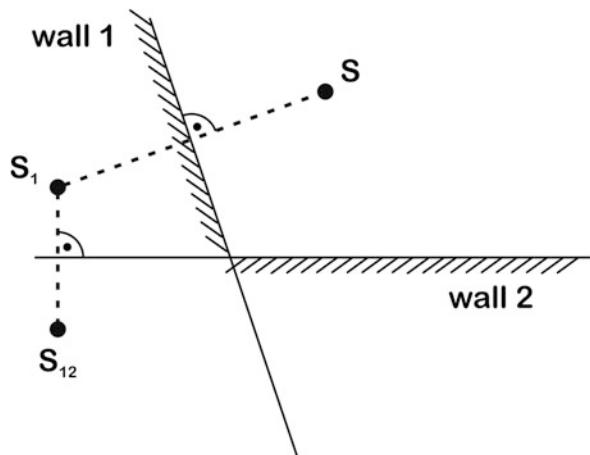
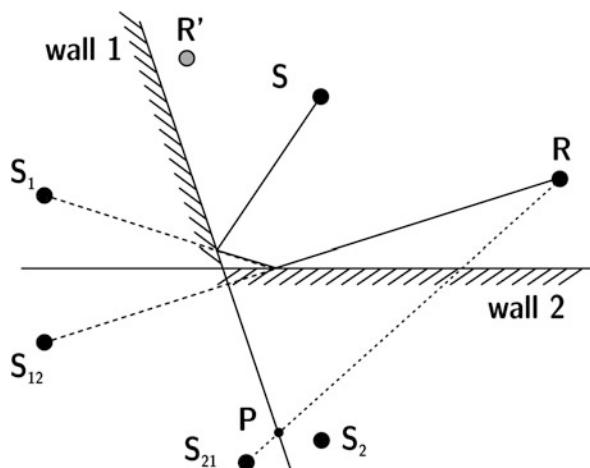


Fig. 11.22 Image source audibility test



with $n_k \neq n_{k \pm 1}$ counting the walls hit, $n_k \in (1, n_w)$. The audibility test is started at the receiver point, and the ray is traced back from this point to the source along the chain of image sources.

As an example, Fig. 11.22 shows the determination of audible image sources up to the second reflection order. The current receiver point is connected with the image source under test, S_{12} . The last index indicates that wall 2 was the last to be hit. If the intersection point of the straight line with the wall polygon 2, \overline{RS}_{12} , is inside the polygon, the result is preliminary positive. We then continue with drawing a line from the intersection point to the mother source¹⁰ of S_{12} , S_1 . This procedure is repeated until the original source is reached. If all intersections hit inside the

¹⁰A mother source is the source located in the chain as predecessor.

polygons, the image source (in this case S_{12}) is in fact audible from the receiver point R. Note that S_{21} is not audible from R since P is outside wall polygon 1. But it would be audible from R'. Thus, each source must be tested specifically in relation to the receiver. Also, the fact that one image source is audible cannot be generalized for its predecessors. It should be noted, too, that the crucial test of this procedure is the point-in-polygon test, like in ray tracing. All strategies to accelerate the back-tracing procedure by spatial substructures may therefore be applied as well.

For audible image sources, the complex amplitudes are stored in the sound pressure impulse response. The delay of the impulse related to each image source is calculated from the distance between the image source and the receiver, r_{IS} .

$$p_{IS} = \frac{j\omega\rho_0\hat{Q}}{4\pi c r_{IS}} e^{-j\omega t_{IS}} \prod_{n=1}^i R_n \quad (11.38)$$

with $t_{IS} = r_{IS}/c$ and R_n being the reflection factors of the walls involved in the respective chain source. No time intervals are required at this stage of processing. The temporal resolution can be infinitely high, and the sampling rate can thus be freely defined.

An estimation of the reflection factor from the absorption coefficient can be done straightforward by using

$$|\underline{R}| = \sqrt{1 - \alpha} \quad (11.39)$$

and an appropriate phase. Rindel (1993) and Mommertz (1996) show that absorption coefficients in one-third octave bands serve well for reconstructing the complex reflection factor. Minimum or linear phases are options for the phase function, to be reconstructed, for instance, from Hilbert transformation.

11.4.3 Limitations

Truncation

The image source model is a strictly deterministic method. Uncertainties are caused by the prerequisite of geometrical acoustics and by the necessary truncation of the image source of a certain order. Although not explicitly implemented, an average maximum time, t_{max} , has to be taken into account. The maximum order is related to the computation time, which is more crucial than in the case of ray tracing. This fact is due to the dramatic increase of the number of sources which need to be treated with higher orders.¹¹ The number of audible sources, however, just increases with t^3 in the impulse response (Cremer 1948).

¹¹Without strategies for excluding sources for geometric reasons, the increase is exponential.

At higher orders of image sources, the ratio between constructed and visible sources is absolutely uneconomic. With i denoting the chosen maximum order, the truncation time is

$$t_{\max} = i/\bar{n} \quad (11.40)$$

and the energy missing is

$$\frac{w_{\text{trunc}}}{\langle w \rangle} = (1 - \alpha)^i. \quad (11.41)$$

The computation time, t_{calc} , is similarly estimated from the time required for the point-in-polygon test (Eqs. 11.22 and 11.23):

$$t_{\text{calc}} = \tau \cdot N_{\text{IS}} \quad (11.42)$$

The classical (Allen-Berkley / Borish) image source model is thus applicable and efficient for the following cases:

- For short impulse responses (second or third order)
- For simple geometries (small n_w)
- For rectangular rooms (since the audibility test can be omitted)¹²

Spherical Waves at Grazing Incidence

Another limitation is given by the validity of Eq. (11.34). The reflection coefficient $R(\vartheta_0) \neq 1$ in the contribution of the image source implies a constant angle of incidence and thus a reflection of a plane wave at one reflection point. This assumption is well suited when a spherical wave¹³ is considered with a large distance between the source and the receiver at the wall.

At smaller distances and corresponding grazing incidence, Eq. (11.34) carries a non-determined uncertainty. It would then be necessary to calculate the reflected wave in the more general form.

Generally, the sound pressure at the receiver point is given by

$$p = \frac{j\omega\rho_0\hat{Q}e^{-jkr_0}}{4\pi r_0} + p_{\text{IS}}, \quad (11.43)$$

whereas the standard model offers an estimate equation

¹²For a rectangular room, a regular lattice of image sources applies, with multiple sources (of indices in permuted order) coinciding at the lattice points. It can be shown, however, that for every receiver point exactly one lattice source is audible.

¹³Relevant here is the temporal coincidence of the wavefront phases hitting the surface.

$$p_{IS, \text{plane wave}} = \frac{j\omega\rho_0\hat{Q}e^{-jkr_1}}{4\pi r_1} R(\vartheta_0) \quad (11.44)$$

with

$$R(\vartheta_0) = \frac{\zeta \cos \vartheta_0 - 1}{\zeta \cos \vartheta_0 + 1}. \quad (11.45)$$

It includes the specific wall impedance, ζ for locally reacting surfaces. The exact solution accounts for spherical wave propagation following Eq. (3.17).

The errors introduced by the plane-wave assumption for impedance and reflection factor have been investigated both by experiment and by field calculation (Suh and Nelson 1999). As a rule of thumb, we should remember that at grazing incidence ($\vartheta_0 > 60$ degrees) and at too close distances, d , of receiver and source to the wall ($d \leq \lambda$) systematic uncertainties of sound pressure, Δp which are clearly audible ($\Delta p_{IS} > 20\%$, which corresponds to 1 dB) must be considered. In the middle of a room, errors are smaller than at positions near the room's boundaries. Small- and medium-size rooms ($50 \text{ m}^3 < V < 200 \text{ m}^3$) at low frequencies ($f < 200 \text{ Hz}$), however, hardly have a centre area which is further away from the boundaries than a wavelength. Here, the validity of standard (i.e. plane-wave impedance) geometrical acoustics is at its limits.

A more specific and precise uncertainty of the auralization result cannot be predicted, since the simulation errors in the end depend not only on the room but on the type of signal. A pure tone simulation is by far more delicate than a broadband simulation over several critical bands. The fact that walls are not locally reacting at low frequencies and that the impedance may be distributed over the entire wall (by distributed screws in studs, for instance) creates further complications.

11.4.4 Diffraction

In the classical image source model (and also in ray tracing), diffraction is neglected. In room acoustics diffraction may happen for two reasons: there can be obstacles in the room space (e.g. stage reflectors), or there can be edges at the surroundings of the finite room boundaries. In the latter case, either the boundary is forming an obstacle, such as columns or the edge of an orchestra pit, or the boundary is forming the edge between different materials with different impedances (and absorption). Since diffraction is a typical wave phenomenon, it is not included in the basic simulation algorithms listed above. In the past, there were some ideas of including diffraction as a statistical feature into acoustic ray models. But the success was quite limited because the increase in calculation time is a severe problem (Fig. 11.23).

In optics and radio wave physics, however, ray tracing models were generalized in the uniform geometrical diffraction theory, UDT (Kouyoumjian and Pathak 1974;

Fig. 11.23 Geometric edge diffraction using secondary sources. (After Svensson et al. 1999)

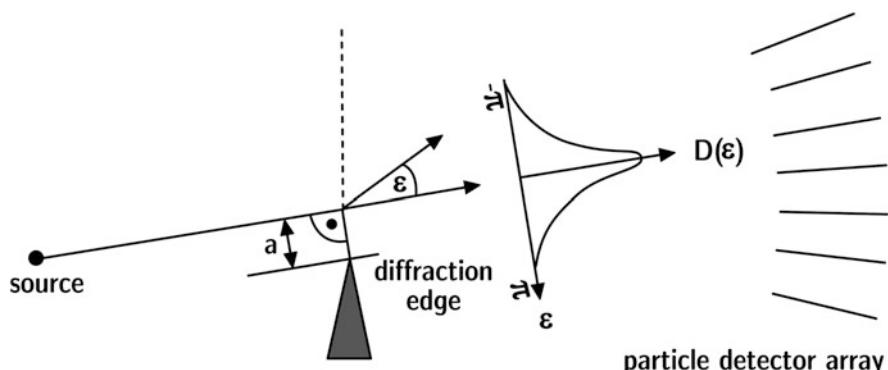
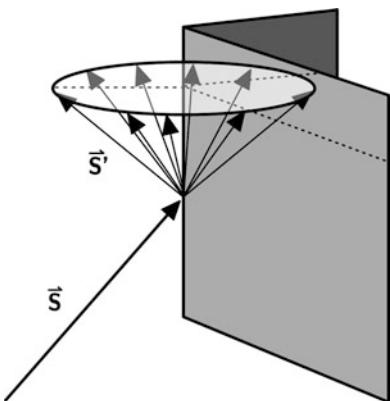


Fig. 11.24 Ray (particle) diffraction model. (After Stephenson 2004b)

Tsingos et al. 2001). Other approaches were presented by Svensson et al. (1999), who applied the model developed by Biot and Tolstoy (1957) to acoustic problems. These approaches are very powerful for the determination of first- or second-order diffraction. But all methods of geometrical diffraction are very time-consuming for the simulation of a multiple-order diffraction and corresponding reverberation, and they also introduce ray-splitting and an exponential factor in the multiple-scattering algorithm. Another possibility is to apply the finite element or boundary element methods, of course.

Stephenson (2004a, b) developed two approaches for edge diffraction. One consists of the implementation of “virtual flags” mounted at the edge, which let the edges affect rays in the room interior (Fig. 11.24).

The flag’s width is about one wavelength representing the midband frequency of the octave or one-third octave band. Rays hitting the flags can be diffracted according to the statistical distribution of diffraction angles depending on the distance from the edge, the wavelength and the ray incidence angle. This distribution is derived from either slit diffraction models or Fresnel’s edge diffraction models.

Both methods show good results in case studies, but they are, however, not perfect in the variety of all general cases in practice.

Diffraction models and their implementation in image source and ray tracing algorithms will be one of the greatest challenges in future developments of geometrical room acoustic simulation methods.

11.4.5 Reduction of Computational Load by Preprocessing

Field Angle

Geometric tests help to reduce the computational load to create and check the audibility of image sources. If inconsistent sources are found, construction of daughter sources can be avoided. Mechel (2002) defines the field angle which is valid for each image source. If the receiver is not located in the field angle, the source is inaudible. Now, if a wall polygon to be used for creation of the daughter source (see 11.4.1) is outside the field angle, the mother source cannot give birth to this daughter. In the example in Fig. 11.25, it is illustrated that the image source S_{70} (the daughter of source S_0 and granddaughter of S) spans a field angle with the “father” wall no 7. Receivers outside the field angle do not receive sound from S_{70} . Furthermore, in this example a convex corner situation is present and, thus, the field angle is affected by a shadow region.

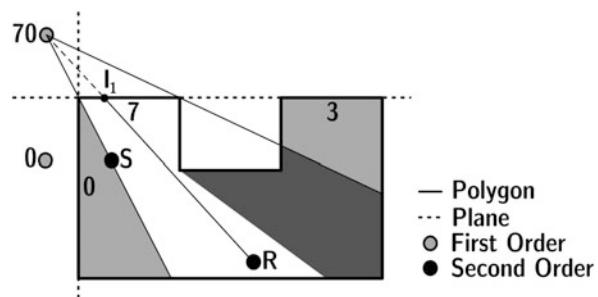
With these and other criteria, strategies for interrupting the creation of image sources can be reduced significantly (Mechel 2002).

Spatial Data Structures

Spatial data structures have been commonly used in applications of computer graphics and have already been applied in room acoustics simulation algorithms (see Funkhouser et al. 1998; Jedrzejewski and Marasek 2004).

Different types of spatial data structures are in use, such as bounding volume hierarchies, binary space partitioning (BSP) trees and octrees. They all have in common that the entire geometry is subdivided into smaller subspaces, which are encoded in the data structure. In contrast to bounding volume hierarchies, which

Fig. 11.25 Field angle of image source S_{70} (reflection path with walls 0 and 7)



separate the scenery into single objects enclosed by bounding volumes like spheres or boxes, BSP and octrees subdivide the scenery into subspaces without taking into account any information about single objects.

Octrees divide the scenery regularly along all three axes. Hence, every division of space results in eight new boxes. BSP trees allow a flexible partitioning of space by means of arbitrary-shaped partitioners. Both methods satisfy the same type of queries. For simple intersection tests, the BSP structure is more appropriate than octrees due to its higher flexibility.

Spatial subdivision helps to speed up the required intersection tests, the fast point-in-polygon test and an efficient data structure with respect to real-time auralization in order to provide a faster determination of the BRIR. This goal is illustrated further by discussing the example of BSP (Shumacker et al. 1969).

Binary Space Partitioning

The aim of BSP is speeding up the determination of the location of an arbitrary point relative to a geometrical scene, which is encoded by a so-called BSP tree.

It is preferable to use planes spanned by polygons as partitioners (Fig. 11.26). Each tree node contains one partitioner that divides the current subspace into two smaller subspaces. In the problem of finding the next reflection point, the point in question can have three possible relative locations: (1) on, (2) in front of, and (3) behind the partitioner. In the first case, the query can be stopped. To pursue the latter two cases, the test is continued by branching to the respective son of the node, left for “behind” and right for “in front”. The tree’s root node therefore refers to the whole space, while leaf nodes refer to a subspace which does not contain any further scene detail. Contrary to the naive approach to test the position of the point against all planes, the tree structure allows the determination of the position by testing only a subset of planes. This subset is defined by the path in the tree. By using a balanced tree, i.e. a tree of minimum height, the number of tests can also be minimized, which drops the complexity from $O(N)$ to $O(\log_2 N)$, where N is the number of polygons.

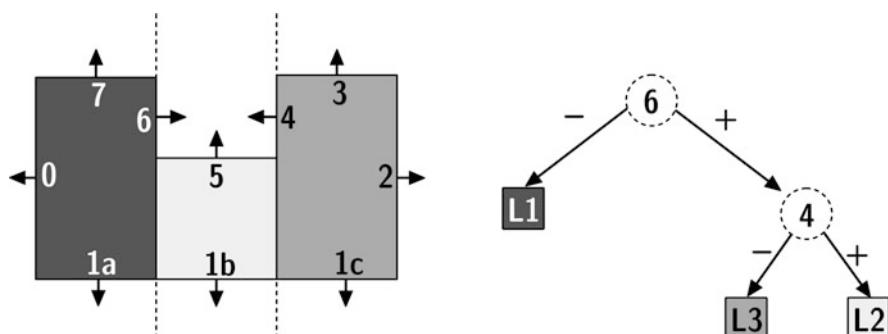


Fig. 11.26 Example of a BSP tree for a given room geometry. (After Shumacker et al. 1969; Schröder and Lenz 2006)

Voxels

Like octrees, voxels are volume boxes. They are created to subdivide the actual room volume, for example, by using cubes. In one pass of preprocessing, it is tested for each voxel with which walls it intersects. Later, in the ray tracing process or visibility test, we look for the next intersection point with a polygon. This test can be speeded up by not testing the complete set of polygons but only polygons in voxels along the straight line representing the ray or image path. Stephenson (2006) derived an analytic equation for the gain in computation which results to

$$t_{\text{calc}}|_{\text{voxel}} \propto \sqrt{n_w}, \quad (11.46)$$

which was a previously linear increase without using voxels (Eq. 11.24).

11.5 Hybrid Image Source Models (Deterministic Ray Tracing)

Ray tracing and image source algorithms have opposite advantages and disadvantages. Therefore it is worthwhile to develop a combination of both in order to obtain:

- Fine temporal resolution in the sampling rate quality
- Inclusion of scattering
- Faster audibility check of image sources

This kind of combination is called “hybrid method”. The term “ray tracing” is also in use, however, somewhat confusing with the stochastic ray tracing described above. The key to combine the methods is the audibility test of image sources in the forward direction.

If we run a specular ray tracing process and find a receiver hit by a ray, the corresponding image source must be audible.

Variants of this approach are often summarized as “ray tracing”. Specific algorithms are also known as “cone tracing”, “beam tracing”, “pyramid tracing”, etc. What all have in common is the idea of a forward audibility test of specular reflections, represented physically by image sources. The specific physical reason is the fact that the energy contributions of the reflections are calculated by using Eq. (11.38). The main difference between this type of ray tracing and the previously discussed stochastic ray tracing method is the way of detecting sound energy (Fig. 11.27).

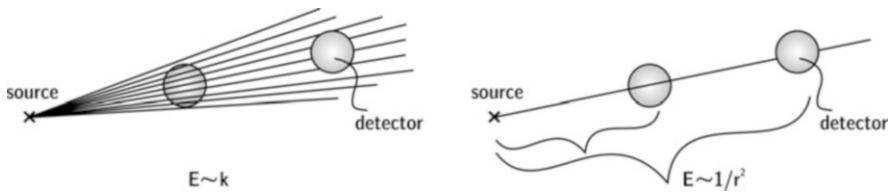


Fig. 11.27 Energy detection in stochastic (left) and deterministic (right) ray tracing

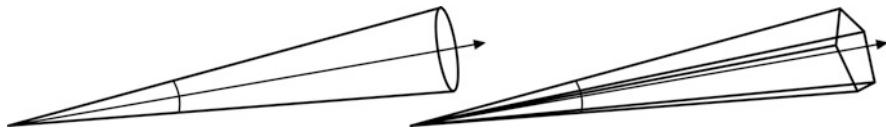


Fig. 11.28 Cone (left) and pyramid (right) tracing

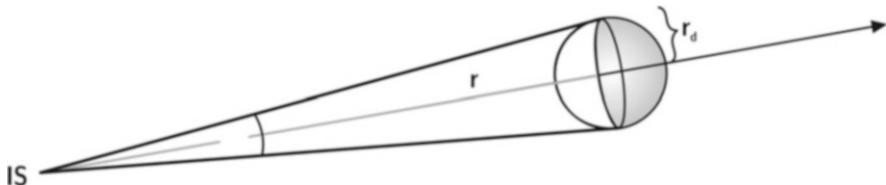


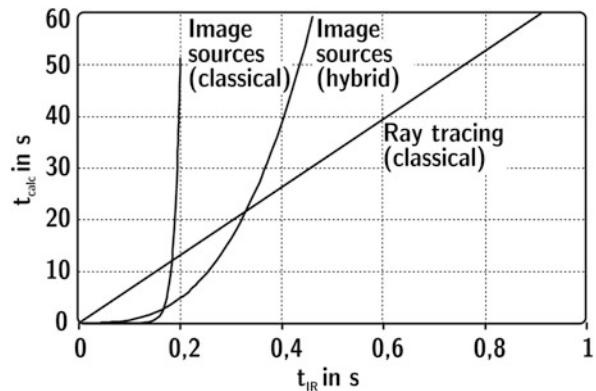
Fig. 11.29 Geometry of source, cone and receiver

Of course, in a running ray tracing process, rays hitting the receiver can have the same history of wall reflections. In this case we observe an energetic overlap between rays, and it must be ensured that each image-source-based reflection is counted just once, for instance, by checking the indices in the wall list. Exactly this point is the reason for the differences in the variants listed above. In the various concepts, rays are extended towards geometrically diverging objects like cones or pyramids (Fig. 11.28).

This way it is also ensured that image source contributions are counted only once. While cone or beam tracing still create slight problems with overlap (Vian and van Maercke 1986), pyramid tracing which uses the wall polygons as pyramid bases avoids double counts (Stephenson 2004b). At the same time, point-like receiver can be because of extended and diverging ray geometries.

The required number of rays or cones depends on the maximum distance between the image source and the receiver and therefore on the order of the image source and the corresponding average delay in the impulse response, as illustrated with the example of a cone (Fig. 11.29).

Fig. 11.30 Computation times of room acoustic simulation algorithms (example)



The figure shows the path between image source and receiver in an expanded version.¹⁴ N cones are started into the full space of 4π . From these

$$k = N\Omega/4\pi \quad (11.47)$$

will hit the receiver, with Ω denoting the spatial angle of the cone:

$$\Omega r^2 = \pi r_d^2 \quad (11.48)$$

for $r \gg r_d$. Hence the probability of cones hitting the receiver is

$$k = \frac{Nr_d^2}{4r^2} = \frac{Nr_d^2}{4(ct)^2}. \quad (11.49)$$

To ensure at least one ($k = 1$) cone hitting the receiver after a time t_{\max}

$$N_{\min} = \frac{4(ct_{\max})^2}{r_d^2}. \quad (11.50)$$

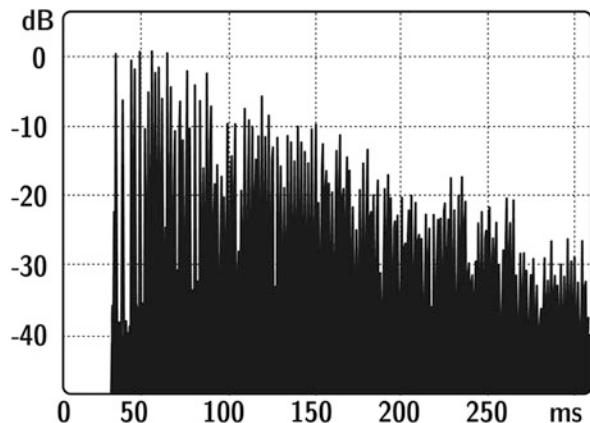
The computation time of this kind of ray tracing (see also Eq. 11.26)

$$t_{\text{calc}} = \frac{4c^2 \bar{n} l_{\max}^3}{r_d^2} \tau, \quad (11.51)$$

increases with the third power of the impulse response length (Figs. 11.30 and 11.31).

¹⁴The real path is folded from reflection to reflection. The total path length is identical in the expanded version.

Fig. 11.31 Results of a hybrid image source algorithm



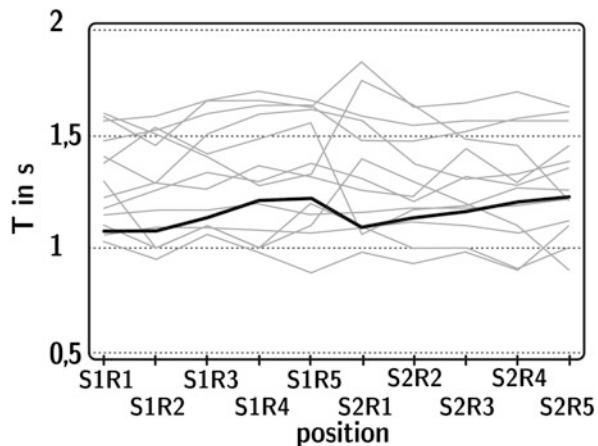
In the last decades, several authors presented strategies to fight against the computational load of the image source method (Mechel 2002; Funkhouser et al. 1999; Stephenson 2004a). Other authors presented interesting approaches for generalizing the image source principle with regard to diffraction and scattering (Martin and Guignard 2006). It should be stressed again that the standard image source method is based on non-absorptive walls or, at least, on absorption in the plane-wave approximation. Approaches of extensions which do not consider wave theoretical extensions still suffer from limits set by the theory of specular reflections.

Before we will focus on strategies for speeding up the simulation algorithms for auralization, the physical limits of room acoustics simulation by ray tracing and image sources should be discussed. Uncertainties of simulation methods applied in practice were checked and documented. The basic material data available and the necessary psychoacoustical rating of the results are also important and need to be considered.

11.6 Systematic Uncertainties of Geometrical Acoustics

Geometrical acoustics (ray tracing, image sources) allows simulation of sound fields in rooms and outdoors. These models yield correct results under specific circumstances (mostly academic cases) and then fulfil the wave equation. But absolute physical correctness is not always a reasonable goal. In by far the most cases,

Fig. 11.32 Round robin I, PTB auditorium (Vorländer 1995). Reverberation times calculated and measured (thick line) for source/receiver combinations



geometrical acoustics offers approximate solutions, solutions which are absolutely sufficient in practice. But it should be kept in mind that the limits are:

- Large rooms, non-grazing incidence¹⁵
- Curved surfaces, focusing (Vercammen 2010)
- Low absorption coefficients
- Broadband signals

Room acoustical results like reverberation time, clarity, strength, etc. are usually expressed in relation to frequency bands. The reason for using frequency bands is that sound signals related to room acoustics are of broadband character, like speech and music. Also, from a physiological point of view, one-third octave bands are a good compromise since they are an approximation for critical bands (Sect. 6.3.1). These arguments are the justification for using energetic methods (neglecting phases). At low frequencies, however, the situation is different. Transmission of pure tones, possibly the fundamental frequency of musical instruments, is surely of interest in optimizing the placement of recording arrangements in studios or in finding the optimal loudspeaker positions. Simulation and auralization of harmonic signals in such cases may largely differ from the sound pressure in the real case.

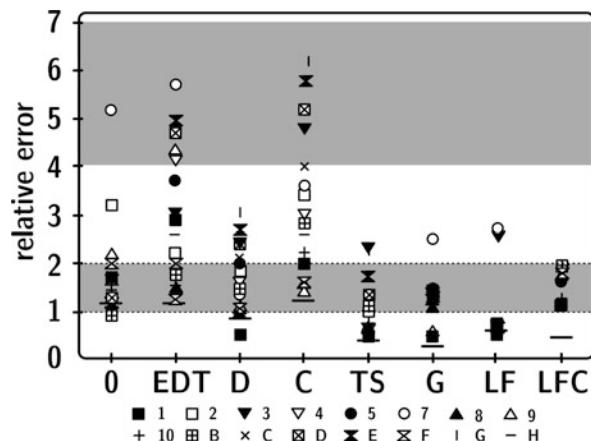
In 1995, room simulation software solutions were tested in an international round robin project. Eight parameters defined in ISO 3382 (T , EDT, D , C , T_S , G , LF and LFC) were calculated at two source positions and five receiver positions in a test room.¹⁶ The results were compared with measurement results, at first only in one frequency band (1 kHz) (Fig. 11.32).

The simulated reverberation time results of the first phase of the project were generally too large, thus indicating that the users underestimated absorption. Then, after performing measurements, the absorption coefficients were published, and the

¹⁵or tangential modes.

¹⁶Auditorium of the Physikalisch-Technische Bundesanstalt, Braunschweig, Germany.

Fig. 11.33 Overall results of the “round robin I” (Vorländer 1995). The ordinate scale refers to a normalized deviation from the jnd (At that time the jnd for clarity was assumed to be round 0.5 dB. Later it was shown that 1 dB is a better approximation; see also Table 6.1)



software was run again with harmonized input data. The agreement was much better, of course, but now the efficiency of the software could clearly be identified. The overall accuracy of the computer simulations can be estimated by a single number rating. This quantity is defined by using the absolute difference between the result from the simulation and the measurement on each of the ten source and receiver combinations. The average of these location-dependent differences is then related to the jnd for the respective acoustic criterion. The resulting relative error for each acoustical parameter is shown in Fig. 11.33. Note that a value of 1 indicates the order of magnitude of the jnd. Only 3 out of 17 programmes could deliver results of similar accuracy, which is given by the standard deviation of the measurements and by the just noticeable differences (Table 6.1).

In the second intercomparison project, Bork (2000a) investigated the results of simulations in a larger room and for more frequency bands. The example room was the ELMIA hall in Jönköping, Sweden. The intervals of the jnd related to the specific psychoacoustic dimensions were refined. The results generally supported the conclusion from 3 years before, namely, that simulation algorithms require a module for the treatment of scattering. Also, in the second project, one CAD room model was provided in order to run software with identical input data.

Finally the third round robin was focused on a smaller room with expected modal effects. The results again support the fact that (a) a good scattering model is essential and (b) the overall prognosis accuracy, when results are compared with experimental results, can be in the range of 1 to 2 jnd. For achieving this accuracy, programmes are available, so that a certain quality of the results can be guaranteed. The operator of the software, however, still has crucial influence on the choice of absorption and scattering coefficients.

After all, room acoustical computer simulation has been proven to yield reliable results. Scattering was identified as an important factor. With careful data choice of absorption and scattering coefficients, it is possible to obtain good prediction results which deviate from the experimental result in the same order of magnitude as the uncertainty of the experiment or the jnd.

More information about open data and open-source software including intercomparisons and benchmarks in computational acoustics is given in Sect. 18.5.

11.7 Hybrid Models in Room Acoustics

Deterministic models like the image source model suffer from inherent systematic errors and from limitations in the software implementation. Stochastic models suffer from poor temporal resolution, but they can handle scattering. Impulse responses from image-like models consist of Dirac pulses arranged according to their delay and amplitude and sampled with certain temporal resolution. However, in the intercomparison of simulation programmes (see 11.6), it soon became clear that pure image source modelling will create too rough an approximation of physical sound fields in rooms, since a very important aspect of room acoustics – surface and obstacle scattering – is neglected. This fact is further supported by the observation that scattered energy will dominate the reverberation process after a few reflections (order 3 or 4), even in rooms with rather smooth surfaces (see Fig. 11.34 and Kuttruff (1995)). Also, the audibility of characteristics of diffuse reflections was proven by (Torres et al. 2002).

Fortunately, the particular directional distribution of scattered sound is not relevant in the first place and can well be assumed as Lambert scattering. Solutions to the problem of surface scattering are given by either stochastic ray tracing or radiosity. Another effect of wave physics – diffraction – can be introduced into geometrical acoustics in principle. Big problems, however, arise from extending diffraction models to higher orders (Svensson et al. 1999). Thus, in applications other than in

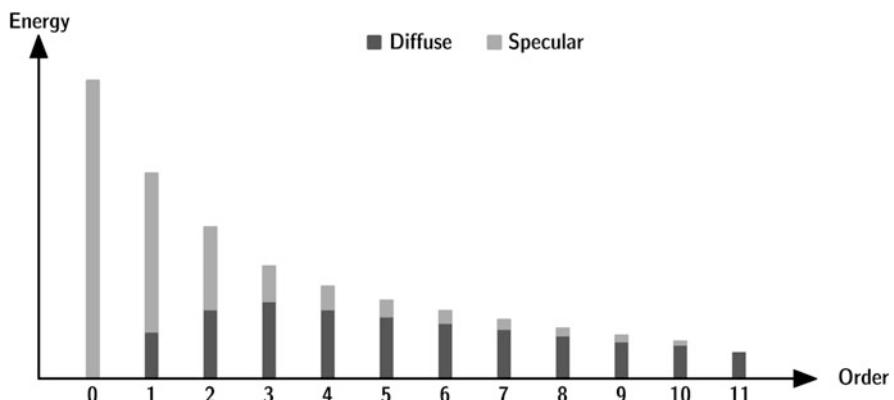


Fig. 11.34 Conversion of specularly into diffusely reflected sound energy, illustrated by an example. (After Kuttruff 1995)

outdoor noise propagation,¹⁷ diffraction models have not yet been implemented. New algorithmic schemes, like the one presented by Stephenson (2004b), have not yet been implemented either. It should be mentioned here that until today numerous algorithmic details have been published in the field of wave field simulation. Sound transmission and diffraction, too, must be implemented in cases of coupled rooms or in room-to-room transmission; see Chap. 12.

11.7.1 Hybrid Deterministic-Stochastic Models

The general principles of room acoustic modelling were discussed in the previous sections. They yield results of high temporal resolution (in scales of μs), typically represented by reflections of directly used or implicitly involved image sources. Due to numerical constraints and also due the fact that it is unnecessary to create such a high resolution for the late reverberation ($>100 \text{ ms}$), stochastic methods like ray tracing and radiosity now move into the focus of our interest. It is obvious to combine the tedious but exact deterministic methods with a quick estimation of the reverberation tail.

In the following, some well-known software solutions are briefly discussed. This overview is not intended to cover all kind of software used today. The particular examples listed here represent the historic process of software development for room acoustic computer simulation using hybrid models in the 1990s.

CATT-Acoustic

In his thesis, Dalenbäck (1995) studied aspects of an extended radiosity approach merged with specular reflection techniques (see also Dalenbäck 1996). The implementation of the software had already been started in the late 1980s and has continuously been improved and extended. The main model is based on predicting high-order reflections using randomized tail-corrected cone-tracing and direct sound, first-order specular and diffuse and second-order specular-to-specular reflections. The reflections are created using deterministic methods and image sources with full temporal resolution and bandwidth. Higher-order reflections are the result of independent ray/cone-tracing for each octave-band taking into account the frequency dependence of diffuse reflections via randomly selecting specular or diffuse via the scattering coefficient magnitude where a diffuse ray is reflected using the Lambert distribution. Dalenbäck and McGrath (1995) also presented tools for dynamic auralization using headphones with head tracking already in 1995 (Fig. 11.35).¹⁸

ODEON

The first version of ODEON was published in (Naylor 1993). It is a hybrid image source model with a stochastic scattering process which uses secondary sources. The

¹⁷Almost free-field propagation with at most one diffraction edge.

¹⁸www.catt.se

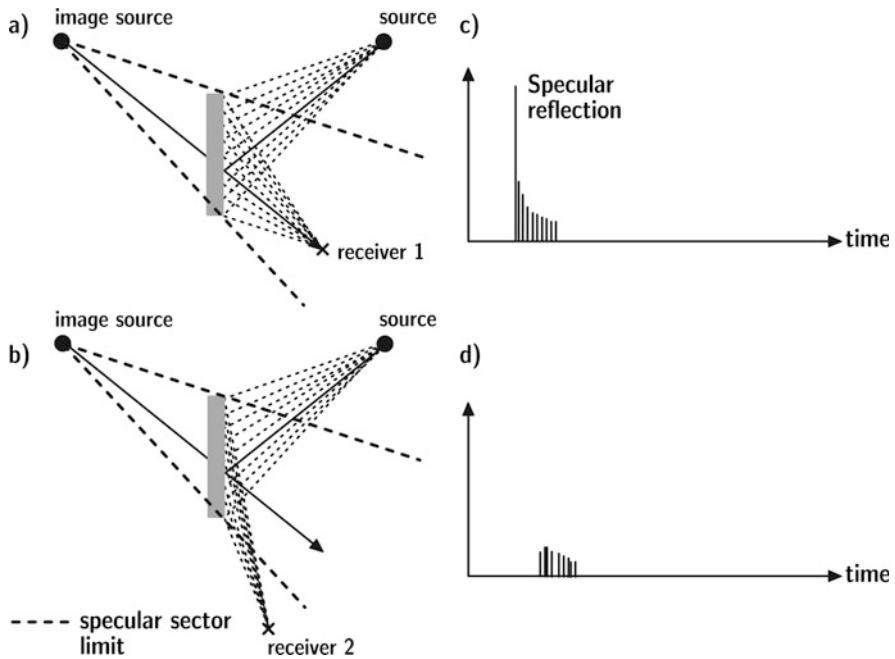


Fig. 11.35 Scattered energy radiated from surface patches in CATT. (a) First-order diffuse and specular reflections, (b) first-order diffuse reflections, (c) schematic echogram at receiver 1 (direct sound omitted), (d) schematic echogram at receiver 2 (direct sound omitted)

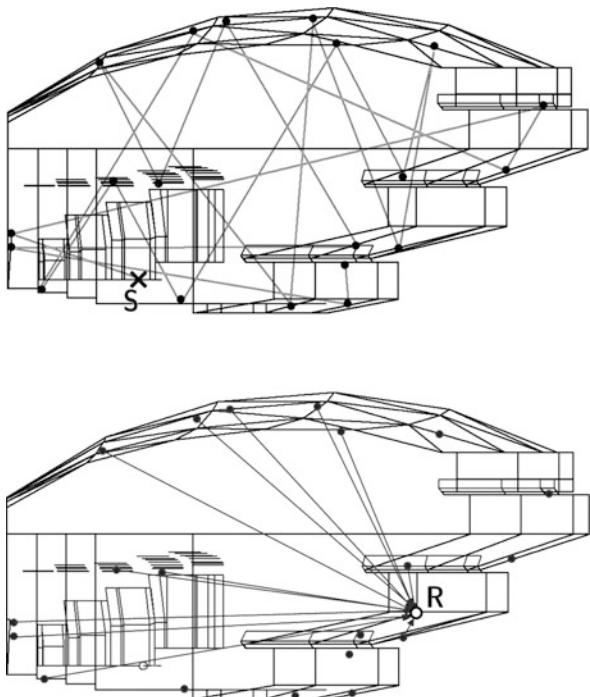
secondary sources are assigned a frequency-dependent directionality, the so-called reflection-based scattering coefficient. This implies that the direction of the ray that created the secondary source is taken into account. The polar scattering patterns are created from a vector-based process which adds Snell's (specular) and Lambert's (diffuse) scattering into one final scattering direction for each ray (Rindel and Christensen 2003).¹⁹ The creation process of secondary sources is shown in Fig. 11.36 (top) for the example of one single ray. The irradiation of sound from the secondary sources to a receiver is shown in Fig. 11.36 (bottom) for the example of a receiver on the first balcony. Some of the secondary sources cannot contribute to this receiver due to the audibility check.

QPBT

QPBT is the abbreviation for Quantized Pyramidal Beam Tracing (Stephenson 1996, 2004a). The geometric principle of the method is pyramidal beam tracing on the basis of the wall polygons which form the pyramidal base area. Successive reflections are created by extending and splitting the pyramids towards the other wall polygons. The problem of exponential increase of pyramids is inherent in the geometric imaging method. In QPBT, this is solved by the unification of spatially

¹⁹<http://www.odeon.dk>

Fig. 11.36 Secondary source model in ODEON.
 (After Rindel 1993). S is the source, R the receiver point.
 Top, creation of secondary sources; bottom, only the
 secondary sources with free
 line-of-sight contribute to
 the sound pressure at the
 receiver on the balcony



closely located pyramids. Furthermore, by also discretizing the time frame, it is possible to recombine sound energy into finite elements of time and spatial angle. The particular way of energy travelling via a pyramid to the discrete element is irrelevant. The model is thus open for inclusion of scattering and edge diffraction. In which way edge diffraction is implemented best in the pyramidal beam has actually been tested in Stephenson and Svensson (2007) (Fig. 11.37).

EASE

The software EASE stems from the application of simulating installations of professional audio and sound reinforcement systems (Ahnert and Feistel 1993).²⁰ The database of loudspeaker frequency responses and directivities is hence very large, and it includes innovative approaches for detailed and accurate loudspeaker representation. It was first implemented in an image source model capable of simulating particularly the direct sound and the early response in particular.

In a later version of EASE, the room acoustic simulation was extended by integrating elements from a hybrid image source/ray tracing model called “CAE-SAR” (Vorländer 1989; Schmitz et al. 2001). The key of the latter method is the parallel running of image source and ray tracing algorithm which delivers the specular reflections precisely and the diffuse reflections via the stochastic approach

²⁰http://www.ada-acousticdesign.de/set_en/setsoft.html

quantization with respect to source position and radiation angles

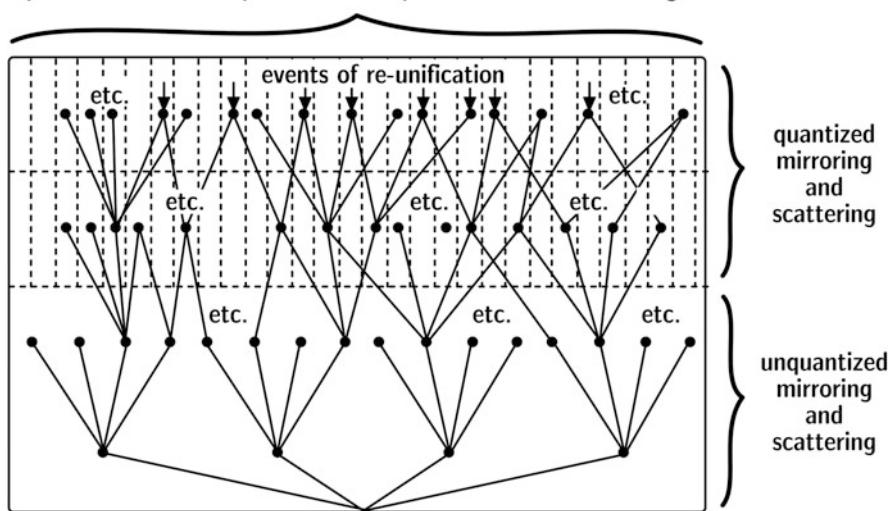
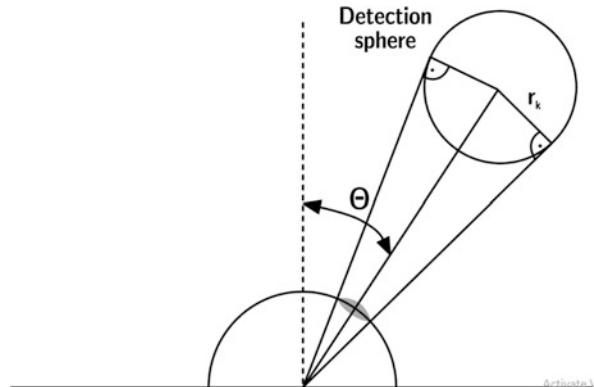


Fig. 11.37 Quantized Pyramidal Beam Tracing and corresponding pyramid tree. (After Stephenson 1996)

Fig. 11.38 Diffuse rain implemented for calculating the portion of scattered energy to a detector sphere (see (Heinz 1993) or the recently extended version in (Schröder et al. 2007), see also Fig. 16.7)



("diffuse rain" after Heinz (1993)) with lower spectral and temporal resolution in the reverberation tail. The main feature of this method is a closed solution of an inherent transition from early specular to late specular/diffuse reflections. It is not required to define transition orders or secondary sources (Fig. 11.38).

RAVEN SketchUp Plugin

The room acoustics simulation module called RAVEN (Room Acoustics for Virtual Environments) is implemented as an C++ library (Pelzer et al. 2014). It is available in open access for educational purposes.

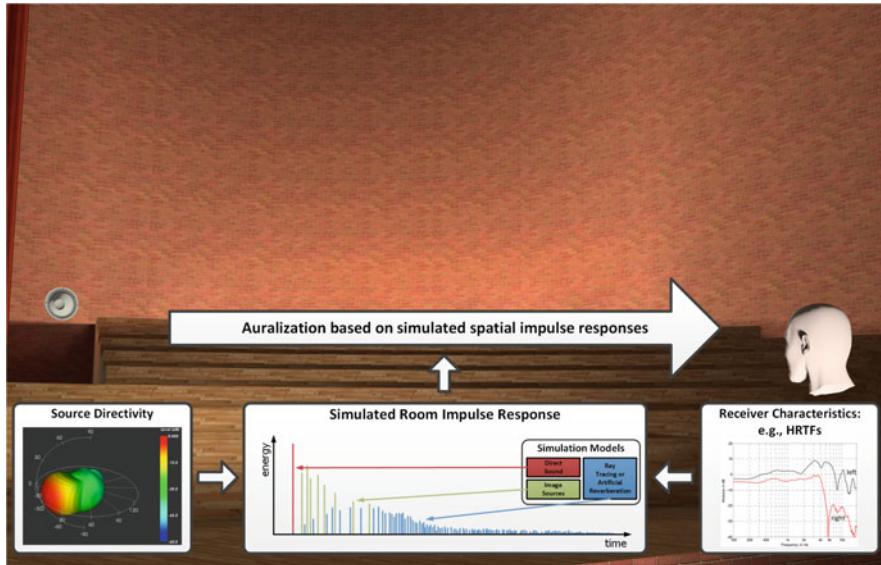


Fig. 11.39 RAVEN SketchUp plugin. (Vorländer et al. 2015)

RAVEN includes several optimization techniques such as multi-core parallelization and utilization of extended instruction sets. This makes it a high-performance library that can take advantage of the full computation power of modern CPUs. The most time-consuming part of the simulation is the ray tracing component, which takes some seconds of computation time, of course depending on the amount of required rays and the complexity of the room model. The calculation time can be reduced significantly by distributing the workload to two or more parallel threads, e.g. on a multi-core CPU. In order to constantly transfer the room model from the 3D modelling GUI to the acoustics simulation environment, a plug-in was developed and integrated into the 3D modeller “SketchUp” with full interactivity. Modifications in the model geometry or in the material properties are processed and auralized immediately. For the early part of the response, which is calculated by image sources, the update rate fulfills real-time performance, while the late response is updated in the order of magnitude of seconds (Vorländer et al. 2015) (Fig. 11.39).

State of the Art

Still today, the programmes listed above are subject to permanent development, not only concerning the component of impulse response computation but also other elements such as source and material databases. Most of them also offer signal processing tools, visualization tools and auralization tools in various reproduction formats.

Algorithms of geometric room acoustics modelling were also published by Farina (1995), Alarcão and Bento Coelho (2003), Lokki (2002) and Camilo et al. (2002), besides others. Many of these programmes were inspired by Kulowski (1985), Stephenson (1985), Vian and van Maercke (1986), Vorländer (1989) and Lewers (1993). In some methods, it was tried to build room impulse responses from reflection statistics (Vorländer 2000; Bento Coelho et al. 2001). Due to the rapid development in acoustic room simulation, however, this list cannot be complete; see also the excellent overview in Savioja and Svensson (2015).

State-of-the-art room-acoustic simulation is still based on geometrical acoustics. Wave models, however, are catching up. They require computation times which are higher by orders of magnitude, but their application is well possible for a limited frequency range.

11.8 Hybrid Wave and Geometrical Acoustics Models

According to the particle-wave dualism, acoustic room responses can be calculated both ways, with geometrical acoustics and with wave models (see Sect. 10.2). For small rooms and for objects in the room which cause explicit wave effects, geometrical methods cannot produce accurate results. This may apply to columns in the interior space, to the edge of an orchestra pit in an opera house or to the seat-dip effect in concert halls. The simple reason for the latter example is that the seats and the audience are not small compared with all wavelengths in the audio frequency range. For 100–200 Hz, where the seat-dip effect occurs, full wave-based models must then be chosen. Due to limitations of computation speed, maximum mesh sizes and memory space, however, the wave models are limited at higher frequencies. Hence, combinations of wave models with geometrical acoustics models were introduced (Kleiner et al. 1995; Summers et al. 2004). This is where the finite element method, FEM, the boundary element method, BEM, and the method of finite differences in time domain, FDTD, come into play.

Wave-based simulations can yield results between lowest frequencies to up to above the Schroeder frequency (Savioja 2010; Hamilton and Bilbao 2017; Hargreaves et al. 2019), so that an overlap in frequency range with results from geometrical acoustics can well be achieved. The results are then merged in a crossover network in the frequency domain and transformed back to time domain in order to obtain the room impulse response (Fig. 11.40).

A criterion which determines the crossover frequency where the simulation can be “switched” from the time-consuming wave models to geometrical acoustics, however, cannot be given in a general rule of thumb. As long as the room transfer function for the given source-receiver combination contains specific wave effects such as diffraction and modes (under a balcony, for instance) wave models are required. In the same room but at other source-receiver combinations, pure geometrical models may lead to results with acceptable accuracy.

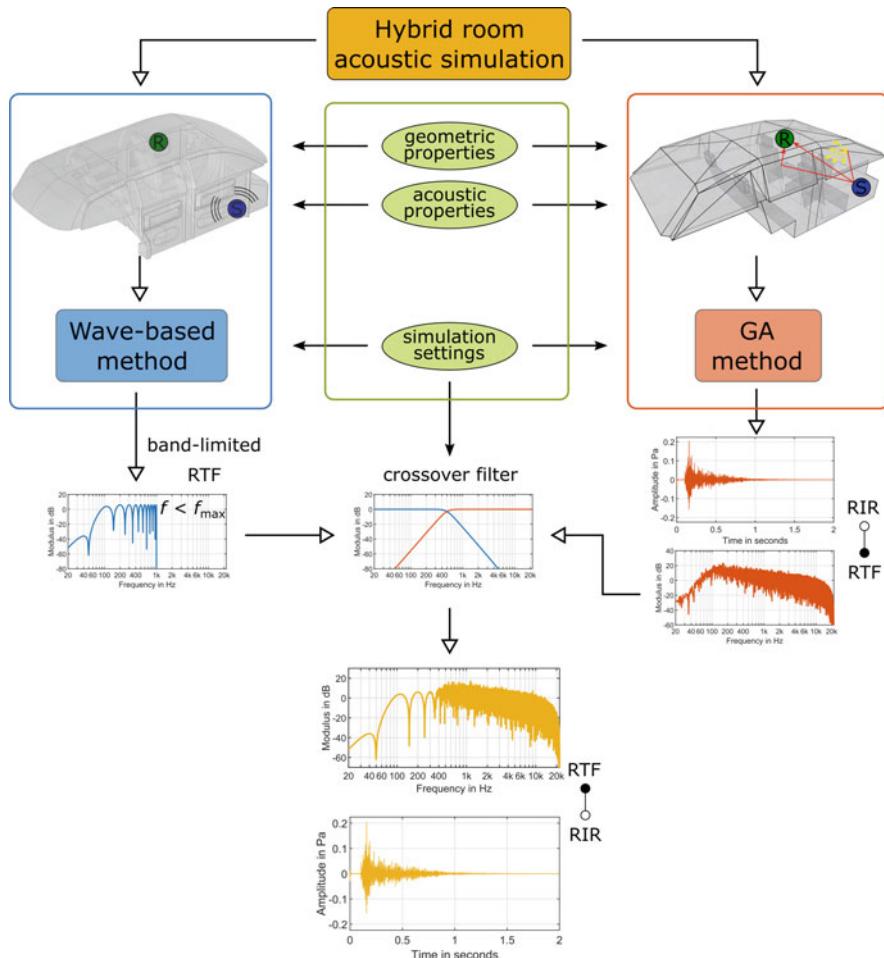


Fig. 11.40 Crossover network for merging results from wave models and from geometrical acoustics. (After Aretz and Vorländer 2014b)

An alternative is to pre-calculate room transfer functions and the corresponding impulse responses with wave models and to implement them in a lock-up table for real-time processing, which was mainly introduced for application in games.²¹

As concerns physics-based models with a demand for correctness in comparison with real space, the largest problem, however, is the identification of boundary conditions for wave models. Real constructions are much more complex than an analytic or numerical boundary condition model (porous material model, resonator model) can imply. A lightweight construction such as gypsum boards screwed on

²¹<https://docs.microsoft.com/en-us/gaming/acoustics/what-is-acoustics>

metal studs, for example, has a complex vibration pattern depending on the screws' positions, the rigidity of the connection, the stiffness of the metal studs, the airgap filling with mineral wool, etc. Its random incidence absorption coefficient can be estimated in standardized type testing for the construction as such but the "real" impedance boundary condition of a specific sample construction *in situ* may differ from the type specification quite a bit. When it comes to precision of simulation results, furthermore, it is not sufficient to model the construction in FEM or BEM with an impedance boundary condition since the construction is non-locally reacting with an impedance depending on the angle of incidence (see Sect. 3.2). Here, also FEM and BEM do produce systematic errors, and only a fluid-structure coupling between air and the construction would lead to better model results.

Compromises must be found, and whether these are appropriate depends very much on the specific case. Car passenger compartments, for example, are among the most difficult environments due to the complexity of surface materials (Aretz and Vorländer 2014a). Similar problems can be expected in train and airplane compartments. These examples demonstrate that wave models may be nominally accurate as they solve the wave equation exactly but the results in room acoustics are heavily affected by uncertain input data.

Now, that the basics of hybrid room acoustic modelling techniques are known, we can discuss the next step: the creation of impulse responses for auralization. This only is reasonable when including binaural or other kind of spatial technology.

11.9 Filter Construction of Spatial Room Impulse Responses

Impulse responses suitable for signal processing and particularly for auralization must have a sampling rate appropriate for the audio frequency range, typically about 40 kHz (Fig. 11.41). Thus, image source algorithms or hybrid models are applicable. The impulse responses are fed into FIR filters for convolution with dry source signals.

Room auralization shall be based on 3D audio formats. Otherwise, the spatial information would be lost. By using HRTF (see Sect. 6.4.1), the directional

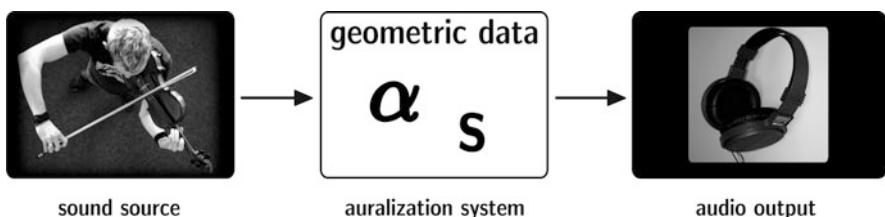


Fig. 11.41 Auralization of sound in rooms

information of a human listener is taken into account. Alternatively, the directional impulse response can be encoded in spatial audio formats such as ambisonics or VBAP; see 11.9.1.

During the historical development of the auralization of rooms, the first studies were based on purely specular reflections. After the principles had been proposed by Schroeder et al. (1962) and Schroeder (1973), one of the first who created a signal processing concept for room acoustics auralization was Pösselt (Pösselt et al. 1986), who used image source models for rectangular rooms.²² The model was further extended by Lehnert and Blauert (1989, 1992), Lehnert (1992) and Vian and Martin (1992). The purely specular reflection model, however, was not sufficient, due to lack of scattering, as we know today. But his important and novel contribution was the rigorous description of the room impulse response on the basis of spatial direct sound and reflections. With this deeper understanding of the details in room impulse responses and their relevance for auralization available to other researchers, extended model approaches including hybrid models and scattering models in particular have been developed.

Since data of air attenuation and particularly the absorption coefficients of walls are available in frequency bands, the spectral representation required for the impulse response components must be created by interpolation.

Interpolation

Several interpolation algorithms are possible candidates. The cubic spline interpolation is one example for a well-qualified method to create interpolated data. For interpolation of reflection factors from octave band data (see also 11.2.2), effects of interpolation in frequency and time domain must be discussed. Too rough interpolation may cause audible ringing. Onset and decay of the interpolation filter creates a kind of tonal reverberation.

Mommertz (1996) showed that absorption coefficients in one-third octave bands serve well for the reconstruction of the complex reflection factor. While the modulus of the reflection factor can be obtained directly, a plausible phase must be added. Minimum or linear phases are options. Only for very few cases, like for focuses, the specific choice will be important (Fig. 11.42).

After inverse Fourier transformation, a spectrum consisting of M complex coefficients corresponds to a signal of M samples. Each convolution of the signal and transfer function of M lines in the frequency domain creates a result of M lines, too. In contrast, in time domain convolution, the result contains $2M$ samples (Fig. 11.43).

Reverberation Tail

As explained above, stochastic effects such as scattering play an important role particularly in the late response. Starting with the idea of spatial and temporal reflection patterns, the contributions of the scattered and late part of the impulse response are represented by a set of noise-equivalent stochastic reflections H_j , the

²²The reason for this is clearly the limitation by processing time in the late 1980s.

Fig. 11.42 Step (dotted line), linear (broken line) and spline interpolation of spectra based on octave band data (example)

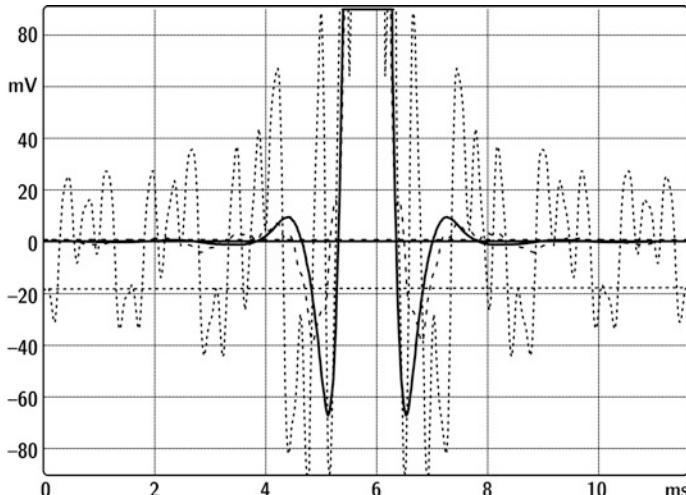
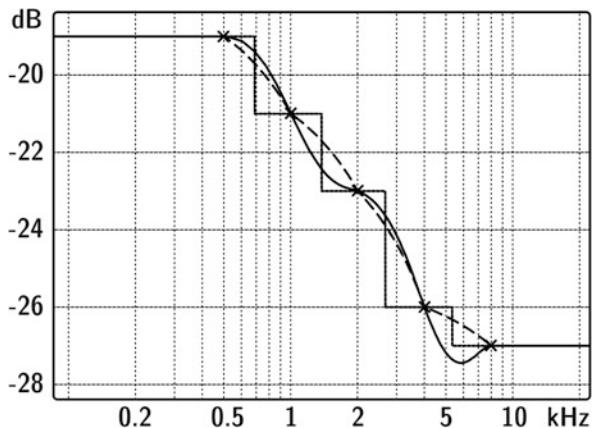


Fig. 11.43 Result of an interpolation of reflection factors created from one-third-octave band absorption coefficients. Plot in time domain (maxima clipped). Spline interpolation between one-third octave bands corresponds to the continuous line. Too rough interpolation (step function) corresponds to the dotted line (ringing effects)

arrangement of which must be constructed in a filter model. The basis for this kind of construction may be stochastic ray tracing, radiosity, free path statistics or an artificial reverberation process. All methods mentioned yield estimates of the late impulse response envelope, $M^2(\Delta t, f)$, a function of frequency and time.²³ By adding

²³Similar to a modulation function.

an adequate fine structure which represents the actual reflections statistics, the binaural impulse responses can be created (Heinz 1993; Schröder 2011) by “noise shaping”.

The parameters of the fine structure are the average amplitudes, \underline{H}_j , and the density, n_j , of the reflections in the time interval Δt . The set of reflections yield the correct modulation if

$$M^2(\Delta t) = |\underline{H}_j|^2 \cdot n_j \quad (11.52)$$

for all frequency bands.

The choice of the specific algorithm for estimating the late response, the bandwidth, the size of the time intervals and spatial resolution of the HRTF are subject of research and development in many places. It is clear that fine details of the late response can usually not be perceived. If, however, the room creates an echo in the late response and this was not detected and modelled in the simulation, the auralization will miss an important feature of the room. Hence, there exists no simple rule on how the parameters of the late response and the transition time between the early and the late response must be chosen.

So far, in this chapter it was referred to the filter construction for sound pressure impulse responses which corresponds to measurements with omnidirectional pressure microphones. As soon as the spatial information shall be taken into consideration, spatial audio standards are introduced into the room impulse response filter; there are two fundamental approaches possible, as will be described in 11.9.1 and 11.9.2 (Fig. 11.44).

11.9.1 Encoding of Spatial Room Impulse Response Filters

A spatial impulse response is created from direct sound, early reflections, scattered components and reverberation by using the concept of plane wave decomposition (Sect. 2.6) in a spatial audio format such as spherical harmonics, for example. In the formulation in the frequency domain,²⁴ this process is described by the multiplication of the plane-wave filter functions representing sound travelling from the source to the receiver via various paths until it arrives at the receiver under the angles ϑ, φ .

In such spatial audio formats, the receiver is a spherical microphone array. The incoming components of the impulse response contributions arriving at the microphone array are decomposed one by one and finally added in a superposition process. The amplitude vector, B_n^m , of a ray from the direction of ϑ, φ reads (Daniel 2000)

²⁴In time domain the same can of course be formulated by convolution, but the spectral product terms are more comfortable to interpret.

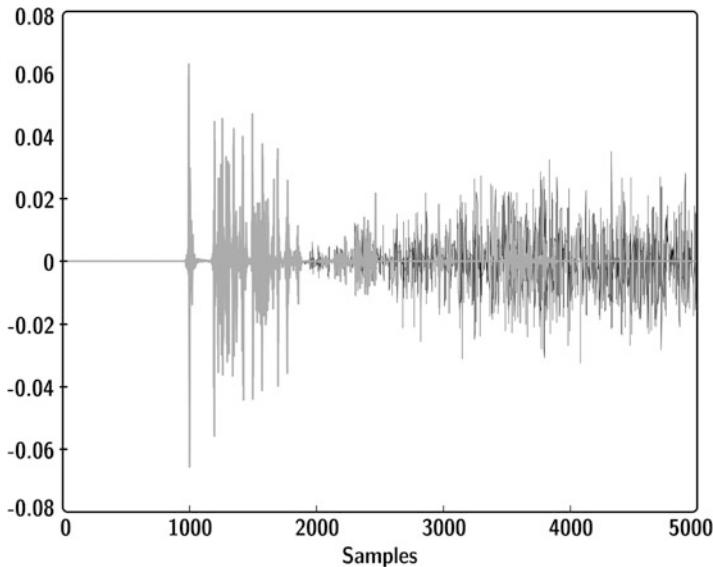


Fig. 11.44 Omnidirectional sound pressure room impulse response created from specular (grey) and diffuse (black) reflections. (Example after Schröder et al. 2007)

$$B_n^m = 4\pi \sqrt{-1}^n Y_n^m(\vartheta, \varphi) p(\vartheta, \varphi) \quad (11.53)$$

$$p(\vartheta, \varphi) = \frac{e^{-j\omega t_j}}{ct_j} \cdot \underline{H}_{\text{source}}(\theta, \phi) \cdot \underline{H}_{\text{air}} \cdot \prod_{i=1}^{n_j} \underline{R}_i \quad (11.54)$$

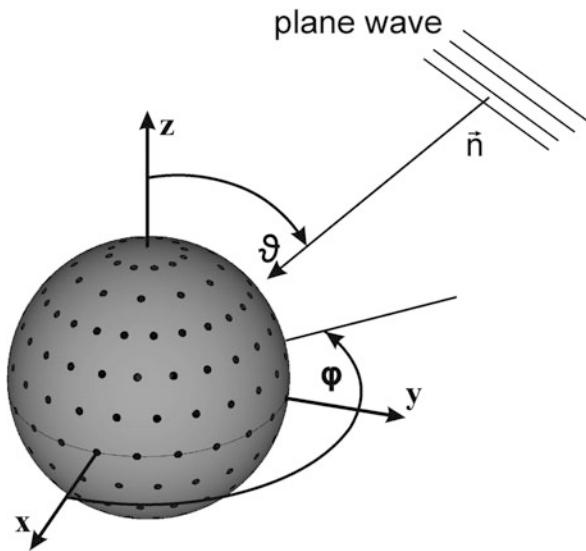
with $p(\vartheta, \varphi)$ denoting the sound pressure spectrum of the j th reflection, t_j its delay, $j\omega t_j$ the phase, $1/(ct_j)$ the distance law of spherical waves, $\underline{H}_{\text{source}}$ the source directivity in source coordinates, H_{air} the low-pass filter of air attenuation and \underline{R}_i the reflection factors of the walls involved.

In higher-order ambisonics, HOA, for example (see 17.3.2), the N -channel signal vector is encoded by using the spherical harmonic coefficients of the specific directions of all impulse response components. Finally, in decoding the signals for reproduction, a surrounding loudspeaker array radiates signals represented by the continuous spatial functions of spherical harmonic base functions, and all loudspeakers together form an approximation of the integral over the surrounding sphere. This corresponds to a discrete Spatial Fourier synthesis with a truncated version of the series of spherical harmonics.

The amplitude vector of the N -channel encoding format in spherical coordinates in a specified geometric arrangement of a (virtual) spherical microphone array is

$$\mathbf{B} = S\mathbf{Y} \quad (11.55)$$

Fig. 11.45 Encoding of a plane wave (ray) into a multi-channel spherical array format



Y denotes the vector of the real-valued spherical harmonics, **S** the source signal and **B** the vector of driving amplitudes of the N loudspeakers. The result of reconstruction of the incident ray is (see also 2.8) (Fig. 11.45)

$$p(\theta, \phi) \approx \sum_{n=0}^N \sum_{m=-n}^n \sqrt{-1}^n j_n(kr) B_n^m(\omega) Y_n^m(\theta, \phi) \quad (11.56)$$

In panning techniques such as VBAP, the N -channel loudspeaker amplitude vector is derived from weights according to the phantom source generation in triangular loudspeaker subsets. The total field is obtained by superposition of all impulse response components involved. More details about HOA and VBAP techniques are given in Sect. 17.3.2.

11.9.2 Binaural Room Impulse Response Filters

In binaural formats the receiver is a human head with its two ears. A binaural impulse response is created from direct sound, early reflections and scattered components by using the concept of binaural synthesis (Sect. 9.4). All components are added. In the formulation in the frequency domain,²⁵ this process is described by the

²⁵In time domain the same can of course be formulated by convolution, but the spectral product terms are more comfortable to interpret.

multiplication of the transfer and filter functions representing sound travelling from the source to the receiver:

$$\underline{H}_j|_{l,r} = \frac{e^{-j\omega t_j}}{ct_j} \cdot \underline{H}_{\text{source}}(\theta, \phi) \cdot HRTF(\vartheta, \varphi)|_{l,r} \cdot \underline{H}_{\text{air}} \cdot \prod_{i=1}^{n_j} \underline{R}_i \quad (11.57)$$

with H_j denoting the spectrum of the j th reflection, t_j its delay, $j\omega t_j$ the phase, $1/(ct_j)$ the distance law of spherical waves, $\underline{H}_{\text{source}}$ the source directivity in source coordinates, H_{air} the low pass of air attenuation, \underline{R}_i the reflection factors of the walls involved and HRTF the head-related transfer function of the left (l) and right (r) ear of the sound incidence in listener coordinates at a specified orientation.²⁶

²⁶Usually, the listeners are oriented towards the original source or a similar reference point.

Chapter 12

Simulation and Auralization of Outdoor Sound Propagation



12.1 Introduction

Auralization of outdoor sound propagation models such as noise mapping is particularly challenging due to wave phenomena which can hardly be solved exactly following the wave equation. Therefore, diffraction and refraction require approximations which are adequate in terms of accuracy but still acceptable as concern computation times. Outdoor scenarios and additionally moving sound sources such as vehicles pose challenges which require special attention. This chapter deals with the question how outdoor sound propagation models and noise mapping models can be enhanced and adapted for auralization.

12.2 Environmental Sound

Environmental noise negatively affects the comfort, well-being and health of the population in urban and rural environments. The assessment of noise impact on society is a topic that involves technical and psychoacoustic aspects. Tools of simulation and auralization enable noise control engineers to present to the various stakeholders of a society a realistic forecast of possible scenarios of an intervention. Similarly, the auralization tools complement these virtual scenarios, by producing the equivalent three-dimensional sound at the site of observation.

Also here, the objective is to enhance the current process of acoustic engineering through advanced noise simulation and auralization tools. Currently, this process is based on the offline simulation and prediction of A-weighted noise levels. Commercial solutions are available and are effectively put into action by acoustic engineers. The well-established prediction tools generate tabularized noise data at sensitive receiver positions and/or depict noise distribution as visualized maps. Such an outdoor sound prediction can be based on a simplified free-field sound propagation

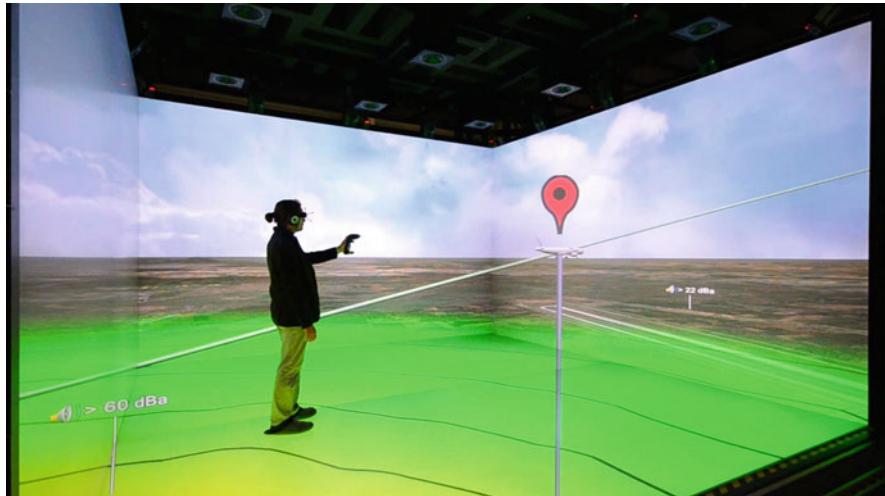


Fig. 12.1 Real-time simulation of outdoor sound propagation with audio-visual VR scene and illustration of the noise footprint on the ground. (Sahai et al. 2016)

model such as ISO 9613-2 and appropriate noise mapping software such as Harmonoise (Salomons et al. 2011). Noise mapping became very popular after national and regional governmental organizations required noise assessment and actions against environmental noise. The adherence to international standards assures a powerful and necessary step during the process of decision making in the context of acoustic pollution. Accordingly, software is available on the market, which in various ways can be used to predict noise levels averaged of daytime, nighttime, etc.

However, to understand the meaning behind those results, basic knowledge on the topic is required. To overcome this obstacle in participation of the public and to include non-experts into the planning stage, advanced noise simulation and auralization tools are very helpful.

Noise mapping involves a landscape or city model including reflection and diffraction models, atmospheric conditions and appropriate source data. Typically all effects can be considered by using energy models and corresponding attenuation factors in frequency bands (Fig. 12.1).

For given noise source positions of vehicles, aircrafts, etc. and a listener within the scene, the intersected volume elements (voxels) on the direct line between both objects are determined within the meteorological model. Similarly to the flanking path energy, the relevant paths connecting the source and the listener, their filter functions and their contribution to the total sound field are superposed (added) with their attenuation factors.

The result of standardized noise mapping, however, does not yield impulse responses. The temporal features such as the propagation delays and corresponding 3D filters must be created similarly as described above, for room acoustics.

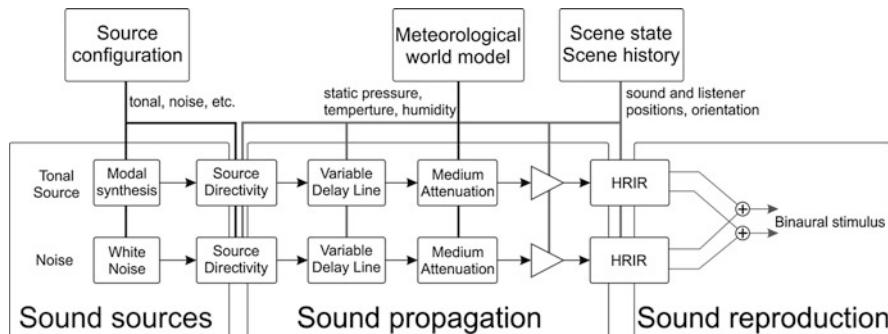


Fig. 12.2 Auralization of outdoor sound

Accordingly ray tracing and image sources and all other numerical methods yielding energy and delays of propagation paths are candidates to determine the physical behaviour of sound in a virtual scenario. The simulation is used to predict transmission paths from a source to a receiver in an environmental situation, which is entirely based on digital input data. This is basically identical with the situation in indoor scenarios (Fig. 12.2).

The main challenges are, however, different between indoor and outdoor simulations. Outdoor scenarios must be considered a large number of relevant propagation paths including diffraction (see 12.3). The number of outdoor sources (street, rail and air traffic) may be significantly larger than in room acoustics, the sources are typically moving (see 12.4), and the atmosphere is not homogeneous and time invariant (see 12.5).

12.3 Implementation of Diffraction Models

Diffraction is an important wave phenomenon that can be observed at sound propagation, when finite objects are located within the vicinity of sound source and recipient. In urban environments, for example, diffraction of sound waves plays a significant role, because a large amount of the incoming sound energy is partly or even fully determined by diffraction. Emergency vehicles with horns, airplanes and helicopters, automobiles and trucks as well as railway vehicles, which can all be heard in distant street canyons – even though they are not visible at the receiver position – are examples for signal sources with high amplitude and no direct line of sight to the listener.

In simulation and auralization of virtual environments, geometrical methods for the simulation are commonly used, which theoretically incorporate the transfer path from sound source to sound receiver on a direct path along the direction of propagation of a plane sound wave. The simulation of the diffraction of sound (which is a wave-based phenomenon) within energy-based geometrical acoustic methods

increases the required computational effort, as will be outlined in the following. Hence, approximations of sound diffraction calculations are required as the discrepancy between high-quality simulation, and required computing time does not allow a satisfying auralization of scenes with outdoor noise.

The fundamentals of diffraction phenomena were presented in 3.5. The integration of approximations of diffraction models into ray tracing was developed (Stephenson 2004b) and later confirmed (Pohl 2014), who showed that Stephenson's model can efficiently be integrated into ray tracing.

For real-time auralization of outdoor noise applications, however, extensions of the image source model are more promising, if they are extended by path search algorithms and by filters representing frequency-dependent diffraction at edges. Analytic methods (Biot-Tolstoy, Medwin, Svensson, etc.) that can fully solve this problem, however, are computationally intensive. Therefore, to simplify the problem, the first step is to find the major propagation paths based on a scene's geometry. Concentrating only on the relevant audible contributions will decrease the computational effort to derive diffraction filter that represent the acoustic sound propagation through the auralized scene. Based on this principle, the next step is the efficient implementation and handling of the previously found sound propagation paths. This means identifying a basic diffraction object (e.g. single finite wedge) the complex scenarios (individual paths potentially including diffraction at multiple wedges). These basic diffraction objects have their algorithmic counterparts in likewise basic operations, which will be solved efficiently using shared-memory parallelization. Based on the complexity and the time when results need to be available, this planning instance partitions the previously identified major propagation paths into its elementary diffraction objects and recombines the distributed basic operations' results to form the best solution. This solution reflects possible real-time constraints and the necessary auralization quality in virtual reality applications.

As an example, diffraction in a complex urban environment consisting of buildings can be discussed. As soon as the faces of the buildings, their edges and vertices are defined, geometric construction algorithms such as the edge rotation algorithm or the image edge method can be used to identify path candidates which connect the receiver and the source.

Whether or not paths lead to audible sound contributions depends on several factors. Physically, diffraction paths may be associated with very small energy compared with direct sound or reflection paths; perceptually, they may be neglected due to spectral masking. Culling methods are, therefore, required in order to reduce the paths to the absolutely relevant contributions in which complex and time-consuming diffraction filters (e.g. unified theory of diffraction, UDT) shall be introduced.

Figure 12.3 shows an example of this kind of complexity reduction. Here, the relative and the absolute sound level of the contribution are predicted from the overall distance between source and receiver, and the level reduction of diffraction is estimated from the diffraction angles (Erradj 2019).

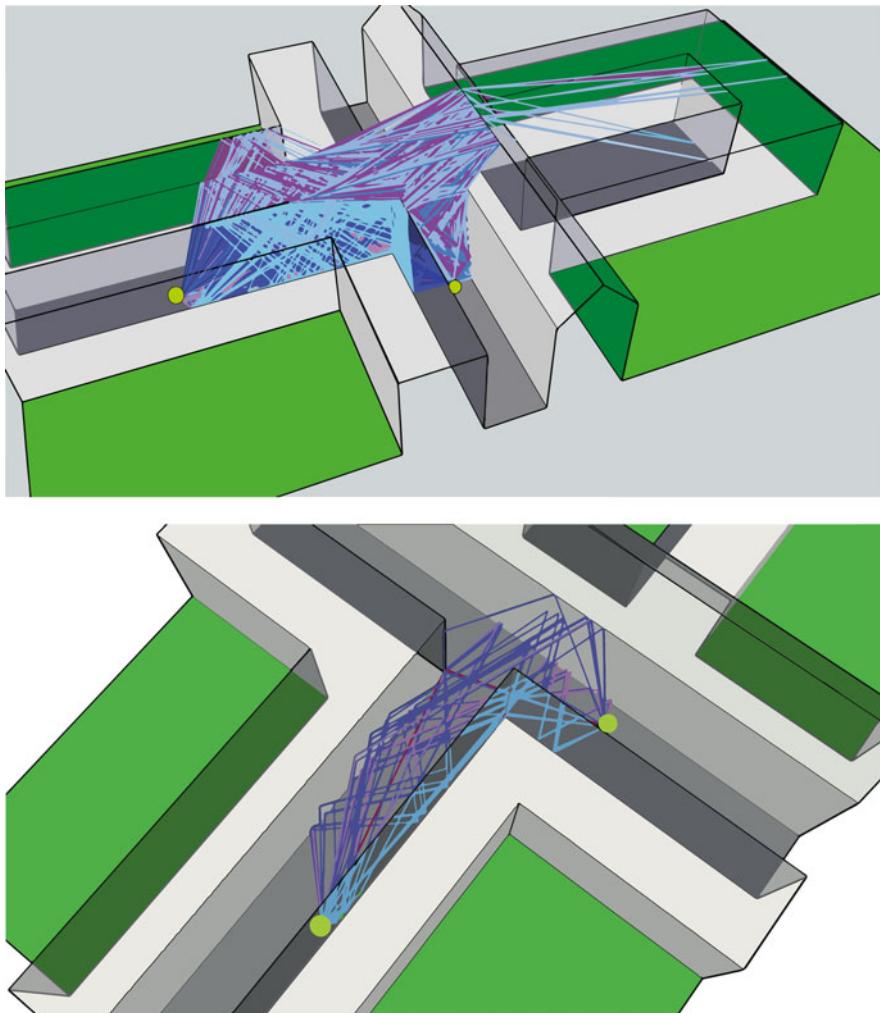


Fig. 12.3 (top) All paths up to fourth order and (bottom) remaining propagation paths in an example scenario after perceptual path length culling

In order to validate pathfinder and diffraction filters, a reference case from the round robin on auralization was used (“scene 5” (Brinkmann et al. 2020); see also Sect. 18.5). The impulse responses measured at a distance of 2 m are shown in Fig. 12.4. The simulated comb filter structure ranging from about 300 Hz up to 2 kHz is very close to the measured result which proofs that the reflection and diffraction filters represent the propagation phenomena well.

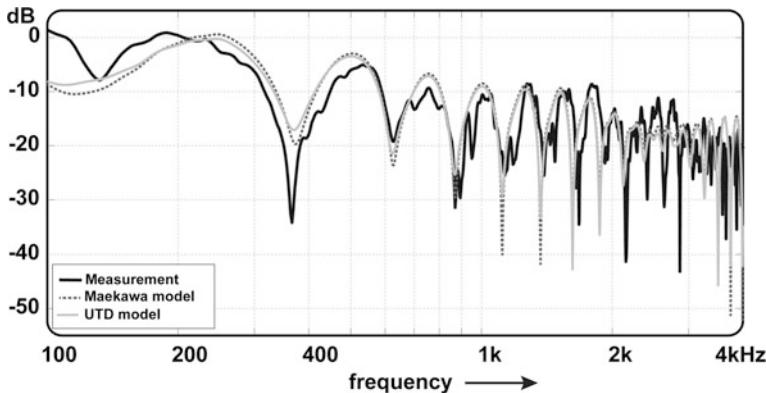


Fig. 12.4 Sound propagation filter for source and receiver separated by a screen, equivalent to “scene 5” of Brinkmann et al. (2020)

12.4 Moving Sources

In outdoor noise scenarios and particularly in traffic noise simulations, moving sound sources may be present, some of which at very high speed (high-speed trains, aircrafts). This causes frequency shifts, the so-called Doppler effect; see 3.8. Depending on the speed, the Doppler effect can be clearly audible, and, thus, it must be simulated properly.

For the auralization in general, linear time-invariant systems are considered, which is an obvious contradiction with the time-variant (moving) source. The LTI processing by convolution, or filtering in general (see 7.3), must be revisited and extended with appropriate updating and fading. Windowing and crossfading are techniques for creating an output signal with smooth waveforms.

The crucial quantity to describe the wave propagation at time-variant source-receiver distances is the effective propagation delay (Strauss 1998), which delay the signal by $\tau(t)$. The spatial motion of sources and receiver can be parametrically described by trajectories $\vec{r}_r(t)$ and $\vec{r}_s(t)$. For an auralization it is needed to know which sound wave arrives at the listener at a specific time t and position $\vec{r}_r(t)$. The propagation delay, $\tau(t)$, is calculated from the time of arrival t at the position $\vec{r}_s(t)$. The retarded time $t' = t - \tau(t)$ is the radiation time at a (previous) source position $\vec{r}_s(t')$. This leads to a fundamental condition, stating that the Euclidean distance between the point of radiation and reception match the effective propagation distance and delay (Wefers and Vorländer 2015).

$$\|\vec{r}_r(t) - \vec{r}_s(t - \tau(t))\| = c\tau(t) \quad (12.1)$$

The propagation delay $\tau(t)$ must be calculated from the trajectories of source and receiver. In case of a simple linear trajectory and a fixed receiver at \vec{x}_r , for example, this leads to

$$\vec{r}_{sr}(t) = \vec{r}_r(t) - \vec{r}_s(t) = \vec{x}_r - \left(\vec{x}_s + \vec{v}_s t \right) \quad (12.2)$$

This can be solved in general for the propagation delay into

$$\tau(t) = \frac{\vec{r}_{sr}(t) \cdot \vec{v}_s + \sqrt{\left(\vec{r}_{sr}(t) \cdot \vec{v}_s \right)^2 + \left(c^2 - \|\vec{v}_s\|^2 \right) \|\vec{r}_{sr}(t)\|^2}}{c^2 - \|\vec{v}_s\|^2} \quad (12.3)$$

The problem is now to estimate $\tau(t)$ from the analysis of the trajectories for the specific motions of source and receiver. For fixed scenarios this can be easily done in preprocessing. But it is clear that this is a significant challenge for interactive real-time auralization, when no a priori knowledge about the motion is available. More is discussed in 16.8.

Another challenge is to predict the propagation delay with sufficient accuracy as concerns the absolute value $\tau(t)$ and its derivative $\partial\tau(t)/\partial t$, in order to avoid artefacts from discontinuities in the interpolated samples.

The filter implementation can be done by using variable delay lines, VDL. VDL systems process the delays also in sub-sample ranges within the processing frames in order to avoid quantization artefacts (Savioja et al. 1999). At increasing distance between source and receiver (increase of $\tau(t)$), the discrete samples on the time axis are stretched, which corresponds to a compression of the discrete frequency spectrum and, hence, to a downwards frequency shift, the Doppler effect (see also 3.8). Vice versa, for source approaching the receiver, the samples are compressed, while the spectrum is stretched.

The resulting signal must now be re-arranged in the nominal sampling scheme. Technically, this process is equivalent to a band-limited interpolation of the samples in a ring buffer and is, hence, analogous to an asynchronous sample rate conversion, ASRC.

12.5 Sound Propagation in the Atmosphere

Outdoor sound propagation modelling usually incorporates the propagation effects of geometric (spherical) spreading, atmospheric absorption according to ISO 9613-1 and ground reflection according to impedance models. For the sound attenuation in the atmospheric, humidity and temperature, dates are the basis for calculation of the absorption in frequency bands. Furthermore, the inhomogenous medium and possibly wind cause distortions of sound propagation paths (Piercy et al. 1977). In a ray tracing model, for instance, this results in curved rays. In a layered medium, such as in the atmosphere, refraction and wind cause ray curvature (Fig. 12.5).

In more detailed modelling, the climatological input data of the atmospheric conditions (temperature, wind) lead to the effective sound speed profile as a function of the height aboveground. With this information refraction effects can be calculated

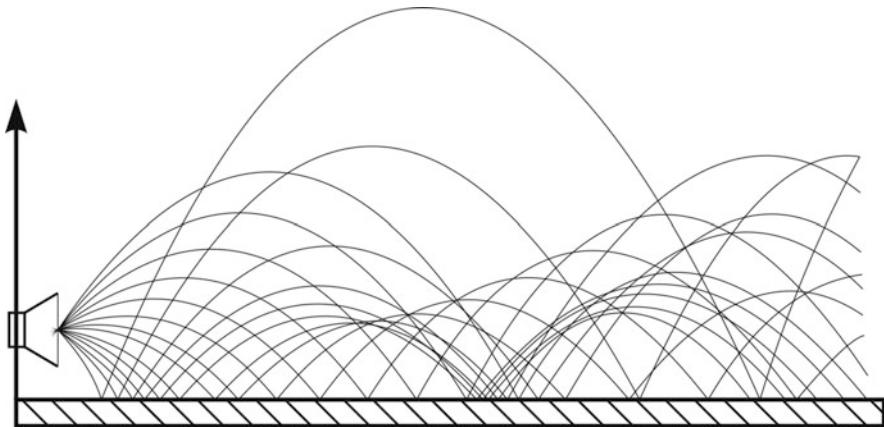


Fig. 12.5 Curved sound paths in the atmosphere

(see also 3.6), which happen to result in upward or downward propagation path curvature. Although powerful wave propagation models are available, the ray model is very useful for illustrating the propagation paths. A ray in this sense is still the (eikonal equation) solution and can be interpreted as following the Fermat's principle, yielding the path with the shortest propagation delay between source and receiver.

This can be illustrated by introducing a modified wave equation

$$\frac{\partial^2 p'}{\partial r^2} - \frac{1}{c_{\text{eff}}^2} \frac{\partial^2 p'}{\partial t^2} = 0 \quad (12.4)$$

with an effective sound speed, which depends on the wind vector, \vec{v} , and \vec{n} the normal of the wave (ray) direction in the vertical plane

$$c_{\text{eff}} = c + \vec{v} \cdot \vec{n} \quad (12.5)$$

In more rigorous numerical techniques, finite difference time domain solutions (FDTD) and parabolic equation (PE) models are used either in a volume-based (voxel) discretization of the wave equation directly or as a modification of the wave equation, respectively (Ostashev 1997; Salomons 2001; Blanc-Benon et al. 2001). According to Whitaker and Norris (2008), PE methods are typically derived from the Helmholtz equation in 2D cylindrical coordinates:

$$\frac{\partial^2 p(r, z)}{\partial r^2} + \frac{1}{r} \frac{\partial p(r, z)}{\partial r} + \frac{\partial^2 p(r, z)}{\partial z^2} + k_0^2 n^2(r, z) p(r, z) = 0 \quad (12.6)$$

with k_0 denoting ω/c_0 , $n(r,z)$ the refraction index acc. to Sect. 3.6. By substituting the pressure as an envelope function, $\psi(r,z)$, multiplied with the Hankel function, the wave equation for the field gets the parabolic equation:

$$\frac{\partial^2 \psi(r,z)}{\partial r^2} + 2jk_0 \frac{\partial \psi(r,z)}{\partial r} + \frac{\partial^2 \psi(r,z)}{\partial z^2} + k_0^2(n^2 - 1)\psi(r,z) = 0 \quad (12.7)$$

The sound pressure is proportional to $1/\sqrt{r} \cdot \psi(r,z)$. Approximations of the solutions of the parabolic equation for the vertical axis z take the form (Whitaker and Norris 2008)

$$\psi(r + \Delta r, z) = e^{jk_0(Q-1)}\psi(r, z) \quad (12.8)$$

where cylindrical shells are stepped towards larger distances r .

Turbulences are local effects of air movement in so-called “eddies”, the size of which is typically metres. These eddies are discontinuities in the medium and cause stochastic scattering. For example, auralization of aircraft noise propagating through the atmosphere is studied at the National Aerospace Laboratory (NLR) of the Netherlands (Arntzen et al. 2014). The main sound propagation effects such as refraction and attenuation are important, but also stochastic atmospheric effects such as wind and turbulence are identified to enhance the sound perception and the plausibility.

12.6 Auralization from Noise Mapping

Auralization has already been applied as a part of the noise mapping approach in many research initiatives. For example, in the EU project called “Sonorus” (Kropp et al. 2016), experts collaborated according to the following quote from the project website “to conduct a research programme that addresses the key issues of modern urban planning with the goal of achieving a good acoustic, to implement a holistic approach to the planning of acoustic environments including diverse areas such as city planning and traffic, to enhance the European knowledge economy by providing trained and mobile researchers equipped with the skills necessary”. The key problem still today is input data acquisition and processing.

The state of the art in sound propagation simulation arrived at a quality and feasibility that urban spaces can be modelled in the computer with high accuracy indeed. The resulting noise maps are prerequisite for urban planning. Experience shows, however, that data mining, programming and a high-quality audio reproduction for auralization is very challenging, tedious and time-consuming and thus very costly, in particular when it comes not only to visual display of noise maps but to auralization.

Accordingly, efforts were taken in various initiatives of harmonizing the source and propagation models and the analysis tools. The so-called “Harmonoise” model was an important outcome (Salomons et al. 2011) in Europe. It is an extended version of propagation models for all propagation phenomena described above: ground reflection, diffraction, air attenuation and refraction in the atmosphere. With appropriate source models or measured source data, noise maps can be created, and with filters of spectral and temporal features for each of the propagation phenomena, the auralization process can be completed. Finally, the implementation of the Common Noise Assessment Methods in Europe (CNOSSOS) (Kephalaopoulos et al. 2012) ensures that software can be evaluated on standardized test cases.

Source data for trains, road vehicles and air vehicles, however, are the critical part; see Chap. 8. General solutions don’t exist, although the physical model approach of primary excitation of wheel or tyre interaction with the contact surface is straightforward (Pieren et al. 2017; Maillard et al. 2019). The quality depends strongly on the assumptions of material properties and their stochastic features of real cases.

Chapter 13

Simulation and Auralization of Airborne Sound Insulation



13.1 Introduction

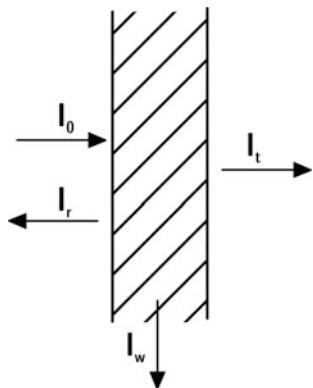
As previous chapters dealt with virtual environments with sound in free-field and in enclosures (room acoustics, Sect. 11.7), situations with partition walls and complete buildings will be discussed in this chapter. This concerns the field of classical building acoustics as part of architectural acoustics. A typical example dealt with by building acoustics is the problem of sound transmitted into a so-called receiving room from the room next door (source room), for instance, sound from a stereo set that is transmitted into the neighbouring sleeping room. The fundamentals of building acoustics and its auralization are described in this chapter. Another example is impact noise created by walking on the floor in the room above, a typical example of a structure-borne source, which will be treated in Chap. 14. Furthermore, sound generated by building service equipment like heating or air condition systems is of interest. Noise from outside the building, like traffic noise or industrial noise, concerns the sound insulation of façades.

13.2 Building Acoustics

Building acoustics is the discipline of sound and vibration transmission in buildings. In contrast to room acoustics, structural acoustics plays an important role in sound insulation. The models used are general models of structure-borne sound. The principles and methods discussed here can also be used in other fields like vehicle acoustics or noise control. This will be discussed later.

In typical room-to-room situations, the signal received sounds quiet (fortunately) and dull. An auralization has to reproduce these properties (level and colouration) in first place.

Fig. 13.1 Sound transmission through a wall



Our goal in this chapter is to auralize the character of *sound transmission* between spaces separated by structures.

Specific spatial effects in the receiving room, such as early lateral fraction, are not relevant. We will therefore focus on the basic quantities used in building acoustics today. Based on these quantities, strategies for designing auralization filters will be introduced. Sound reduction indices or standardized level differences will serve as input data for auralization. They may be obtained by calculation or by standard building acoustic testing. That is why standard building acoustics have to be introduced in detail.

13.3 Definitions of Airborne Sound Transmission

The basic quantity to describe the performance of construction elements¹ is the sound reduction index. It is defined as 10 times the logarithmic ratio of the incident sound intensity, I_0 , and the transmitted intensity, I_t .

$$R = -10 \log \tau = 10 \log \frac{I_0}{I_t} \quad (13.1)$$

with τ denoting the transmission coefficient, $\tau = I_t/I_0$ (Fig. 13.1).

In a free space with the partition separating two domains, R is identical with the sound pressure level difference, $D = L_S - L_R$. In buildings, however, the sound

¹Construction elements are typically walls, floors, doors, windows, etc.

reduction index is related not to plane waves at one specific angle of incidence but to diffuse and reverberant sound fields. This yields two consequences: (a) the angle of incidence must be averaged and (b) reverberation changes the level difference between the source and receiving room.

The sound power, P , will create a diffuse sound field in the source room with energy density²

$$w_S = \frac{4P}{cA_S} \quad (13.2)$$

with A_S denoting the equivalent absorption area in the source room. The energy density will lead to an irradiation strength, B , of the partition wall. The irradiation strength is identical to the incident sound intensity, I_S , in the source room.

$$B = I_S = \frac{c}{4} w_S \quad (13.3)$$

By definition, the sound reduction index is the ratio (in decibels) of the incident and the transmitted intensity. Thus, the power radiated from the partition walls is

$$B = S \cdot I_R = \frac{cA_R w_R}{4}. \quad (13.4)$$

From Eq. (13.3) and Eq. (13.4), we can derive the sound reduction index, R , for two adjacent rooms with diffuse sound fields:

$$R = 10 \log \frac{I_S}{I_R} = 10 \log \frac{w_S}{w_R} + 10 \log \frac{S}{A_R} = L_S - L_R + 10 \log \frac{S}{A_R} \quad (13.5)$$

These definitions are most simple if the sound is transmitted via exactly the one element the transmission coefficient is referred to. In building practice, however, the situation is much more complicated. Sound and vibration and their coupling create energy transmission over multiple paths. Therefore we have to divide the problem into the elements involved. Building elements have finite size. They are coupled in a specific way, rigidly coupled or mounted with elastic connections. All these conditions have specific effects on the total sound insulation. Before we will discuss the details of element coupling, we will study the basic effects of single partition sound transmission.

²The direct field is neglected, thus we assume the distance to be larger than the reverberation distance.

13.4 Sound Insulation of Building Elements

Partition walls react to an excitation with airborne sound waves. The adequate parameter to describe the amount of movement caused by sound pressure (or force) is the wall impedance. Generally, the impedance may contain effects of inertia, of stiffness and of damping. Furthermore, excitation and radiation depends on the size of the building element, as explained in 5.3.1.

A two-port model (Sect. 10.3) of the partition wall separating two domains, source side and receiving side, offers a very effective and illustrative study of basic effects. The partition wall is accounted for by its impedance, Z_p (Fig. 13.2).

The ratio of the transmitted pressure side to the sound pressure of the incident wave is

$$H_{\text{trans}} = \frac{p_0}{p_2} = \frac{1}{2} \left[a_{11} + a_{22} - \rho_0 c a_{21} - \frac{a_{12}}{\rho_0 c} \right], \quad (13.6)$$

with a_{ij} denoting the matrix elements of the two-port (see 10.3),

$$\underline{a} = \begin{pmatrix} p_1 & \frac{\det Z}{p_2/v_1} \\ p_2 & p_2/v_1 \\ \frac{v_1}{p_2} & \frac{v_1}{v_2} \end{pmatrix}. \quad (13.7)$$

In this formulation, the sound reduction index of the partition is

$$R = 10 \log |H_{\text{trans}}|^2. \quad (13.8)$$

Detailed information about the content of the two-port yields the matrix coefficients and the transmission function.

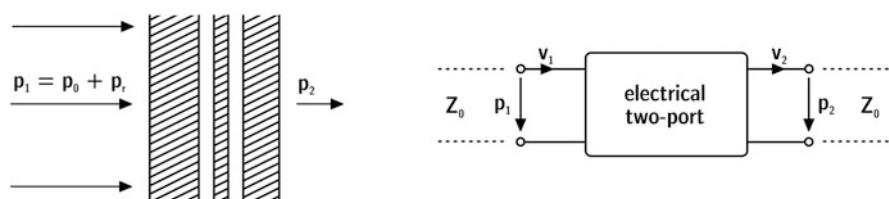


Fig. 13.2 Two-port model of partitions

13.4.1 Single Layers

With only an effect of inertia, the sound insulation of a single layer can be calculated easily (Fig. 13.3). The impedance Z_p is given by $j\omega m''$, with m'' denoting the mass per unit area (see also 3.2.1). These results are, at least in case of buildings, a good estimate of the trend of insulation.

$$a_{11} = a_{22} = 1, a_{12} = -Z_p; \quad a_{21} = 0 \quad (13.9)$$

and according to Eq. (13.6), we get

$$R = 10 \log \left| 1 + \frac{Z_p}{2\rho_0 c} \right|^2 \quad (13.10)$$

and

$$R = 10 \log \left[1 + \left(\frac{\omega m'' \cos \theta}{2\rho_0 c} \right)^2 \right] \approx 20 \log \left(\frac{\omega m'' \cos \theta}{2\rho_0 c} \right), \quad (13.11)$$

the latter approximation for $\omega m'' \cos \theta \gg 2\rho_0 c$.

Stiffness effects like bending waves (Sect. 5.3), size effects (Sect. 5.3.1) or losses (Sect. 5.3.2) are accounted for by adding more detailed terms into Z_p .

$$Z_p = j \left(\omega m'' - \frac{Bd\omega^3 \sin^4 \theta}{c^4} \right) \quad (13.12)$$

In case of an incident diffuse field, the effective sound reduction index is about 5 dB lower than the result of Eq. (13.11) (Fig. 13.4). The resulting sound reduction index is

$$R = 20 \log \left(\frac{\omega m''}{2\rho_0 c} \right) - 10 \log \left(\frac{1}{2\eta_{\text{tot}}} \sqrt{\frac{f_c}{f}} \right) \quad (13.13)$$

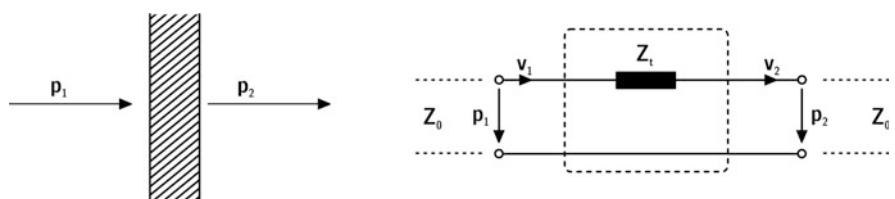
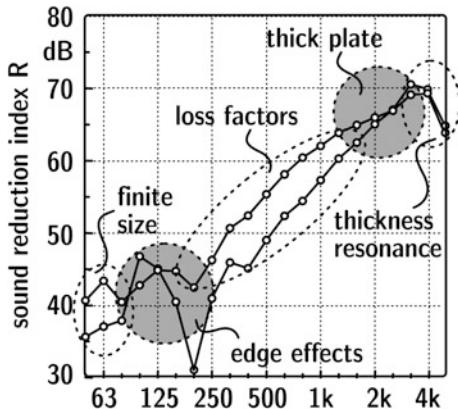


Fig. 13.3 Sound transmission through a partition with mass impedance

Fig. 13.4 Typical results of measured sound insulation of massive walls. Range of variation found in intercomparison tests. (After Meier 2000)



13.4.2 Double Layers

The principle of double wall is well known for lightweight constructions and for glazing (Fig. 13.5). Double or multiple wall systems are not only used in building acoustics. Their installation is used for other applications as well since it is an efficient way to improve the sound insulation at high frequencies. We come back to the two-port impedance model and extend the separation impedance towards two mechanically decoupled masses. The inner part represents an air gap or a gap filled with viscoelastic material. The air gap as well as the material filling acts as spring (Sect. 3.2.1, Eq. 3.13).

For the frequency range above resonance ($\omega \gg \omega_0$), the movement in the system (vibration velocity) is concentrated in the first mass layer and the stiffness element. The stiffness element produces a short-circuit, so that the second mass stays in rest. The sound reduction index can thus be approximated by (see also Fig. 13.6)

$$R = 10 \log \left(1 + \left(\frac{\omega(m''_1 + m''_2) \cos \vartheta}{2\rho_0 c} \left(1 - \frac{f^2}{f_r^2} \right) \right)^2 \right). \quad (13.14)$$

At frequencies below the resonance frequency f_r , the spring shows an infinitely large impedance, the energy bypassing the masses is almost zero and, thus, the behaviour of the double wall is identical with that of a single wall of mass $m''_1 + m''_2$. At resonance frequency, the separation impedance $Z_p = 0$ and accordingly $R = 0$. In practical cases, viscous losses in the air gap and friction in the mass layers limit the minimum of Z_p to a real value.

The resonance frequency is given by

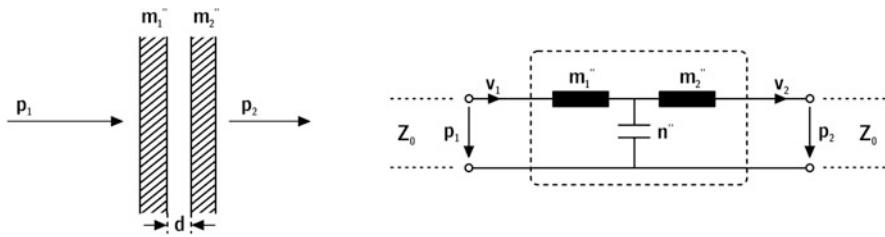
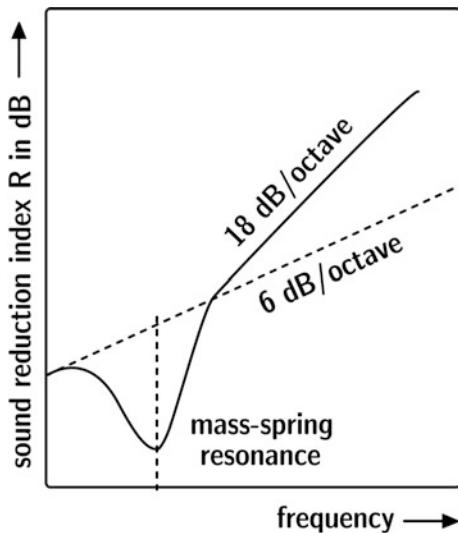


Fig. 13.5 Sound transmission through double wall

Fig. 13.6 Sound reduction index of a double wall



$$f_r = \frac{1}{2\pi} \sqrt{\frac{1}{M'' n''}} \quad (13.15)$$

with the reduced mass M''

$$M'' = \frac{m_1'' m_2''}{m_1'' + m_2''}. \quad (13.16)$$

13.5 Sound Insulation of Buildings

The most important difference between academic studies of sound insulation and the effective sound insulation in a real building is the presence of energy transmission via several paths. Particularly in the case of very high sound insulation of the

separating wall, sound energy will flow significantly through other building components, such as small elements, doors, frames, flanking walls, suspended ceilings, access floors or façades. Like in a strong flood of water, any dam will be only efficient if the bypass water flow will be prohibited. The characterization of the sound energy flow is, therefore, not sufficiently covered in case of incident and transmitted sound waves for adjacent rooms are separated by the partition. The tangential energy flow in the structure or plates and beams involving all structural wave types, their interaction and transformation at junctions, and the structure-to-air radiation must be considered as well. Simulation and auralization of this kind of sound and vibration generation and transmission is a complex problem of transfer path identification and superposition.

The results are usually measured or predicted in one-third octave bands between 100 Hz and 3150 Hz or in the extended range between 50 Hz and 5000 Hz. It is clear that the frequency range below 100 Hz may play a significant role for the effective sound insulation. This might become even more important since low frequency sound sources are increasingly relevant in residential buildings (TV and Hifi equipment).

By definition, the resulting sound reduction is denoted by R' , the so-called apparent sound reduction index,

$$R' = -10 \log \tau' = L_S - L_R + 10 \log \frac{S}{A_R}, \quad (13.17)$$

with τ' denoting the apparent transmission coefficient including all transmission paths. The reference to account for the sound transmission is still the partition wall with its surface area S , although the energy might be transmitted via flanking paths.

More elegant and more useful for auralization is the definition of sound level differences with reference to reverberation times in the receiving room, T_R , and standardization to $T_0 = 0.5$ s.

$$D_{nT} = L_S - L_R + 10 \log \frac{T_R}{T_0}, \quad (13.18)$$

which is called standardized sound level difference. The difference between R' and D_{nT} is a constant geometric term ($\approx V_R/3 S$) introducing the receiving room volume, V_R ,

$$D_{nT} = R' + 10 \log \frac{V_R}{3 S}. \quad (13.19)$$

Note that this equation allows the prediction of receiving room sound from source room sound and data from the transmission in the building structure. To achieve this form, we rewrite Eq. (13.19) into

$$L_R = L_S - D_{nT} + 10 \log T_R + 3, \quad (13.20)$$

with all L and D in decibels and T in s. It will be used in Sect. 13.7.

13.5.1 Flanking Transmission

Similarly to direct sound transmission the energy flow via flanking paths can be studied. The conditions for characterization of flanking paths are defined in measurement standards. The direct path must be blocked and only one tangential energy path is active. Test facilities for flanking sound transmission are used to determine transmission coefficients in case of the specific energy flow from the building element in the source room, i , to the element in the receiving room, j . Its definition is thus equivalent to that of the direct transmission,

$$R_{ij} = -10 \log \tau_{ij} = 10 \log \frac{P_i}{P_j}, \quad (13.21)$$

with P_i and P_j denoting the sound power incident on the element i in the source room and radiated by the element j in the receiving room, respectively.

The definitions and measurement procedures are specified for sound insulation test facilities. The energy flow in parts of the junctions is blocked by gaps, usually filled with resilient material. Hence the remaining energy transmission path can be controlled and studied with regard to the building constructions tested (Fig. 13.7).

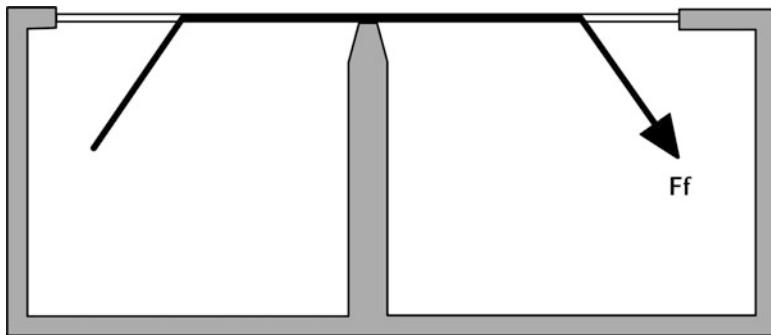


Fig. 13.7 Test facility for determination of flanking transmission

13.6 Sound Transmission Prediction Models

The specific transmission coefficients obtained by separated measurement conditions or calculations have to be combined into a resulting sound transmission coefficient to combine all relevant transmission paths. This calculation consists of addition of energy transmission with weighing factors proportional to the area of sound irradiation and radiation in the receiving room. It can be considered as energy balance approach between subsystems. In a consequent and theoretically sound approach, the statistical energy analysis, SEA, is a method of interest. It requires a high modal density and is thus applicable in many building situations with low critical frequency. In these cases, SEA can be applied in a wide frequency range (see also Sect. 10.2.2). In SEA the energy flows between subsystems are calculated by using the energy losses in the systems and the corresponding coupling losses (coupling loss factors). “Systems” in the sense of the method are the statistical modal fields in the room cavities and on the plates (walls, ceilings).

The fundamental equations of the transmission model appropriate for sound insulation in buildings were developed by (Gerretsen 1979). Although not explicitly referred to as SEA approach, the equation system is equivalent to SEA, with the energy flow limited to cover transmission via the direct path and paths via one junction (Fig. 13.8).

The total portion of the sound power transmitted is (see also Eq. 13.17)

$$\tau' = \sum_{i=1}^N \tau_i \quad (13.22)$$

with

$$\tau' = \tau_d + \sum_{f=1}^n \tau_f , \quad R' = -10 \log \tau' \quad (13.23)$$

and

$$\tau_d = \tau_{Dd} + \sum_{F=1}^n \tau_{Fd} , \quad \tau_f = \tau_{Df} + \tau_{Ff} , \quad \tau_{Dd} = 10^{-R_{Dd}/10} \quad (13.24)$$

In general, all specific transmission coefficients are related to their corresponding flanking sound reduction indices, R_{ij} .

$$\tau_{ij} = 10^{-R_{ij}/10} \quad (13.25)$$

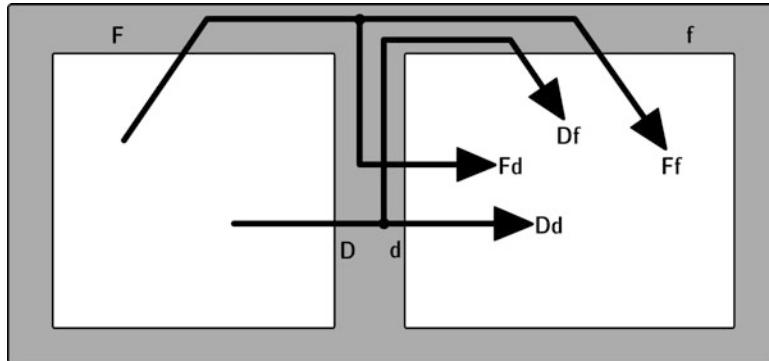


Fig. 13.8 Notation of flanking paths in sound transmission between rooms

Accordingly they can be measured in test facilities. But the power reduction can also be calculated by summing the insulating effects of irradiation in the source room:

$$R_{ij} = -10 \log \tau_{ij} = \frac{R_i}{2} + \Delta R_i + \frac{R_j}{2} + \Delta R_j + K_{ij} + 10 \log \frac{S_S}{l_0 l_{ij}}. \quad (13.26)$$

In this formulation, reciprocity is used, and, thus, the vibration level difference between the connected plates, i and j , are accounted for in a symmetric way and the result is combined. To derive this estimation of R_{ij} , the power flow balance is calculated between two adjacent rooms coupled by plates in L or T junction (Gerretsen 1979). The power transmitted over the junction is one of the crucial parts of the model. Thus, the vibration reduction index, K_{ij} , is specifically important. It describes the energy transmitted via functions of building elements. Its definition complies exactly with the energy balance equation used for airborne sound transmission (Eq. 13.17),

$$K_{ij} = \frac{D_{v,ij} + D_{v,ji}}{2} + 10 \log \frac{l_{ij}}{\sqrt{a_i a_j}}. \quad (13.27)$$

The latter term represents the total losses in the plates (see also 5.3.2). a is the equivalent absorption length of the plates' perimeters and l_{ij} the length of the common junction. As in decaying sound fields in rooms, the equivalent absorption length can be related to the decay time referred to as structural reverberation time. The relationship between the equivalent absorption length, a , and decay of vibration energy, T_s , in a plate of area S is

$$a = \frac{2.2 \pi^2 S}{c T_s} \sqrt{\frac{1 \text{ kHz}}{f}}. \quad (13.28)$$

Finally, the structural reverberation and the loss factor can be divided into three parts, according to the basic information in Sect. 5.3.2. For the application discussed here, sound insulation in building structures, all loss effects are relevant, depending on the construction material. It is well known in building construction that losses may play a significant role in enhancing sound insulation, and this can be achieved either by internal material losses or by energy flow over junctions away from the receiving room (equivalent to electric current being grounded).

The final adjustment of absorption has to be carried out with respect to the actual energy flow in the field situation. This procedure is necessary in case of laboratory data of a and T_s under certain boundary conditions, while the mounting method or connection between building elements in the real building might be different.

13.7 Auralization of Airborne Sound Insulation

Before discussing the procedure for creating an auralization filter, we divide the problem into the possible variants of source signal recording (Fig. 13.9). The interface between source signal and filter has to be defined accordingly. The source can either be recorded in the source room, thus containing the source room's reverberation, or it can be recorded in free field. In this case, the directivity in notation of polar pattern or as coded data can also be taken into account.

13.7.1 Source Signal Recording in the Source Room

Starting with the input data of sound insulation in frequency bands which are obtained by prediction or measurement, we can create auralization filters. A prerequisite, however, is that the transmission paths are clearly separated and that the significant contributions are captured in the model. Based on the calculation of the effective sound insulation including all relevant transmission paths (Eq. 13.22), the sound level reduction associated with each path is obtained. Since the energy model

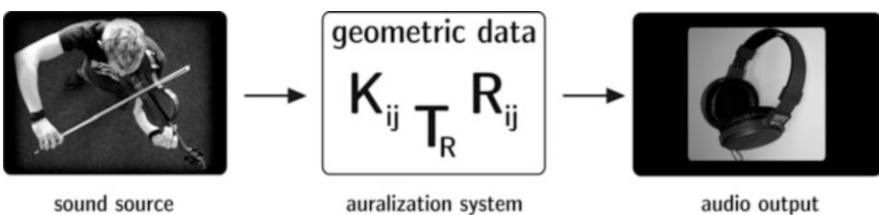


Fig. 13.9 From input data to sound: Auralization of buildings

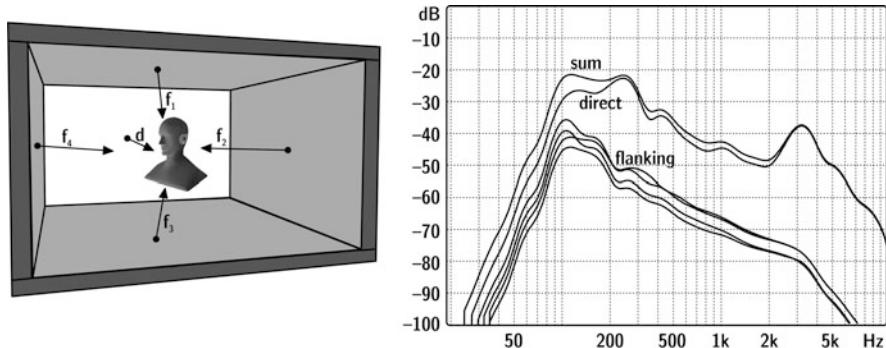


Fig. 13.10 Geometric situation and interpolated filters of the sound transmission paths involved

is making use of the fact that we have statistical modes and energy summation, it is easily transferred into software. The resulting standardized sound level difference, D_{nT} , is

$$D_{nT} = L_S - L_R + 10 \log \frac{T}{0.5 s} = -10 \log \tau' + 10 \log \frac{V}{3 S} = -10 \log \tau_{nT} \quad (13.29)$$

with V denoting the receiving room volume in m^3 and S the partitions walls surface in m^2 .

The key of getting from simulation to auralization is, again, mapping the problem onto a problem of signal processing. By introducing the sound pressure signals, p_S and p_R , in the source room and the receiving room, respectively, we rearrange Eq. (13.29) to

$$p_R^2 = p_S^2 \frac{\tau_{nT} T}{0.5 s} \quad (13.30)$$

and

$$p_R(\omega) = p_S(\omega) \cdot F_{\text{total}}(\omega) = p_S(\omega) \sum_{i=1}^N F_{\tau,i}(\omega) e^{-j\omega\Delta\tau_i} F_{\text{rev},i}(\omega), \quad (13.31)$$

with $F_{\tau,i}$ denoting interpolated filters which are obtained from the energy transfer spectra of the paths involved. $\Delta\tau_i$ are delays corresponding to the geometric situation of the walls and the observation point (Fig. 13.10). $F_{\text{rev},i}$ is the reverberation excited by each of the sound transmitting elements in the receiving room. The absolute sound pressure level in the receiving room is correct if the input sound pressure

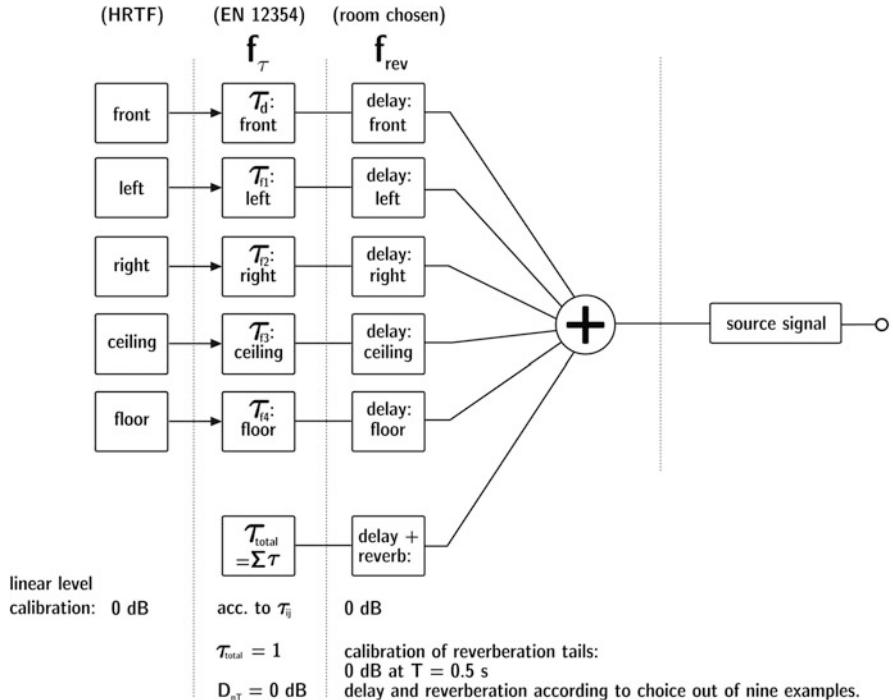


Fig. 13.11 Flow chart of auralization of airborne sound insulation

signal in the source room is calibrated with reference to $2 \cdot 10^{-5}$ Pa. Note that $p_S(\omega)$ is nothing but the complex spectrum of the recorded source time signal. Eq. (13.31) can hence be expressed in the time domain by

$$p_R(t) = p_S(t) * f_{\text{total}}(t), \quad (13.32)$$

with $f_{\text{total}}(t)$ denoting the transmission impulse response, see Figs. 13.11 and 13.12.

Except for the phases, the total set of transfer functions is represented quite accurately, as long as the frequency interpolation does not smooth the exact physical behaviour too much.³ This situation is absolutely acceptable since phases in reverberant sound fields cannot be recognized by human hearing. This does not apply, however, to the discrimination of direct sound and the first (early) reflections related to the direction of sound incidence and the spatial aspects of the early part of the impulse response. Those phases are well covered by $\Delta\tau_i$.

³If the interpolation was too rough, statistical energy analysis would not be an appropriate model. The model has its limits at low frequencies (at low modal densities).

Fig. 13.12 Impulse response of the transmission source room – receiving room

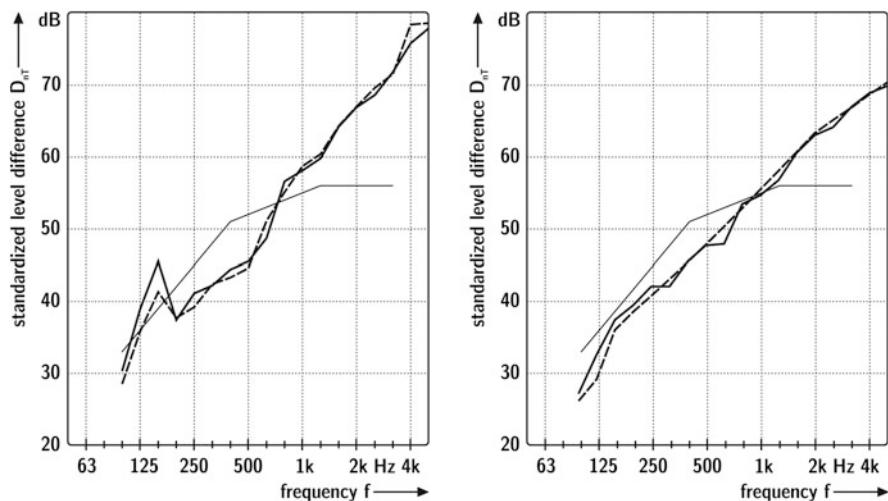
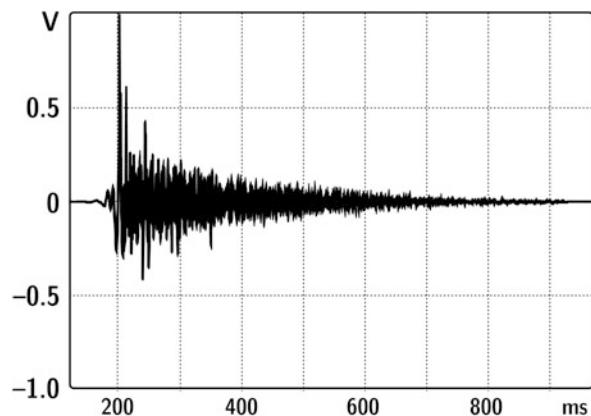


Fig. 13.13 Comparison of the input data of D_{nt} and D_{nt} measured from the auralized signals. The auralization stimuli were created by using pink noise and a measurement by feeding the signals to headphones on an artificial ear

The deviation of the auralized level differences from the measured level differences is mostly below 1 dB. Only at low frequencies a deviation of more than 2 dB occurs. These uncertainties are acceptable since they occur in measurements at these low frequencies anyway due to modal effects. For the auralization “measurement”, only one point in the receiving room was used. With appropriate spatial averaging, the accordance would be even better (Fig. 13.13).

13.7.2 Source Signal Recording in Free Field

For the examples described in the section above, it is assumed that reverberation affects the result primarily in the receiving room. This approach is inherently consistent since the auralization filter of that kind was related to the source room level, thus including the source room reverberation. The source room, however, may also contribute to the listening experience by its reverberation, as pointed out by (Rindel 2006). An extension in this respect is implemented into the model by using the sound power, P , of the source as a reference instead of the source room level.

With reference to the sound power of the source, L_w , the fundamental equation of sound insulation auralization reads (see also Eq. 4.39)

$$D_{nT} = L_w - L_R + 10 \log T_S + 10 \log T_R - 10 \log V_S + 17 \quad (13.33)$$

with $D_{nT} = -10 \log \tau_{nT}$ denoting the standardized sound level difference, L_w , the sound source power, L_R , the receiving room sound pressure level, T_S and T_R the reverberation times of the source room and the receiving room, respectively, and V_S the volume of the source room.

In energetic notation and rearrangement to obtain the sound pressure in the receiving room this reads⁷⁰

$$p_R^2 = \left(2 \cdot 10^4 \frac{\text{Pa}}{\text{s}} \right) \cdot P \cdot \frac{\tau_{nT} T_S T_R}{V_S}. \quad (13.34)$$

As expected, the sound power, P , of the source determines the squared sound pressure level in the receiving room. The factor $2 \cdot 10^4 \text{ Pa/s}$ stems from the combination of various constants like the reference level for sound pressure ($p_0 = 2 \cdot 10^{-5} \text{ Pa}$), the reference sound power level ($P_0 = 10^{-12} \text{ W}$) and from the reference reverberation time ($T_0 = 0.5 \text{ s}$) in the definition of D_{nT} .

Eq. (13.34), however, is not directly applicable when it comes to creating a source-filter model, since the source pressure is somehow hidden in the sound power. For a source characterization, however, the far field pressure in well-defined direction and distance and the directivity pattern can be used (see Sect. 8.2). For omnidirectional sources and with reference to signals recorded in free field in 1 m distance yields⁴

⁴All quantities to be expressed by using their numerical values in SI units.

$$p_R^2(\omega) = \left(2 \cdot 10^4 \frac{\text{Pa}}{\text{s}}\right) \cdot \frac{4\pi}{\rho_0 c} \cdot p_S^2(\omega)|_{1m} \cdot \frac{\tau_{nT} T_S T_R}{V_S} \quad (13.35)$$

which can further be expanded into

$$\begin{aligned} p_R(\omega) &= p_S(\omega)|_{1m} \cdot 24.6 \cdot \sqrt{\frac{\tau_{nT} T_S T_R}{V_S}} \\ &= p_S(\omega)|_{1m} \cdot F'_{\text{total}}(\omega) \\ &= p_S(\omega)|_{1m} \cdot \frac{24.6}{\sqrt{V_S}} \cdot F_{\text{rev},S}(\omega) \cdot \sum_{i=1}^N F_{\tau,i}(\omega) e^{-j\omega\Delta t_i} \cdot F_{\text{rev},i,R}(\omega) \end{aligned} \quad (13.36)$$

The filters to be created in this situation must be related to a free-field calibrated sound source, $p_S(\omega)|_{1m}$,⁵ to a reverberation filter, obtained by simulation or measurement in the source room, $F_{\text{rev},S}$, reverberation determination by simulation or measurement in the receiving room, $F_{\text{rev},i,R}$, and to the standardized sound insulation quantity D_{nT} , as described above. The reverberation time filters have to be calibrated with regard to a total energy of 1 in case of $T = 1$ s, respectively. As above, the same can be expressed in the time domain by

$$p_R(t) = p_S(t)|_{1m} * f'_{\text{total}}(t), \quad (13.37)$$

with $f'_{\text{total}}(t)$ denoting the transmission impulse response.

In contrast to the example of source recording in the source room, a directional source has to be treated in a specific way. The direction used in the recording situation has to match the directional reference in the determination of the source room impulse response or the corresponding transfer function, $F_{\text{rev},S}(\omega)$. This situation can be understood by considering the differences of the intensities radiated into the various directions into the source room. This process feeds the source room reverberation with a total power of $P |\Gamma|^2$, which can also be introduced in Eq. (13.34).

The methods described in this chapter are explained by the example of airborne sound transmission in buildings. Nevertheless they are applicable in equivalent problems of transmission paths between two cavities or enclosures. Such applications might be airborne sound transmission in vehicles, in ships or similar cases in noise control engineering; see Chap. 15.

⁵Note the different reference for the recorded signal compared with Eq. (13.31).

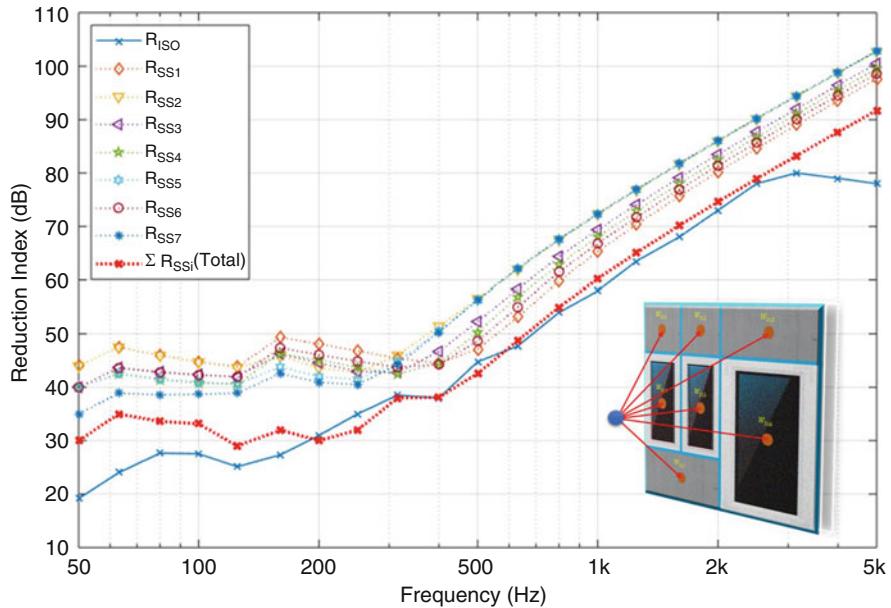


Fig. 13.14 Angle-dependent element sound reduction indices for an outdoor source in front of a façade. Total sound reduction index, R_{SSi} (■), and standard sound reduction index, R_{ISO} (x)

13.7.3 Simulation with Distributed Secondary Sources

In more complex scenarios, the walls radiating the transmitted sound can be modelled with distinct amplitude and phase information which stem from angular characteristic of sound incidence. In cases of façade sound insulation, for example, vehicle pass-by creates a time-dependent angle of incidence and distance to façade and flanking elements. On the source side, the incident sound is decomposed into a more complex sound field on the walls consisting of a direct and, optionally, a diffuse field component in case if an indoor source room. The sound energy transmitted via direct and flanking paths to the adjacent receiving room varies for all surface elements, called “patches”, depending on the sound pressure hitting the corresponding building elements on the source side, due to the source position and directivity.

The influence of the reverberation on the source and the receiver side, rooms and the balance between direct and reverberant energies in the receiving room are incorporated into sound insulation transfer functions as well. These transfer functions are developed for extended radiating walls treated as many point sources (secondary sources on the patches, see Fig. 13.14) rather than considering these walls as single point sources on the receiving room walls (Imran et al. 2019a). It is

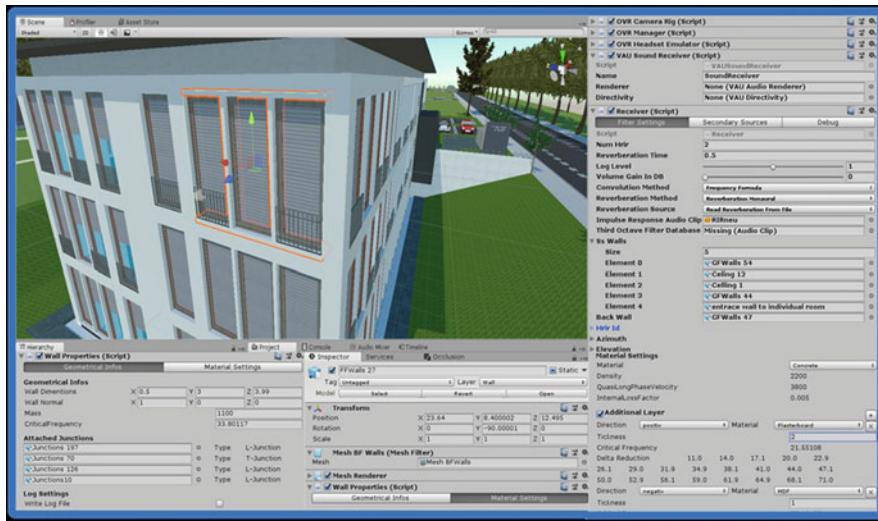


Fig. 13.15 User interface for real-time sound insulation simulation: assigning material parameters to building elements and calculations for geometric input data. (After Heimes et al. 2019)

also adopted a procedure to synthesize an estimated exponential reverberation tail from the reverberation time to include the effects of absorption of room boundaries as well as to simulate a plausible real room (Fig. 13.15).

Chapter 14

Simulation and Auralization of Structure-Borne Sound



14.1 Introduction

The straightforward general extension of the propagation and auralization model is discussed now with regard to the excitation of solid structures by forces. The source, therefore, has to be characterized by its force output or by its velocity injected into the medium. With the velocity in the structure being known, transmission and radiation of sound to a receiver can be considered a solved problem. The first example to illustrate this kind of model extension is impact sound in buildings. The basic definitions and methods for prediction and evaluation of impact sound are given in this chapter.

So far, the prediction and auralization focused on airborne sound propagation in free field (Sect. 9.4), in a room (Sect. 11.9), through a system of partition walls (Sect. 13.7). These applications have in common that the filters generated are intended for an interface with a recorded “dry” airborne source signal. With proper adjustment of the sound power, directivity or the volume velocity of a source membrane, this approach is unambiguous at first sight. If one takes a closer look at the source mechanism, it might be necessary to consider the acoustic load on the source. This, however, hardly applies to the applications listed above.

14.2 Definitions of Impact Sound Transmission

The characterization of sound insulation against the force impact of persons walking on floors into the room below is based on representation of the primary source by a technical source. This so-called tapping machine excites the floor with a standardized momentum. The tapping machine should represent the force injected into a floor by a walking person. It contains a mechanical construction including five hammers with weight of 500 g each. They are lifted by 4 cm and released to a freely falling

condition. Their hammer speed at impact, hence, is fixed to 0.886 m/s. The total repetition rate is 10 Hz. In case of an elastic impact, the impulse force will be transmitted completely into the contact surface, provided the input impedance is very high compared with the impedance of the hammer. This kind of force source represents an ideal case. The contact condition is:

- Point contact
- Large contact impedance of the transmitting system

These conditions are not valid for lightweight or resilient transmitting systems like wooden floors or soft floor coverings. Nevertheless we continue to discuss the impact based on this assumption, before some modifications will be introduced in Sect 14.5 which extend the validity towards smaller floor impedances. Before the problem of structure-borne excitation is discussed in a more general way, we use the tapping machine as a reference. The tapping machine is the result of a standardization process over many years, and it is widely used today in international measurement standards and rating schemes.

Due to direct excitation with standardized force, the source components do not have to be determined. The result of impact sound insulation is, thus, given by the receiving room level, normalized to a reference condition of reverberation:

$$L_n = L_R + 10 \log \frac{A}{A_0}, \quad A_0 = 10 \text{ m}^2, \quad (14.1)$$

where L_n denotes the normalized impact sound level and A the equivalent absorption area of the receiving room. Alternatively, the standardized impact sound level, L_{nT} , is in use.

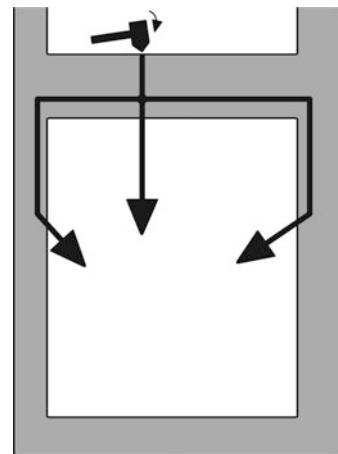
$$L_{nT} = L_R - 10 \log \frac{T}{T_0}, \quad T_0 = 0.5 \text{ s}, \quad (14.2)$$

The above-mentioned problems of flanking transmission do not occur to the same extent as in airborne sound insulation, because mainly one building component is excited by the impact source. Furthermore, significant parallels between D_n and L_n on the one hand and D_{nT} and L_{nT} on the other hand can be observed. A quantity denoting a transmission loss in terms of an intensity ratio, however, does not exist here.

14.3 Impact Sound Model

Compared to airborne sound insulation the insulation, modelling and auralization of impact sound generated by walking on a floor is more difficult. The methodology listed here is therefore based on preliminary results of ongoing research. At first, it must be noted that all data on the impact noise levels of floors are defined on the

Fig. 14.1 Sound transfer path from force source to receiving room



basis of the tapping machine. This prerequisite is not quite suitable for lightweight floor constructions or coverings like wooden floating floors (for more information on how to deal with these problems, see 14.5). Furthermore, nonlinearities have to be taken into account in case of timber floor constructions excited by low-frequency heavy impacts.

As in the case of airborne sound insulation, the estimation presented in this model is also based on the approach of SEA. The modal (resonant) response is the basis of statistical energy flow between the subsystems of floors, walls and rooms. The frequency range is thus limited to the range above critical frequencies, where the energy balance in various building and laboratory situations provides a good estimate of the impact sound level. Specific singular modal effects, forced transmission and nonlinearities are neglected.

The total vibration and radiation into the receiving room is obtained by adding the energy for all transmission paths, (Fig. 14.1)

$$L'_n = 10 \log \left(10^{L_{n,d}/10} + \sum_{j=1}^N 10^{L_{n,j}/10} \right). \quad (14.3)$$

As in airborne sound transmission, the model parameters L_n are to be corrected with regard to the losses in the actual field situation (represented by the structural reverberation times T_s).

$$L_{n,situ} = L_{n,lab} + 10 \log \left(\frac{T_{s,situ}}{T_{s,lab}} \right), \quad (14.4)$$

with $L_{n,lab}$ denoting the reference laboratory condition.

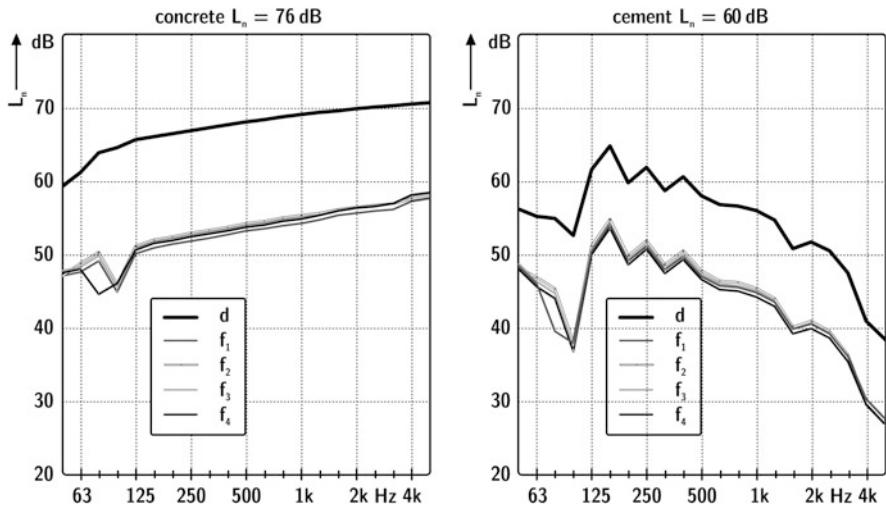


Fig. 14.2 Simulated direct and flanking impact sound level of a concrete floor without (left) and with (right) floating floor. (After Thaden 2005)

Additional layers such as suspended ceilings or wall linings are accounted for by subtracting the airborne or impact sound improvement, ΔR or ΔL_n , from the impact sound level.

$$L_{n,d} = L_{n,situ} - \Delta L_{situ} - \Delta L_{d,situ}. \quad (14.5)$$

For the flanking paths holds

$$L_{n,ij} = L_{n,situ} - \Delta L_{situ} + \frac{R_{i,situ} - R_{j,situ}}{2} - \Delta R_{j,situ} - \overline{D_{v,ij,situ}} - 10 \log \sqrt{\frac{S_i}{S_j}}. \quad (14.6)$$

Equation (14.6) also includes influences of junction level differences in the actual field situation. It has to be identified in each specific case which component is most relevant. Furthermore, the impact sound reduction of linings, floating floors and suspended ceilings cannot be transferred easily from one bare floor system to the next, particularly when comparing massive and timber bare floor constructions. The example presented in Fig. 14.2 illustrates the result of the impact sound model for a robust field situation of a massive bare floor with and without cement floating floor.

In case of unknown floor construction or in case that only a rough estimate of the floor construction is required, the normalized impact sound pressure level of a monolithic massive floor can be calculated from the mass per unit area, m'' ; the losses expressed by the structural reverberation time, T_s ; the radiation efficiency, σ ; and the frequency f ,

$$L_n \approx 155 - 30 \log m'' + 10 \log T_s + 10 \log \sigma + 10 \log \frac{f}{1 \text{ kHz}}. \quad (14.7)$$

14.4 Impact Sound Auralization

The interaction of the impact source and the floor construction will be discussed in Sect. 14.5. For now, we consider the impact source as being ideal. With the impact sound level of the floor construction known, the procedure of creating a filter for auralization is quite similar to that described above (airborne sound). Results in frequency bands are interpolated and treated as spectral filters (with phase 0). From the definition of L_n , for instance, we obtain

$$p_R^2 = \sum_i 10^{L_{n,i}/10} p_0^2 \frac{T}{0.5 \text{ s}} \frac{V}{30}, \quad (14.8)$$

equivalent to the procedure discussed in 13.7.

If one attempts to model and auralize the noise of a person walking on the floor above on the basis of standardized impact sound levels, the tapping machine excitation has to be extracted from the data. This can be achieved by dividing the impact sound spectra by the force excitation spectrum of the standard tapping machine. Thus, a transfer function, $H_{Fp}(\omega)$, can be defined by assuming the injected force to be invariant on various floor constructions. In terms of signal spectra and generalized to other force sources, it reads

$$p_R(\omega) = F_{\text{walker}}(\omega) \cdot \frac{p_{\text{TM}}(\omega)}{F_{\text{TM}}(\omega)} \cdot \sum_{i=1}^N f_{\tau,i}(\omega) f_{\text{rev},i}(\omega) = F_{\text{walker}}(\omega) \cdot H_{Fp}(\omega), \quad (14.9)$$

with F_{walker} denoting the spectrum of the force-time signal of the actual excitation, p_{TM} deduced from the normalized spectrum (L_n) of the tapping machine excitation and F_{TM} the force spectrum of the tapping machine. $f_{\tau,i}$ and $f_{\text{rev},i}$ were defined above in 13.7.

The two floor constructions were, for instance, auralized and analysed with regard their sound pressure levels by (Thaden 2005). In this case the impulse response for the transmission between the force signal in the source room and the sound pressure signal in the receiving room was calculated from the impact sound levels as shown in Fig. 14.2. The forces of the tapping machine and a rubber ball according to (Tachibana et al. 1998) were measured (Fig. 14.3).

Fig. 14.3 Force spectra of the tapping machine and rubber ball

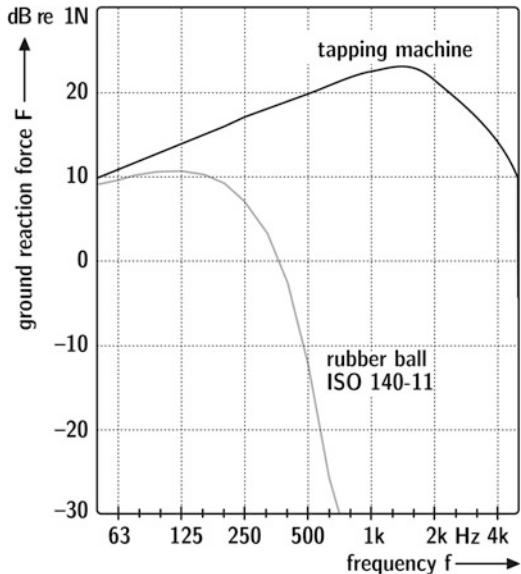
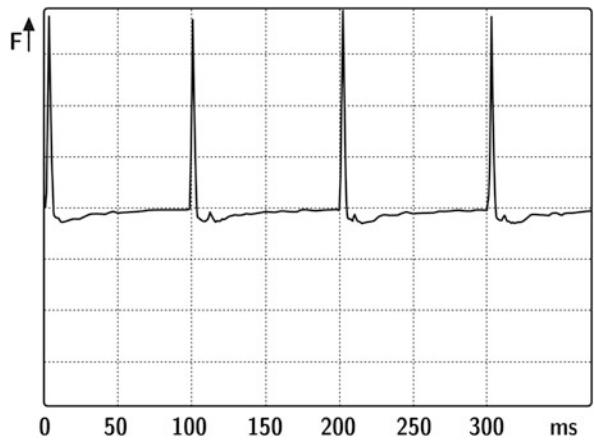


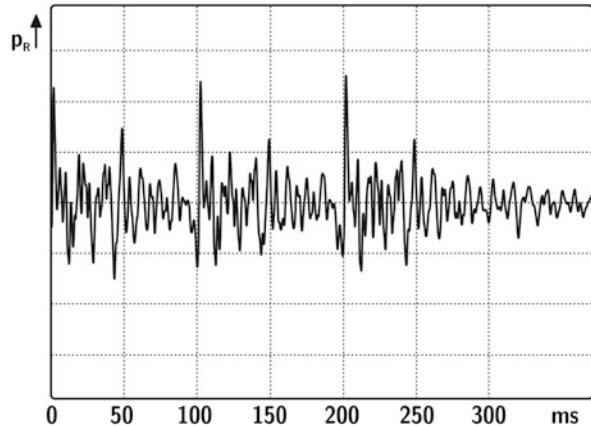
Fig. 14.4 Force-time signal of the tapping machine



Force time signals of the tapping machine are constructed as follows. To obtain suitable time signals, hammer force pulses are recorded and added at 10 Hz rate. Jitter in time and amplitude is introduced to get a more natural impression. A convolution of this primary force signal with the impulse response yields the sound pressure signal (Fig. 14.4).

Other force sources can be treated in the same way, based on recording of the force-time function (Fig. 14.5).

Fig. 14.5 Example of time signals of the sound pressure in the receiving room, tapping machine excitation of a bare concrete floor



14.5 Structure-Borne Interaction Model

The direct method described in the section above, however, is only an approximation since the injected force and the resulting velocity in the (upper layer of the) floor construction depends on the floor mobility. It is difficult to solve this problem in practice, even in case of a linear transmission (Fig. 14.6).

The open-circuit (free) or blocked force and the inner impedances of the source and the floor have to be determined. Impedances and forces can be obtained according to the method described in Sect. 14.4. Based on this model of vibration sources, the interaction of the source with the attached floor system can be calculated. The transfer function between the source and the receiving point (receiving room) is again derived from the normalized or standardized impact sound level (according to 14.2).

$$\begin{aligned}
 p_R(\omega) &= F_{0,\text{walker}}(\omega) \cdot \frac{p_{\text{TM}}(\omega)}{F_{0,\text{TM}}(\omega)} \cdot \frac{Z_{i,\text{TM}}(\omega) + Z_{a,\text{floor}}(\omega)}{Z_{i,\text{walker}}(\omega) + Z_{a,\text{floor}}(\omega)} \\
 &\quad \cdot \sum_{i=1}^N f_{\tau,i}(\omega) f_{\text{rev},i}(\omega) \\
 &= F_{0,\text{walker}}(\omega) \cdot H_{\text{Filter}}(\omega),
 \end{aligned} \tag{14.10}$$

with $F_{0,\text{walker}}$ denoting the blocked-force spectrum of the footstep source, rubber ball of tapping machine, p_{TM} the sound pressure spectrum of the tapping machine (derived from L_n) and $F_{0,\text{TM}}$ the blocked-force spectrum of the tapping machine. Z_i and Z_a are the impedances of sources and construction, respectively.

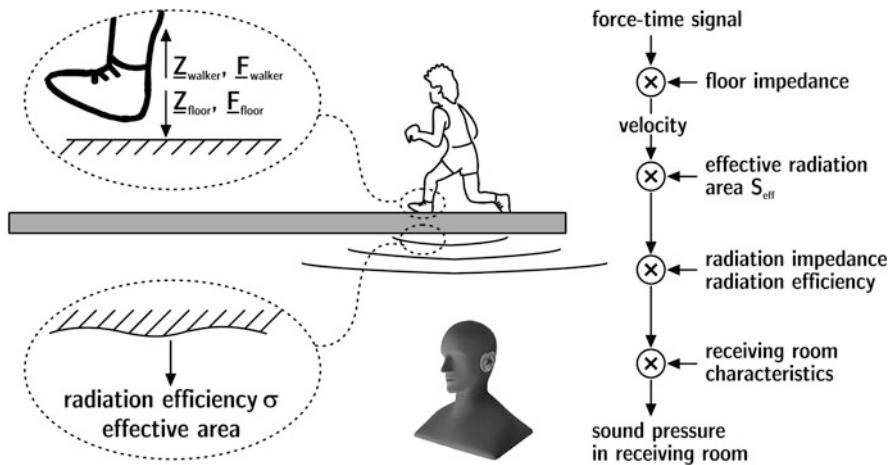


Fig. 14.6 Auralization model for impact noise

Impedances of sources and of floor systems are currently subject to research. At the time being, no standardized definition or measurement method is available. Preliminary results were presented by (Scholl 2001; Brunsøg and Hammer 2003; Lievens and Brunsøg 2007), among others. The curves show some agreement but they differ in details. Examples are shown in Fig. 14.7.

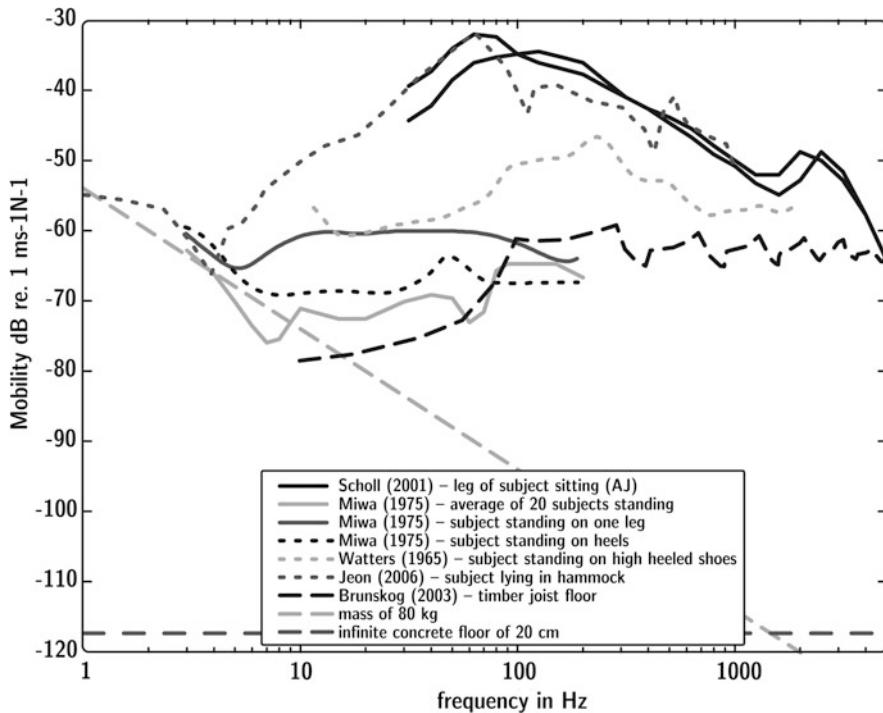


Fig. 14.7 Impedances of feet and floor constructions. (After Watters 1965; Miwa 1975; Scholl 2001; Brunskog and Hammer 2003; Jeon et al. 2006)

In case of footsteps, the model of point contact, however, is just the first step. It must be analyzed to which extend it is important to consider the actual contact area and the temporal dependence of contact forces on the area. And after all, the problem of a nonlinear contact due to changes of stiffness caused by the weight balancing on the heel and toe while walking are factors that result in uncertainties. Further research will have to determine whether these effects are perceptively noticeable in auralized signals.

Chapter 15

Transfer Path Analysis and Synthesis



15.1 Introduction

In this chapter, the acoustic model of airborne and impact sound transmission and the subsequent auralization technique are generalized towards other applications in acoustics and noise control engineering, particularly to vehicle acoustics. It was shown that the resulting sound pressure signal at the listener's ears can be constructed by using binaural filters. Sound transmitting elements are accounted for by their transfer function in order to achieve the correct level and colouration. Phase aspects can often be neglected in case of diffuse field conditions in the listener's environment. Components of early, primary sound, however, must be modelled with their phase or group delay which enhances the presence and immersion and the correct localization. From these components, sources and transfer paths can be constructed and combined into an efficient auralization model.

The Binaural Transfer Path Synthesis (BTPS) is closely related to the binaural transfer path analysis.¹ The analysis is an indispensable tool of acoustic engineering. Applications are manifold. It can, for instance, be used to identify and characterize sources and weak elements of sound transmission and relevant components of sound radiation. When the acoustic system of generation and transmission is identified and separated into the main sound transmitting paths, the synthesis process can be started. It is not only possible to reproduce the analysed situation but also to alter the entire system and to exchange its components. Sound insulation in buildings is one example. The paths of direct and flanking transmission were treated exactly as described above. Adding linings or elastic interlayers in one of the paths is a typical variation of the system. This system can easily be studied with regard to the effective sound insulation performance, and it can be auralized to compare it to other systems and their variations.

¹So-called BTPA.

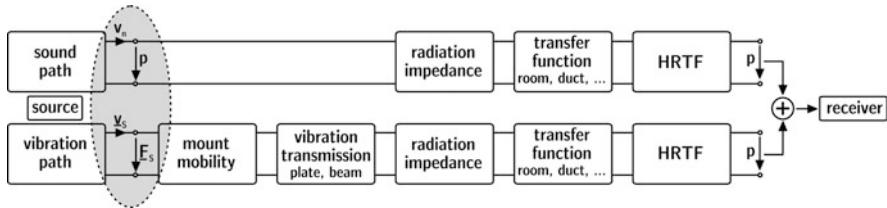


Fig. 15.1 Block diagram of BTPS

The concept of binaural synthesis is an important component of the concept of BTPS, and it was introduced in Sect. 9.4. Any monaural sound can be linked to a specific direction of incidence. In addition, sound carrying the characteristics of transmission paths can be modified in terms of their loudness and colouration. Based on these elements, the binaural transfer path synthesis is defined.

Tools for a transfer path analysis and synthesis are standard methods in the automotive industry. At the early stages of designing and developing automobiles, the engine and the car body and in particular combinations of both are studied with regard to the structural and acoustical performance of a car under development. The discipline of vehicle acoustics will therefore serve as leading example in this chapter.

The method of BTPA and BTPS (Genuit and Xiang 1997) is used to optimize products (automobiles and others). The crucial point is that the human perception of sound and vibration is part of the process. Thus, feedback from BTPS can help the engineer working on the combustion motor to study the acoustic and vibration performance in relation to the car body and the final recipient (the driver). The aim of BTPS is to determine the sound pressure at the ear canals of a test person (typically called driver's ear), as a sum of all sources and transfer paths. To obtain this result, the resulting signals are separated into their components of primary source signals and transfer functions. The crucial part of BTPS is the definition of appropriate interfaces.

The boxes shown in Fig. 15.1 are two-ports. The upper path represents airborne sound generation and transmission. The source may be modelled as volume flow source which acts on the radiation impedance. The effective velocity is calculated by coupling source and radiation impedance. Alternatively, the source can be described by using the sound power or the near-field sound pressure. The source quantity chosen is then coupled to a transfer function and to the listener's ears to obtain a binaural representation. Binaural transfer functions can be measured, predicted or calculated. Interface problems are accounted for by coupling the two-ports, as described in Sect. 10.3.1. Sound radiation and propagation in a room can as well be simulated by using geometric models and binaural filters as introduced in Sect. 11.5. In general, the lower path contains three degrees of freedom (x, y, z) or up to six in case rotational states are being included.

The two-ports include circuits of concentrated components such as springs and masses. For higher frequencies, this approach might not be adequate any longer. The system might require field quantities distributed in the geometry (for instance, the

velocity pattern on a plate or membrane), but the velocity distribution may still serve as input quantity if it is derived into a modal functional basis.

The boxes on the right side represent signal processing from the model of physical wave propagation into listener's dimensions. The sound pressure characteristics of the waves arriving at the ears, addressed to simplified wave fields (plane, spherical, etc.) with well-specified angles of incidence, levels and delays are transformed into a head-related binaural signal; see 9.4.

The efficiency of BTPS was proven in practice, particularly in vehicle acoustics (Sottek and Müller-Held 2007). In automotive engineering, BTPA and BTPS are standard methods to analyse and optimize acoustic signals. In noise control engineering for trains and aircrafts, it is used increasingly. The methodology can be applied to any other machine, household appliances and devices of daily life (such as, personal computers, dishwashers, or vacuum cleaners). But for a better understanding of these concepts, the following discussion focuses on the example chosen, which is automotive engineering.

15.2 Source Identification and Characterization

The specific behaviour of sources in general is related to the degree of freedom of vibration and to the radiation of airborne sound. In automobiles, the main source contributing to the resulting car cabin sound is the combustion engine, including gearbox, intake and exhaust system.² Simulation and auralization of combustion noise is a challenge, particularly, for modern engines with specific electronic injection control and optimized efficiency. The vibration and sound radiation is extremely complex with regard to spatial attributes, near-field effects of noise cancellation and amplification and the complex vibration injected into the car body through the engine mounts. The final result of interest is in any case the sound pressure in the driver's ear canals, as represented in the binaural signal. Figure 15.2 shows that the paths are separated into structure-borne and airborne paths. Structure-borne paths (solid lines) contain interfaces of mounts and parts of the car frame body and the radiation of beams and plates into the driver's cabin. Airborne paths (broken line) represent direct airborne radiation of the primary source of the powertrain and sound transmission to the driver's ear.³

Transfer path determination and interfacing to the source is the key to create BTPS. Here we have to face two problems:

- In the component of airborne sound the complexity of vibrational modes may be much larger than the number of transfer functions practically available. This is the main problem discussed in 15.2.1.

²When driving at high speed, of course, wind noise is another important sound source.

³This example illustrates a BTPS for a vehicle with rear engine. For cars with front engine it is also applicable, of course.

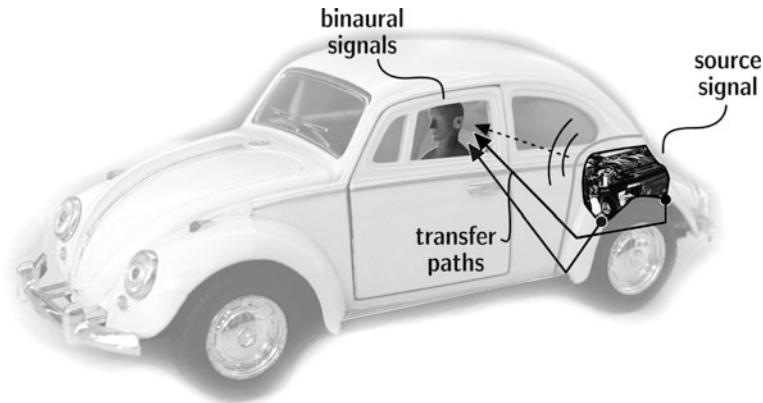


Fig. 15.2 Binaural transfer path synthesis for simulation of in-cabin sound

- In the component of vibration force, feedback between source and transfer function is significant. This is the main problem discussed in [15.2.2](#).

15.2.1 Airborne Sound Sources

We start again by discussing vehicle sound and the combustion engine. For the airborne sound transmission, the motor has to be characterized as primary source. The motor block is heavy and stiff. Vibration of the engine body induced by the engine operation is not affected by the acoustic load of the fluid medium (air). Thus, the source acts as ideal velocity source. We do not have to take feedback at the interface between source and transfer function into account. Instead severe difficulties will occur when the specific pattern of the source vibration has to be described in detail. The engine surface neither moves in a purely breathing mode nor like any other multipole. The modal pattern is a complex phenomenon which can only be studied by using structural models like FEM or a detailed experimental modal analysis.⁴ These methods can be used to determine the surface velocity distribution directly (Fig. 15.3).

The actual vibration pattern can also be approximated by using parameter sets of multipoles ([2.5](#)), spherical harmonics ([2.6](#)) or the technique of pressure mapping for acoustic holography (Williams 1999). The advantage of these techniques is the reduction of parameter complexity. They have in common that sound pressure measurements in well-chosen field points are used as input data for mathematical reconstruction of the surface velocity. In some cases this technique is called “acoustic camera”. It is clear that special conditions like an anechoic environment are

⁴By using laser Doppler vibrometry, for example.

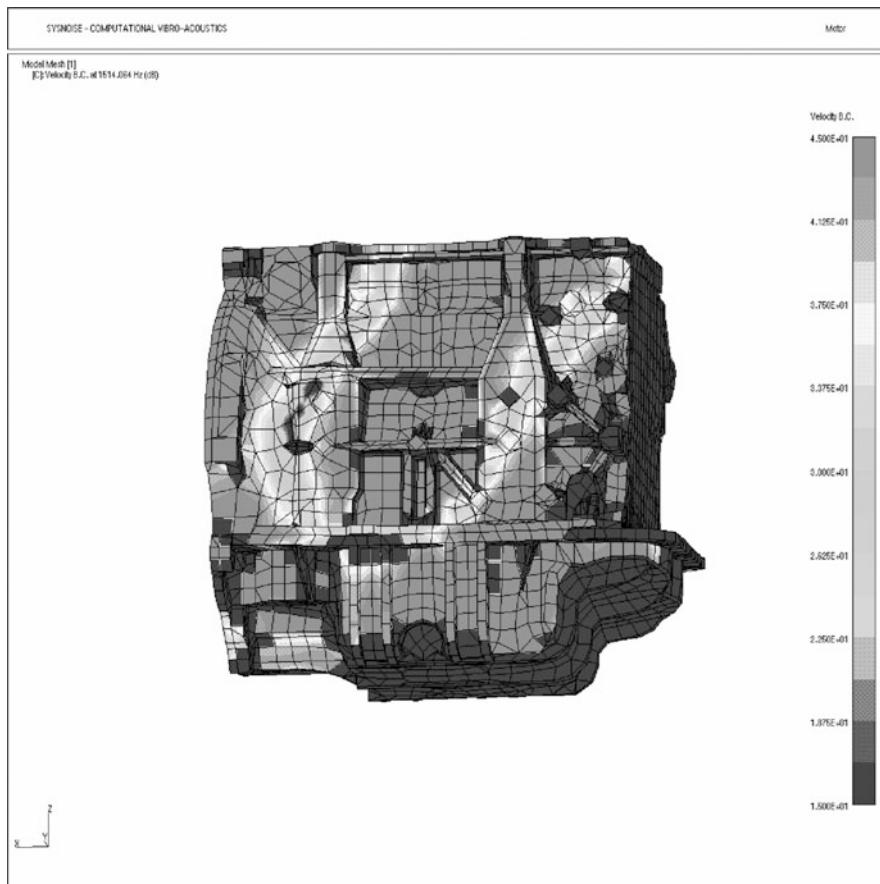


Fig. 15.3 Vibration distribution on a combustion engine. Measurement result obtained by using a scanning laser vibrometer

required. For further processing in BTPS, the parameter sets of multipoles, spherical harmonics or pressure maps are used as intermediate source data. But BTPS of airborne sound sources in vehicles is difficult as the distributed acoustic fields must be matched to a model of concentrated two-port elements for the transfer paths. Unless the approach is not extended to a finite element approach, the method applied is physically not consistent, but an approximation for acoustic engineering, the validity of which must be checked for each application.

When engine data are obtained *in situ*, i.e. in the engine compartment at the engine mounted into the car, the vibration on the surface and the source signal can only be estimated. This simplified approach is possible when the acoustic field in the engine compartment is dominated by cavity modes. The forced response excited by the engine is then close to the resonant cavity response which allows to model the

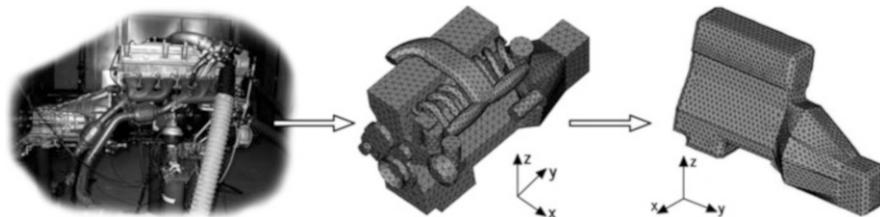


Fig. 15.4 Volume-conservative steps to create an engine model for BTPS. (Römer 2004)

source by estimation of the total volume flow, while the actual vibration pattern is actually ignored. However, the validity of this approach has to be checked carefully, too (Fig. 15.4).

In any case the engine must be represented by its actual volume, inserted into the model of the engine compartment and coupled to the transfer functions. Until a closed solution of coupling is feasible, proper transfer functions and interfaces must be chosen so that all relevant acoustic effects are captured. A modelling solution for the overall system is the only consequent approach with high precision. In this case we insert the engine mesh into a finite element model of the whole car including numerical wave prediction in air and structural modes which creates quite a challenge for numerical calculation.

Other sources of airborne sound in the vehicle are (a) smaller and (b) less complex regarding the modal response. These sources have to be described by the by an air flow, like in the intake and exhaust system, or by tonal components like in the electric generator or in the turbo compressor. In a (slowly) pulsating DC flow, the (acoustic) AC component of sound is difficult to measure. In the exhaust pipe, the “effective” source may also be given by an axial line source due to pulsating jet development. Tonal components require more precise phase responses and transfer functions. Numerous other practical aspects and difficulties are to be considered when dealing with the specific problem which cannot be discussed in detail here.

One important task to be explained in more detail is that of measuring sources and transfer functions *in situ*. When it comes to *in situ* measurement, two aspects are important. The interface between the source signal and the transfer function must be defined. If source signals from free field conditions in test facilities are used as input signals, the transfer function must match this condition. On the one hand, access to an equivalent measurement point in the engine compartment is difficult, and the choice of the set of measurement points is thus restricted to the accessible points. On the other hand, the relative and absolute calibration of the transfer function is difficult due to crosstalk between the measurement points and due to a somewhat arbitrary distance and near-field effect.

Nevertheless it is possible to use BTPA and BTPS on the basis of *in-situ* measurements, if the calibration is adjusted to test facility and *in situ* conditions. These calibration spectra can be obtained by experiment, too.

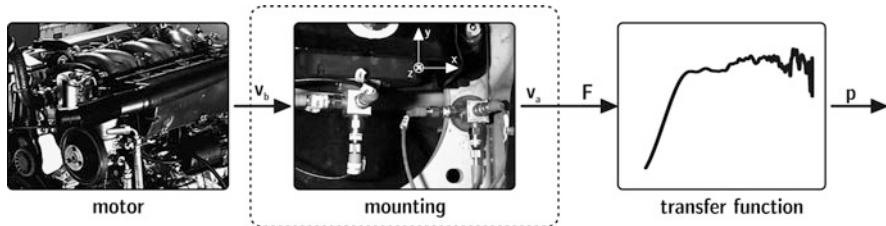


Fig. 15.5 Primary vibration source connected to mount and transfer function. (Sottek et al. 2005)

15.2.2 Structure-Borne Sound Sources

Structural paths are not as multidimensional as airborne sound fields.⁵ The number of paths and degrees of freedom are relatively low, and the signal flow over these paths can be modelled consistently inside BTPS by using the two-port approach. The engine is typically fixed on three points by using rubber or other viscoelastic mounts. With three degrees of freedom for translational and other three for rotational movement, the total force and torque can be modelled. The impedance of the car body, however, affects the vibration injected. The signal flow, thus, is to be modelled by taking feedback into account (see 10.3.1) (Fig. 15.5).

Well-defined load impedances are a prerequisite for a proper transfer path characterization. The primary engine vibration is usually measured in a free condition, assuming an ideal velocity source. The mount mobilities and other two-port parameters are determined in a special test stand for free and blocked-force conditions. Examples are shown in Fig. 15.6.

15.3 Transfer Path Characterization

For BTPS, sound and vibration propagation from the source to the receiver is modelled by transfer functions or transfer impedances. They are defined in each specific case on the basis of sound pressure spectra at the binaural receiver to the source quantity defined. For airborne sound components, the transfer functions are Green's functions (see 10.2); for vibration components typically two-port matrices are sufficient. Binaural filters are introduced at the end to account for the direction of sound incidence.

In straightforward approach, transfer function filters can also be measured between the source and the receiver's ears directly. The source arrangement (see 15.2.1 and 15.2.2) is thus connected to the receiver's ear by filters, the number of which coincides with the number of source signals recorded.

This measurement must be performed separately and independently for each source point to avoid crosstalk between the paths. In the example of a transfer

⁵Which is true at least for structures with point or line contacts.

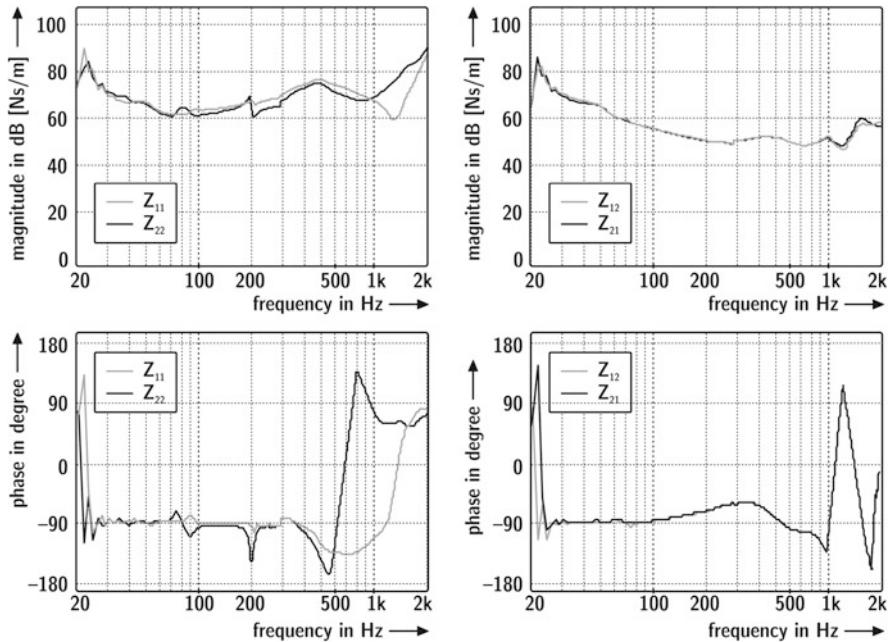


Fig. 15.6 Transfer impedance of engine mounts. (Dohm 2004)

function between the vibration source velocity and the ear sound pressure, the direct measurement requires a calibrated excitation of vibration, such as excited by a shaker or an impulse hammer, for instance. The sound pressure at the receiver is measured using a dummy head for all paths one by one.

In reciprocal arrangement, however, all paths can be measured in parallel. Due to vibroacoustic reciprocity (see also 10.2 and 10.3), the ratio

$$H = \left. \frac{P_{\text{receiver}}}{F_{\text{source}}} \right|_{v=0} \quad (15.1)$$

is equivalent to

$$H' = \left. \frac{v_{\text{eq},\text{source}}}{Q_{\text{eq},\text{receiver}}} \right|_{F=0} = H \quad (15.2)$$

With a calibrated volume source (reciprocal dummy head) at the receiver and accelerometer placed at the structure's source point(s), the transfer function is obtained (see Fig. 15.7).

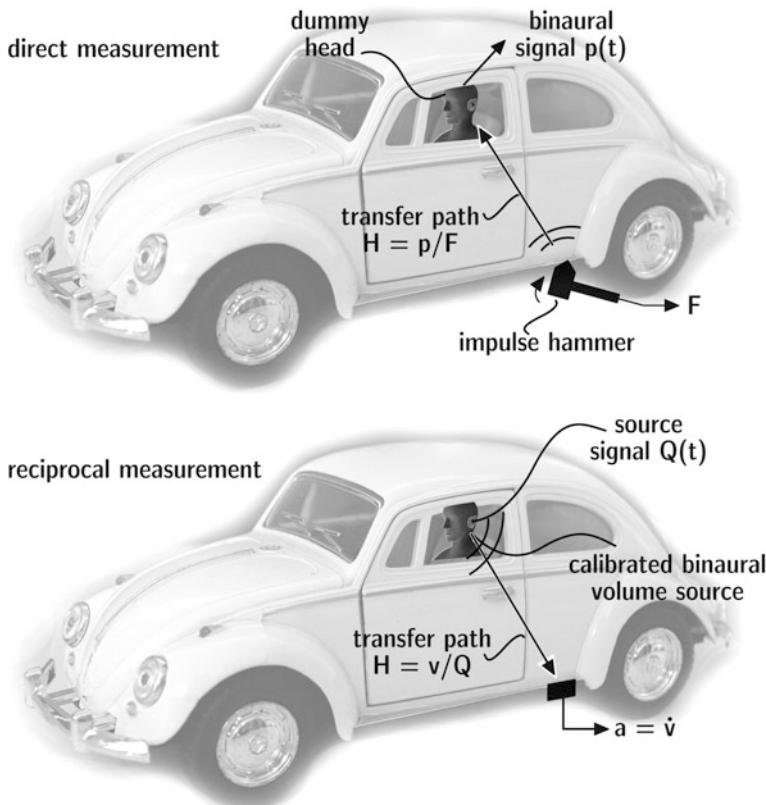


Fig. 15.7 Measurement of vibroacoustic transfer functions. (See also Sottek 2004; Sellerbeck 2003)

15.4 Auralization in BTPS

When the primary source is recorded in operational condition, for instance, an engine running at increasing speed (rpm), this signal serves as input for binaural transfer path auralization. The filters and coupling impedances are used in network components in the two-port mode described above. The signals are processed in convolution units.

The big advantage for the sound engineer is that transfer paths can be added, switched off or modified, depending on the specific choice of system parameters. Other sources may be added too, like wind noise or tire noise. Integrated into a driving simulator, the binaural transfer path auralization allows detailed psychometric tests on sound character and sound design targets (Fig. 15.8).

As described above, the synthesis is based on input measurements in the engine compartment of cars or on an engine test rig. It is typical, however, that during the

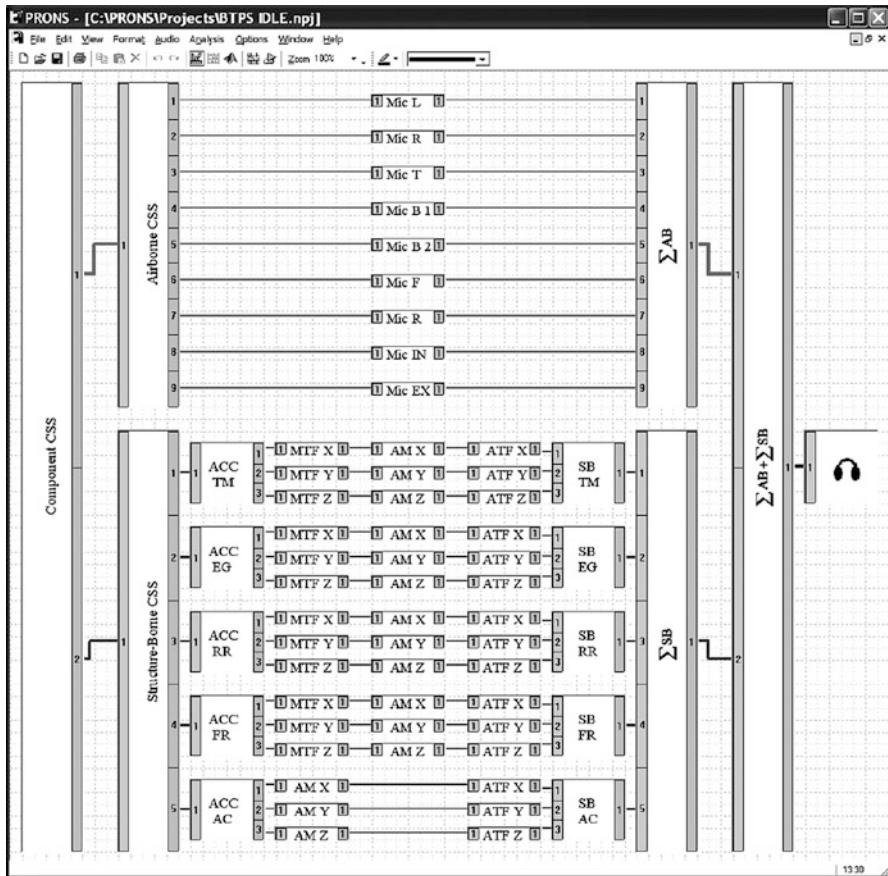


Fig. 15.8 Example for a BTPS software user interface. (After Sottek et al. 2004)

development of new engines, car prototypes are not yet available. That is why recent extensions of BTPS involve more and more simulated sound transmission data for airborne and structure-borne paths or a combination of measured and simulated data. The challenge for future work is the combination of finite-element models, SEA models and two-port models in the different frequency ranges. Also of interest is the variability of input data on material properties and the variability of junctions, related to the variance of those parameters in the production process.

The increasing sloped curves in Fig. 15.9 illustrate an increasing rpm. A combustion engine is mounted into the car body by using two different engine mounts (different stiffness). The time and the frequency, respectively, are depicted on the abscissa and the ordinate, respectively. The sound level in dB(A) is presented in grey scale. On the right-hand side, the resonance effect illustrates a sub-optimal impedance match between engine mount and car structure. Between 800 Hz and 1 kHz, an

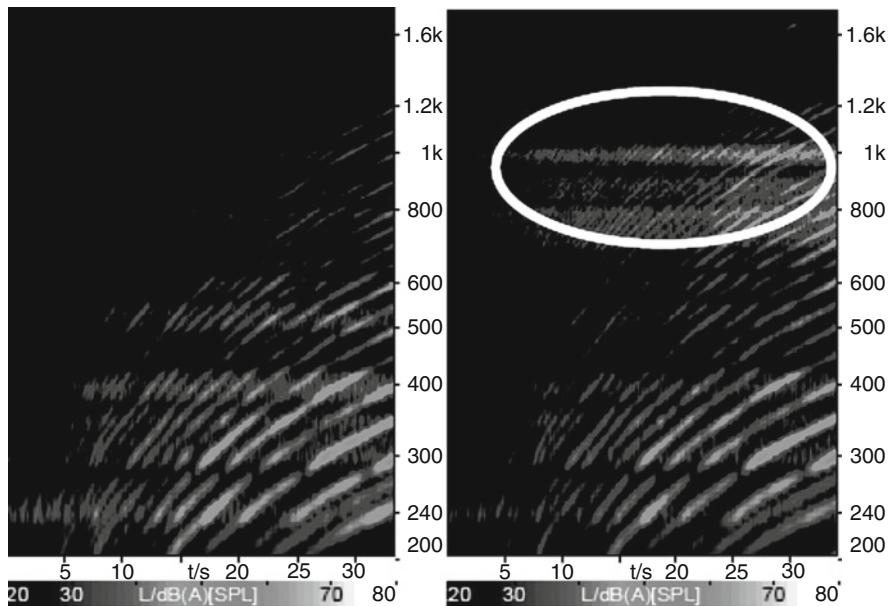


Fig. 15.9 Example for spectrograms showing the effect of changing engine mounts by using BTPS. (After Behler et al. 2006)

rpm-independent spectral maximum appears which creates an unwanted vehicle sound character. Hence the target sound is better achieved by using the mount with the result shown on the left side.

Chapter 16

Filter Construction for Real-Time Processing



16.1 Introduction

Virtual environments should be multimodal and interactive. The technology used for simulation and reproductions must take into consideration the aspect of several sensory inputs in real time and the interaction with the person who is embedded in the virtual environment. In order to achieve a real-time performance, specific runtime conditions must be taken into account in order to stay within acceptable limits for latency and update rates. For the acoustic component of VR, we can use information from psychoacoustics on these limits. In this chapter, the implementation of software and hardware tools is introduced which provide real-time performance and interaction: head tracking and adaptive filtering.

It should be noted that some requirements in developing VR scenarios may be contradictory. To illustrate this fact, we assume a sound source on a circular rotation path around a listener. The source is permanently moving. A high update rate will create a smooth and continuous transition from location to location on the circular path. A high update rate, however, allows only little calculation and processing time for the simulation of sound propagation and binaural synthesis. The new geometric situation (new source position) may be available in one step of simulation; the delays are included between the actual position and the reproduced binaural sound pressure. Interaction will be disturbed since the system's feedback is delayed. However, if the computation time available for simulation is larger, the update rate is slower, and this might affect the smoothness and continuity of the sound stimuli may be affected.

Update rates of 60 Hz and total delays of 50 ms are considered acceptable for acoustic virtual sound. But system components of VR such as head trackers, audio hardware and filters already introduce latency. Accordingly, there is only little time left for the acoustic simulation (acoustic rendering and reproduction). Limited computer performance therefore prevents the acoustic component of VR from reaching its full authenticity. At the moment, however, the performance is sufficient to create plausible acoustic scenes.

There are indications that the subjectively perceived timing and coincidence between visual and auditory cues allows additional delays for the auditory path. At 30 ms delays of the acoustic path, a good perceived coincidence of the stimuli is still found. This might be the result of a faster neural processing for the visual path, compared with the auditory path. In other words, acoustic rendering may require a 30 ms longer latency than visual rendering (Kohlrausch et al. 2006).

What is “real-time processing”? An important factor to be discussed in this context is the above-mentioned interactivity. This means that the dynamic changes introduced by the active user must have an immediate effect on the perception. The latency, accordingly, must be sufficiently small so that the listener is not disturbed or irritated (Brungart et al. 2004). Ideally case, they are below the just noticeable differences (see 6.3). Furthermore, the frame rate (update rate), defined as the time the acoustic input is changed, the re-calculated simulation and the output presented must be sufficiently high to ensure a smooth and continuous scene. These requirements are interconnected with the parameters responsiveness and smoothness. If the update rate of the simulation and the perception of continuous movements are just set to the limit of responsiveness, the reciprocal is the actual allowable processing time for one step of the simulation. For example, if we set the update rate to 50 Hz, the maximum processing time for simulation is 20 ms.

The key technology for real-time systems can be divided into the following components: sound rendering (creating of auditory cues by software algorithms) and sound reproduction (3D audio technology). From the viewpoint of the user, the technical system is a closed and abstract box of numbers. The user interface which provides the translation of numerical data into stimuli of sound and vision is therefore crucial. It should be noted that rendering and reproduction are closely linked, and depending on the data formats chosen, the spatial audio stream is created as a sound pressure signal at the two ear canals of one (binaural) listener or for a certain area in the listening space as a pressure field without reference to a specific listener point. The binaural rendering technique is discussed in more detail in Sect. 16.3. Before this, spatial audio rendering will be introduced in the example of ambisonics-like technology, where the simulation part is used to deliver sets of plane waves from various directions representing the acoustic response of the environment and the source present in this environment.

16.2 Real-Time Spatial Audio Rendering

According to the considerations in Sect. 11.9.1, acoustic environments can be decomposed into sets of plane waves arriving from various directions. The amplitudes and phases of the plane waves must represent the environment as such, including the direct paths from sources and the reflections. This concept is obviously not applicable in too close distances between sources and receivers. Directional plane waves are the basic components of decompositions in spatial audio representations such as Spherical Harmonics (Sect. 2.6).

Changes required in real-time processing therefore include updated plane-wave decompositions by spherical Fourier transform of the incoming sound waves. The loudspeaker filters are then obtained from the new set of encoded surround sound. Cross-fading of filters is required for the case of a dynamic (changing) incoming sound wave field due to changes of the environment or source movements, for example. There is no need for receiver-related updates, as the sound field is reproduced (at least in approximation) in the listener space inside the surrounding loudspeaker array.

16.3 Real-Time Binaural Rendering

Before we take a look at full interactive binaural simulation in real time, an intermediate step is required. Binaural sound can be preprocessed or recorded for some locations and orientations of the listener. Those situations can be created in virtual environments, or they can correspond to real rooms. In both cases the binaural impulse responses are available in a certain grid in lateral and spherical coordinates. In a replay situation,¹ dry sound is convolved with the valid binaural impulse response for the actual position and orientation of the listener. The listener movement is tracked. The best matching binaural filter for the position and orientation is therefore chosen for convolution.

An important aspect, of course, is inaudibility of fading between the filters and a sufficient resolution of the impulse response database. Usually dummy head HRTF are used as a basis for simulation or measurement.

Head rotation is the quickest and most abrupt movement. In this case, translational movement occurs at the user or at the source. Translational movements are in a speed range of 30–50 cm/s; head rotations may be as quick as some 45 degrees per second. If we use this data as a limit, the filter update rate and the necessary resolution can be estimated. The azimuthal resolution is then about 1 degree, the time available for filter update 20 ms (Lentz et al. 2007). As the movements are limited in their speed, required update rates can be estimated. Table 16.1 shows an example for a set of update rates in the specific region temporal regions of the binaural room impulse response.

16.3.1 HRTF in Multiple Degrees of Freedom

Advanced developments allow dummy heads also to introduce multiple degrees of freedom in terms of the head-to-shoulder orientation, which is best illustrated by

¹Also referred to as “walkthrough” (Dalenbäck and Strömberg 2006).

Table 16.1 Proposal for update intervals for some interaction situations of listener and source movement

Action	Time available for filter update	Filter content to be updated
Source rotation	35 ms	Source directivity
Head rotation	35 ms	Binaural processing in listener coordinates – New HRTF
Translational listener or source movement > 0.25 m	700 ms	Binaural processing in listener coordinates and specular part of the room impulse response
Translational listener or source movement > 1 m	3 s	Binaural processing in listener coordinates, specular part and scattering part of the room impulse response

After Lentz et al. (2007)

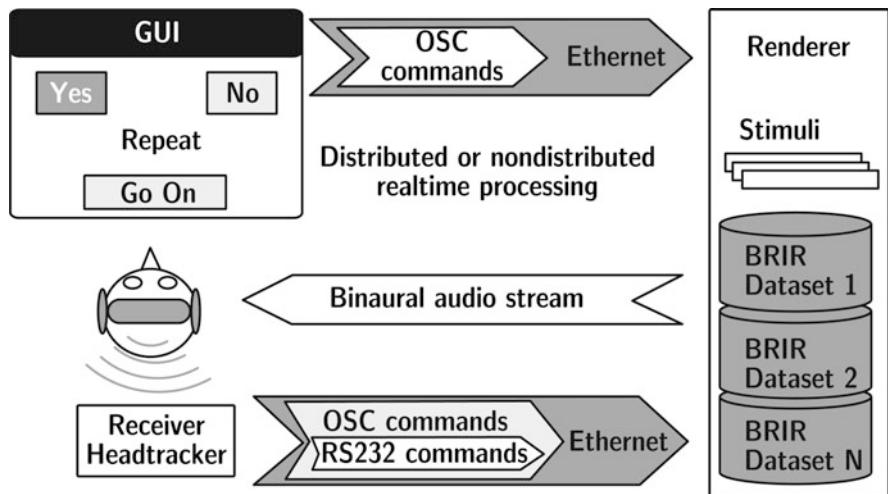


Fig. 16.1 System for dynamic real-time auralization of pre-calculated or measured binaural impulse responses. (After Moldrzyk et al. 2004; Lindau et al. 2007)

allowing the “yes” and “no” movement of the head, compared with a standard fixed dummy head (Moldrzyk et al. 2004).

Multi-DOF heads were used in dynamic (adaptive) replay situations of binaural recordings. In terms of the plausibility and naturalness, the results were excellent. The test subjects could on average not identify whether a stimulus was presented via auralization (head-tracked binaural real-time convolution), or whether it was the original. This test was performed in-situ in the real hall (see (Lindau et al. 2007) and Chap. 18) (Fig. 16.1).

Headphone reproduction without head-tracking would be head-related and not room-related (Dalenbäck and McGrath 1995; Dalenbäck and Strömberg 2006). To obtain sufficient presence in the virtual environment, only a coupled system of head

tracking and binaural synthesis of room-related sources, image sources and reverberation is appropriate. In fact, an extension towards room-related coordinates creates a very big effect of enhancing plausibility and immersion, even if non-individualized HRTF are used. This finding underlines the importance of dynamic cues of localization² and the necessity to implement this feature into binaural systems (Mackensen 2003).

In order to account for dynamic cues, interaction and adaptation of filters, the position, orientation and motion of the listener must be known. Systems which provide this information are head trackers. Systems for measuring the head position and orientation are well known. In dynamic virtual reality systems, head trackers are used for calculation of the reference point for the stereoscopic projection. Like for the stereoscopic projection, the room environment is the invariant coordinate system.

16.3.2 Head Tracker

For listener-centred 3D video projection and binaural reproduction, several tracking devices are available. The technology can be based on electromagnetic, optical, mechanical and also acoustic (ultrasonic) principles. Today, more and more camera-based systems using markers attached to the stereo video glasses (for IR camera) and systems with image recognition and eye tracking are of interest since they do not require any sensor or antenna device mounted onto the head. Precision and latency of the head-tracking system are the most relevant parameters for the applicability in VR technology.

16.4 Fast Convolution

Real-time convolution is a specific problem, especially in rendering and synthesis systems where many sources or many channels must be processed at the same time. In dynamic room auralization, impulse responses must be changed rapidly without audible transitions or “clicks” (McGrath 1996). Segmented convolution (see 9.3.1) is a good candidate for this task, typically in conjunction with the overlap-add (OLA) or overlap-save (OLS) scheme (Oppenheim and Schafer 1989). Its input-to-output latency, however, depends on the length of the filter impulse responses, so that the block length must be optimized with regard to update rate and latency. Low latency convolution can also be implemented on multi-processor systems which actually follows the trend in PC hardware platforms. Furthermore, fading between impulse

²Human beings also use these cues under normal listening conditions, when they slightly move the head in order to solve the problem of front-back confusion.

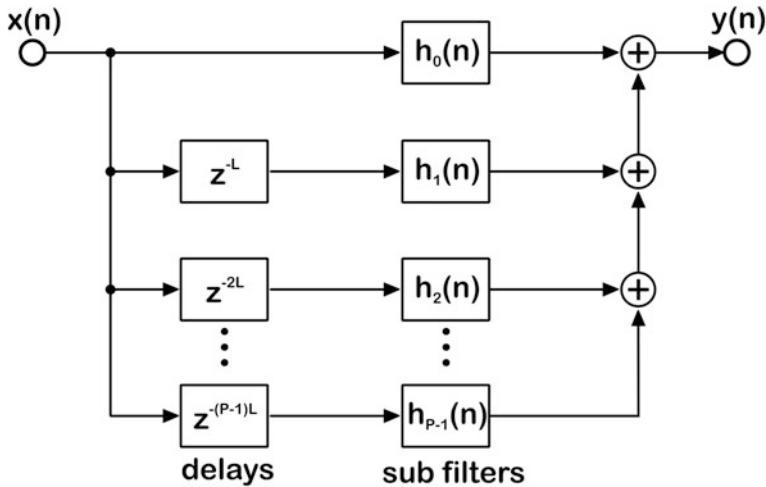


Fig. 16.2 Partitioned filtering. (After Wefers and Vorländer 2014)

responses must also be optimized to avoid coloration (comb filter effects), clicks and abrupt changes in distance or direction of the direct and early sound.

16.4.1 Block Processing with Partitioned Convolution

Low-latency real-time filtering is usually implemented by splitting the filter impulse responses into a set of subfilters and processing these subfilters by FFT-based convolution. In so-called “uniformly” partitioned subfilters, both operands, signal and filter, are usually partitioned with the same block length (Wefers and Vorländer 2014). A low input-to-output latency is achieved by processing short block lengths. In order to obtain significant benefit from using FFT techniques, larger block lengths are preferable, but this, however, results in larger latencies. These methods are well-matched if the filter length N is in the same order of magnitude as the block length B of the signal. For the case of much longer filters, $N \gg B$, zero-padding must be incorporated, making these algorithms inefficient (Fig. 16.2).

A significantly more efficient method is the non-uniformed partitioned convolution. The implementation, however, is more complex as it involves problems of data scheduling of the sub-convolutions. If the filtering is subject to changes due to real-time interaction, a dynamic scheduling is required. More information and in particular very useful programming guidelines are found in (Wefers 2014) (Fig. 16.3).

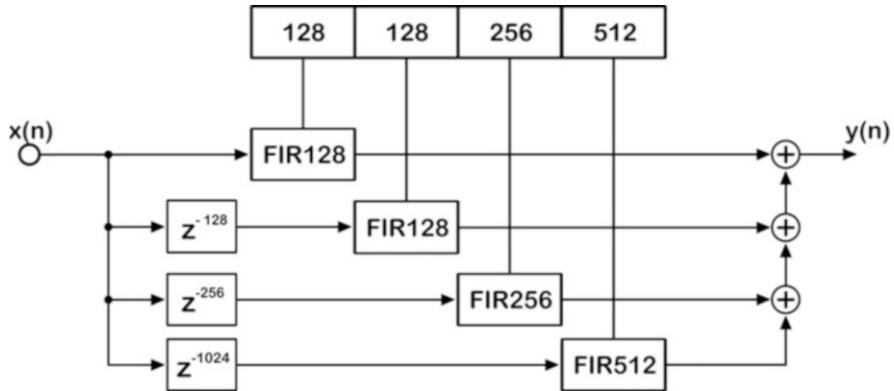


Fig. 16.3 Example of FIR filtering with a non-uniformly partitioned impulse response. (After Wefers 2014)

16.5 Room Acoustical Real-Time Auralization

In cases where the sound event is related to a virtual room, the geometric situation is characterized by a number of polygons, the acoustic data of which represent various coefficients of absorption and scattering. A “wall” is considered as infinitely large plane, and it is described by a point in the plane and a normal vector. The coordinate system defined by the plane itself can be used for the process of ray tracing and image source simulation.

The valid room boundaries are defined by polygons which include the acoustic data of absorption and scattering. The polygons as such are defined by a list of vertices. It is advantageous to divide the polygons into convex parts, to speed up the process of ray intersection test.³

16.5.1 Source and Receiver

The source and receiver characteristics are determined by their position, orientation and directional pattern (Fig. 16.4). The information on the source should contain data on the distance to the receiver, thus allowing the delay of sound propagation and the air attenuation to be known. Source and receiver have six geometric degrees of freedom, three data of translation and three of rotation. The determination of the position is best expressed in Cartesian coordinates as is advantageous for the room boundaries. The orientation and the directional dependence are defined by three angles. HRTF are implemented in the same way as described above in 9.5.

³Test of “point in polygon”.

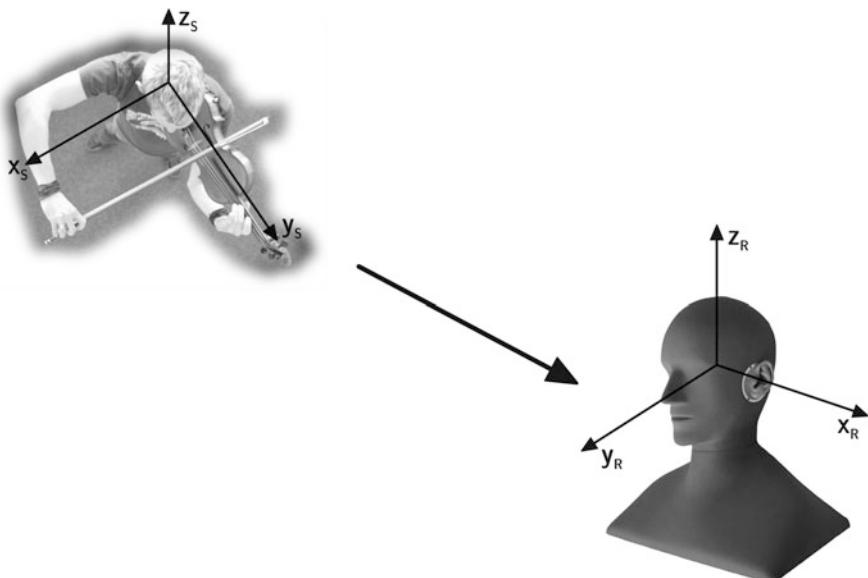


Fig. 16.4 Directional characterization of source and receiver

16.5.2 Real-Time Processing of Image Sources

Image sources are given by their position in 3D coordinates and by a set of parameters including the order, the total time delay, the wall reflections represented and the corresponding reflection factors or absorption coefficients. A source-related coordinate system for translational or rotational transformation is important as well (see below).

The specular part of the binaural room impulse response is composed of direct sound and specular reflections (see 11.4). The image source contributions can be determined best in the frequency domain. The spectral resolution of input data like absorption coefficients, however, is hardly useful for signal processing. HRTF data is stored in discrete spectra⁴ with a resolution of ≈ 300 Hz. The spectral density of HRTF must be identical with the spectral density of the transfer function. If coefficients of wall absorption, air attenuation or source directivities were used from octave band data, the missing frequency lines must be inserted by interpolation. Towards higher octave bands this requires more and more interpolation finally resulting in a linear resolution with constant frequency spacing.

The same problem occurred when a modal response of a transmission system was calculated with high-frequency resolution (possibly pure-tone excitation makes this

⁴HRIR in FIR filters of 128 taps represent discrete spectra of $\Delta f = 344$ Hz (at sampling rate of 44.1 kHz).

high resolution necessary); other spectra like for HRTF must be interpolated or represented by longer FIR filters.

The room impulse response is composed of direct sound and reflected sound. These components derive from simulations (see 11.9). In frequency domain, the superposition of the components reads:

$$\underline{H}|_{\text{left,right}} = \sum_{i=1}^N \frac{e^{-j\omega t_i}}{ct_i} \cdot \underline{H}_i(\theta, \phi) \cdot \underline{H}_{i,\text{air}} \cdot \underline{\text{HRTF}}(\vartheta, \varphi)|_{\text{left,right}}, \quad (16.1)$$

or converted into the corresponding binaural impulse response:

$$h(t)|_{\text{left,right}} = \text{IFT}\left\{\underline{H}|_{\text{left,right}}\right\}. \quad (16.2)$$

with $h(t)$ denoting the binaural room impulse response, $j\omega t_i$ the phase lags due to retardation of the i th reflection, $1/(ct_i)$ the distance law of spherical waves, \underline{H}_i the source directivity (Sect. 2.5) in source coordinates (θ, ϕ) , $\underline{H}_{i,\text{air}}$ the low pass of air attenuation (Sect. 3.7) and HRTF the head-related transfer function (Sect. 6.4.1) of the sound incidence in listener coordinates at a specified orientation (ϑ, φ) .

In this equation, the binaural impulse response is valid for the time frame corresponding to the input parameters used in the simulation. It is clear that in dynamic scenes the parameters are changed. The binaural impulse response has to be changed accordingly.

16.5.3 Room Subdivision

It is vital that the update of reflections is as quick as possible, even in complicated and detailed environments. Strategies for speeding up the image processing are crucial. To this end, approaches to reduce of the computational load are available. They can be interpreted as tree structure either in a logical link between image sources or in a spatial sense (see 11.4.5).

16.5.4 Dynamic Interaction

With the cloud of image sources being known, a translation or rotation of source or receiver can be calculated by elementary geometric transformations. The positions of image sources are created for a fixed position (time invariance) of the original source. If this position changes, which is quite obvious in dynamic situations in VR environments, all image sources must be recalculated. However, we can make

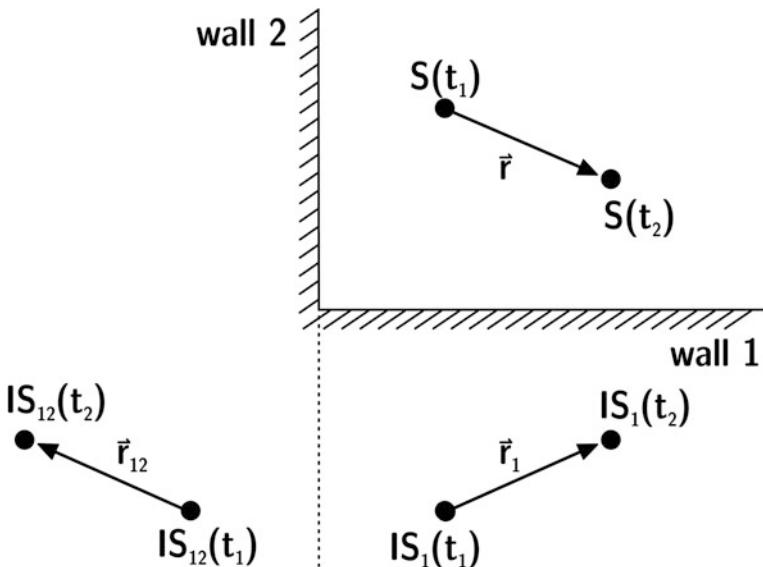


Fig. 16.5 Translation of image sources

use of geometric facts. It is not required to start the imaging process all over again. Instead it is only necessary to rotate and translate the image sources given.

The difference vector between source and receiver⁵ is used as a translation vector.

Similar to the mirroring of vertices at walls (see Sect. 11.4.1), we can calculate the mirrored translation vectors. As illustrated in the example in Fig. 16.5, the source movement from time frame t_1 to time frame t_2 is described by a translation vector, \vec{r} .

A three-dimensional vector can be modified in location and orientation by using a 3×3 matrix operation. A matrix has to be found which provides the first-order image source being placed to the position corresponding to the actual movement of the original source.

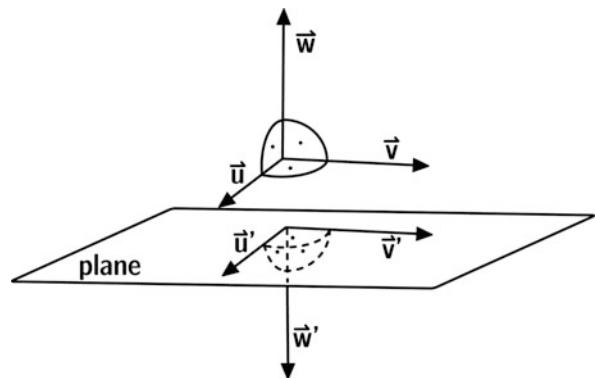
$$T = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix} \quad (16.3)$$

This 3×3 matrix contains nine unknowns. The nine required equations are constructed from the actual movement of the original source in the orthonormal system of the mirroring wall plane.

The vector elements \vec{u} , \vec{v} and \vec{w} , shown in the figures, can be used to calculate the coordinates of the image space.

⁵Representing the direct sound path

Fig. 16.6 Mirroring the wall orthonormal system



$$\begin{aligned}\vec{u}' &= \vec{u} \cdot T = \vec{u} \\ \vec{v}' &= \vec{v} \cdot T = \vec{v} \\ \vec{w}' &= \vec{w} \cdot T = -\vec{w}\end{aligned}\quad (16.4)$$

Hence, vectors parallel to the wall plane are invariant, while the normal vector is mirrored (Fig. 16.6). With these nine equations the transformation matrix of any shift vector can be determined. Figure 16.5 shows as well that the shift vector of image source 12 is obtained by mirroring the shift vector of image source 1. The total shift vector is thus the result of matrix multiplication of all wall matrices involved.

$$T_{\text{total}} = \prod T_{\text{walls}} \quad (16.5)$$

and accordingly

$$\vec{r}_{i,\text{new}} = \vec{r}_i + T_{\text{total}} \cdot \Delta \vec{r}_{\text{source}} \quad (16.6)$$

As long as the geometry of the environment is invariant, the transformation matrices are also invariant. Using the matrix operation to calculate new image source positions is therefore faster.

16.5.5 Real-Time Modelling of Reverberation

It is unavoidable that image source processing in real time is limited so that the impulse response must be truncated. The late reverberation has to subsequently be added using other techniques, such as reverberation processors. Reverberation processors are well known in audio engineering (Griesinger 1989; Blesser 2001). In order to get a first effect and a proper reverberation time of the scene, these reverberation machines can be used.

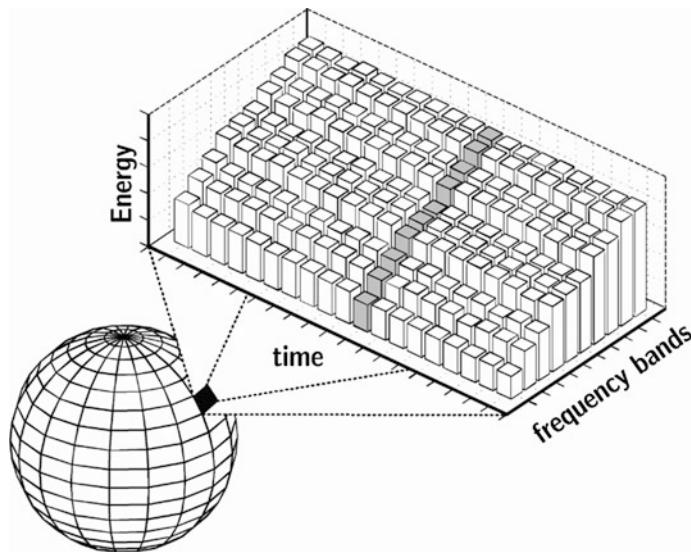


Fig. 16.7 Receiver-related sphere divided into directional groups with energy decay histograms for every group and frequency band. (After Schröder et al. 2007)

The first digital⁶ reverberation processors (Schroeder 1961) were based on allpass and comb filters. The problem, however, is that periodicities may occur due to recursive implementation. Impulsive sounds fed into these algorithms have a somewhat “metallic” character. Modifications of the allpass filter approach involved integration of early reflections by delay lines and artificial binaural cues.

For a physically consistent concept, the reverberation should have a clear correspondence to the room under test, not only to its size and absorption but also to its shape and absorption distribution. The reverberation process may even be distorted in a sense of “non-Sabinean” behaviour⁷. In these cases, artificial reverberation is not sufficient, and a more detailed estimation of the specific reverberation process is required.

In many algorithmic solutions of reverberation estimation, the temporal resolution or the spatial resolution is affected by limitations by computation time. In best case, the spectral, temporal and spatial distribution can be predicted with acceptable accuracy, but not in its fine structure. The distributions are present as envelopes of the actual energy decay, averaged over some 10 ms and in certain sections in 10×10 degrees of the spatial angle. These resolutions are actually sufficient to describe the late reverberation (see Fig. 16.7).

Methods of ray tracing, radiosity, waveguides, randomized tail-correction or other stochastic methods are well qualified for obtaining late reverberation spectral

⁶Historically also implemented in analog hardware

⁷This would apply for coupled rooms or rooms with flat or long shape

and temporal envelopes of energy decays. The necessary sample-related fine structure in the sound pressure impulse response required for creating convolution filters, however, must be generated in post processing (see 16.6). To achieve a good approximation, certain assumptions must be made concerning the reflection density and spatial distribution. It is known that the reflection density in rooms increases with t^2 in a diffuse field impulse response. Information about the perceptual aspects of the spectral, temporal and spatial cues in room impulse responses will be of interest for future simplifications of real-time reverberation modelling (Blesser 2001). Models of binaural neural processing may also help in this respect (Braasch 2005).

Figure 16.7 shows an illustration of the late reverberation energy decay plotted in time and frequency domain for a set of spatial angle intervals. The simulation model used is a mixed approach of physical room simulation with stochastic estimation of the late decay envelope. The fine structure is generated by using a Poisson process of pulses.

16.6 Real-Time Auralization of Sound Propagation in Buildings

In architectural applications such as a virtual walk in a complete building, the rendering process can hardly be performed on the whole physical model of the rooms and the wall partitions. An apparently simple case when a person is leaving a room and closing a door requires complex models of room acoustics and sound insulation.

As described in Chap. 13, a well-suited method for the determination of the transfer function between the source and the receiving room is the SEA approach (Gerretsen 1986). The sound energy is considered by its magnitude, its flow through the building elements, the energy exchange between adjacent building elements and the respective energy losses. With an appropriate model of the radiation path from the wall to the listener in the receiving room, the total signal can be obtained after superposition of the direct and flanking paths (Fig. 16.8).

The approach of sound transmission paths directly leads to the concept of filter networks. For each path, the sound pressure in the receiving room can be approximately described by the sound pressure source signal multiplied by frequency-dependent filters. Subfilters correspond to the transfer functions between the source room and the radiating walls. Spatial information such as binaural transfer functions apply to the final propagation path between the radiating element in the receiving room and the listener.

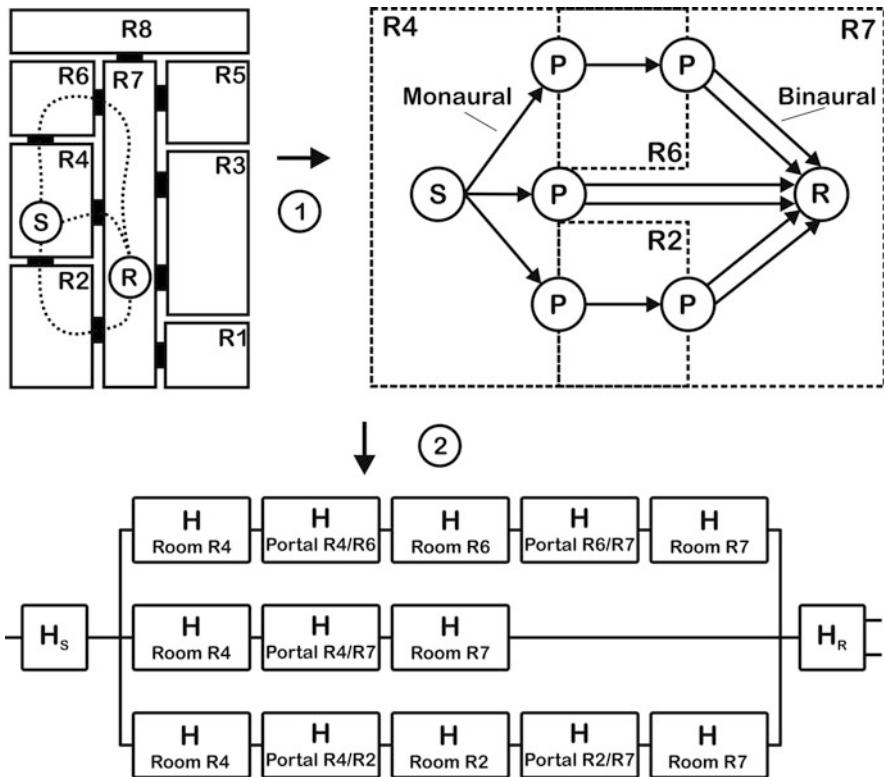


Fig. 16.8 Example of tracking sound propagation paths throughout an office floor and constructing a corresponding filter network. (After Schröder 2011)

16.7 Hybrid Real-Time Auralization Algorithms

To start with the discussion on hybrid real-time models, we should study the psychoacoustically relevant temporal resolution in the impulse response and the spectral details in the frequency response.

At least four regions can be distinguished: the early part and the late part in the impulse response and the frequency range below and above the Schroeder frequency (Fig. 16.9).

In all regions, the complexity must be restricted to the just required minimum to achieve a certain plausibility. Statistically many overlapping reflections and modes denote the degree of diffuseness. Accordingly, mean values or envelopes in time intervals or in frequency bands are sufficient to describe the perception, while the details in the sampled data can well be inserted by using random data with physically consistent statistics such as reflection or modal densities.

In the early part of the impulse response, however, human hearing is sensitive to details, as this part determines the loudness and the localization. The low-frequency

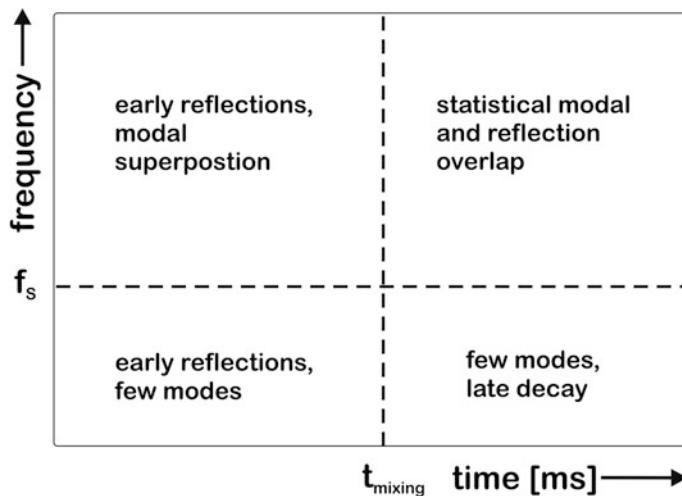


Fig. 16.9 Sound propagation model regions. Numerical methods are preferred for the lower left region, whereas statistical methods are candidates for the upper right region

component with distinct modes determines the perceived sound character significantly. In all parts, different algorithms might be superior in terms of computational performance and efficiency.

Furthermore, capability of interaction, instantaneous system reaction (responsiveness) and smoothness of the stimulus presentation in real-time processing must be kept at a high level, to ensure a high level of immersion.

(Lokki 2002) presented an auralization method with dynamic rendering based on the image-source method and the late reverberation modelling by a time-invariant recursive digital filter structure. An important part of this work is the evaluation by means of auditory models and listening tests while comparing at the same time auralization with the real room under test. This and other studies are oriented towards a psychoacoustic evaluation rather than focusing on the technical aspects of simulation and auralization. As discussed above, the psychophysical limits are more relevant than limits characterizing the system performance in a purely physical sense. As result of decay perception tests in (Meesawat and Hammershøi 2003), a limit of 140 ms is identified after which the specific receiver-related reverberation is irrelevant and can be replaced by a statistical decay approach. When analysing the user's behaviour in the VR system, several aspects can be noted (LaViola 2003). These results are not only interesting when it comes to determining the absolute speed of the movements and permitted latency but also for predictive algorithms of filter updates.

These data are of course subject to future development, which will allow faster dynamic scenes. Current research therefore deals with aspects such as faster convolution processors and the limits of time-invariant system theory.

16.8 Real-Time Signal Processing for Moving Sound Sources

Moving sound sources are characterized by their sound emission data and additional metadata describing the trajectory of the movement. The distance and velocity with regard to the listener are, hence, time-variant. The approach of auralization can still be based on convolution as long as the corresponding audio stream and the propagation filter are processed in blocks with an approximatively time-invariant approach within each block. The main challenge is now to provide a smooth transition of the resulting sound from block to block, including re-sampling for the compressed or stretched frequency spectrum according to the Doppler effect (see 3.8).

A physically consistent data structure is required for calculation of the propagation delay and the effective arrival time of sound at the listener position. Just very few trajectories can be described analytically (e.g. straight lines or circles). Mostly, in fully interactive systems, source trajectories are represented in samples of the position as a function of time (see 12.4). In a simple example, the source might move on a straight line at constant speed. It is then located at the discrete time and position samples $(t_n, x_n), (t_{n+1}, x_{n+1})$. The velocity, v , can be calculated from $v = (x_{n+1} - x_n)/(t_{n+1} - t_n)$. A generalized a solution for the propagation delay, τ , is

$$\tau(t) = \frac{\vec{r}_{SR}(t) \cdot \vec{v}_S(t) + \sqrt{\left(\vec{r}_{SR}(t) \cdot \vec{v}_S(t)\right)^2 + \left(c^2 - \|\vec{v}_S(t)\|^2\right) \vec{r}_{SR}^2(t)}}{\left(c^2 - \|\vec{v}_S(t)\|^2\right)} \quad (16.7)$$

$\vec{r}_{SR}(t)$ denotes the relative position vector between source and listener. The vector $\vec{v}_S(t)$ is the direction and velocity of the source. The solutions of the equation above are then found in each sampled part of the trajectory. The problem, however, is that linear interpolation of the trajectory may lead to discontinuities of the source velocity and acceleration. The propagation delay and the corresponding Doppler frequency shifts get inherent discontinuities, too (Fig. 16.10).

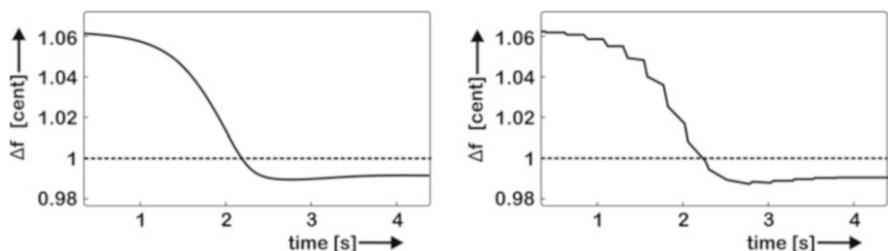


Fig. 16.10 Doppler frequency shifts Δf of a car trajectory in a street curve. Left, analytic; right, constant velocity

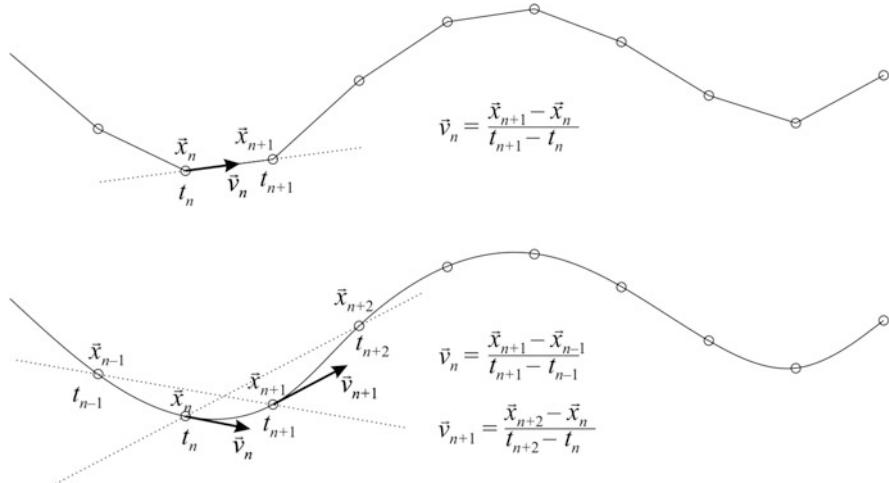


Fig. 16.11 Top linear interpolation (CV model); bottom, Catmull-Rom spline interpolation (CJ model). (After Wefers and Vorländer 2015)

Movements with rapidly changing velocities cannot be adequately described with such constant velocity (CV) interpolations, as the discontinuities lead to audible frequency artefacts. Better suitable are “constant jerk” models (CJ) which are based on spline interpolation involving the position x , the velocity v and the acceleration a . j is the derivative of the acceleration, the jerk.

$$\vec{r}(t) = \vec{x}_0 + \vec{v}_0 t + \vec{a}_0 t^2 + j t^3 \quad (16.8)$$

The improvement of determination of the instantaneous velocity is illustrated in the figure (Fig. 16.11).

Concerning applications, it can be concluded that railbound vehicles with relatively small changes in the velocity and direction can well be auralized by using CV models, whereas scenarios with aircrafts and cars require more accurate approaches such as the CJ model.

16.8.1 Latency Reduction by Prediction

For real-time auralization, it is essential that the propagation delay, $\tau(t)$, is determined for each frame of samples in order to take measures (estimates) of the position at the beginning and at the end of the frame. Audio processing frame rates, however, are typically much faster than the frame rates from position tracking systems (100 ~ 1500 Hz vs. 30 ~ 240 Hz). This means that the next information about the

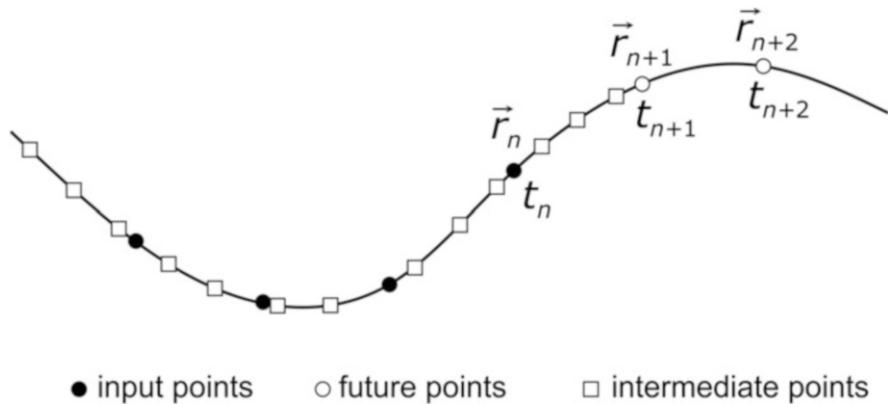


Fig. 16.12 Extrapolation of source positions. (After Wefers and Vorländer 2015)

current source position is delivered too late. It is thus required to predict the next source state from the last source states of position and velocity (Fig. 16.12).

The listener is observed in the current state, which corresponds exactly to one frame in the past. The source trajectory, however, is observed from several frames in the past because the sound radiation from the source occurred the amount of the propagation delay in the past. The larger the distance between source and receiver, r_{SR} , the more past states must be analysed.

Assuming the trajectories are sampled with a motion tracker system with sampling rate f_T . The source trajectory is then required to be sampled back for all positions for which $\tau > r_{SR}/c$ holds. For shorter distances the sound has already arrived at the listener.

With sampling rates of motion tracking systems (typically 60 Hz), the critical distance is

$$r_{SR} > \frac{c}{f_T} \approx \frac{340 \frac{\text{m}}{\text{s}}}{60 \text{ Hz}} = 6.67 \text{ m} \quad (16.9)$$

With such a condition, no prediction is required as the propagation delay provides an automatic retardation between radiation time and reception time. This minimum distance can be reduced by increasing the tracking sampling rate.

16.8.2 Real-Time Data Structures

The auralization of moving sound sources and the listener requires the observation of the objects involved at different points in time. It is not sufficient to describe the virtual scene only in the current state. The exact calculation of the propagation time under consideration of the sound field characteristics to their respective radiation and

propagation times require storage and re-use to data from the past states. To make this possible, the solution can be a tree-based data structure for highly dynamic acoustic virtual scenes, which is specifically tailored to real-time conditions. It describes the properties of a virtual scene, including all its parameters over time and not just the current state.

The bases of the scene description are states of a single object or multiple composite objects at a time. Basic conditions such as scalars, vectors and matrices can be assembled into high-order structures (hierarchy). Such compositions are realized by references, whereby unnecessary data copying is avoided and the compact representation remains in their memory requirements. A major problem in real-time systems is the concurrent access to the data. This usually requires the explicit synchronization of reciprocal access. Real-time audio processing with low latency is particularly sensitive to delays of programme execution. Classical synchronization concepts can block the programme flow and thus lead to the problem of priority inversion or acausalities. Specifically, this means that a simple change in the scene, where appropriate, may block the audio processing. To prevent this problem from scratch, an appropriate data structure defines a specific life cycle for all states. After their creation, the conditions can be modified as desired. But once the processing is complete, they are fixed and are then set to read-only. The creation and modification of conditions takes place exclusively in the context (programme sequence or thread) of the user (e.g. he or she sets positions of objects). An efficient procedure of allocation/de-allocation and re-states is achieved by so-called pools. These manage a free list of objects and regulate their use and reuse (Wefers and Vorländer 2015).

16.9 Final Remarks

We close the discussion on real-time processing and real-time rendering at this point. In dynamic systems, the rendering process is integrated into the sound reproduction system. Both components can be treated separately if some basic introduction is required, but the implementation and complete system performance can only be evaluated when rendering and reproduction are discussed as a unit. In the next chapter, sound reproduction systems are therefore introduced and examples for complete VR system architectures are presented.

Chapter 17

3D Sound Reproduction



17.1 Introduction

Computer simulations of acoustic scenes are an important prerequisite for rendering. Technology for 3D sound reproduction, the so-called audio front-end or the acoustic human-machine interface, is an essential component of VR systems, which must be capable of fulfilling high-quality standards concerning the psychoacoustically relevant cues (Fig. 17.1). These cues may differ from one VR application to the next. Some applications require an exact localization, while for others monaural spectral features like reproduction with exact loudness and timbre are more important. In this chapter we focus on electroacoustic technology for surround sound reproduction (Fig. 17.2). It concerns headphone and loudspeaker technology and audio formats which can be used as data interface between systems.

In visual analogy, modern shutter glasses based on polarization filters or green-red filters in connection with high-definition video displays provide excellent stereoscopic reproduction results. A 3D audio reproduction system for VR applications, however, should not be confused with surround sound systems in consumer electronics. The main difference is that VR applications are related to physics-based models and a high degree of realism in the components of sound and vibration generation, transmission and reproduction. A recording engineer for classical music has a different goal. Even in the case of live recordings, he or she will use recording techniques and strategies of microphone placement and mixing to obtain the best result for home environments, replayed by stereo or 5.1 equipment. The balance of instruments or instrument groups is thus manipulated with the purpose of aesthetic effects.

First of all, for the purpose of acoustic virtual reality, we have a strict and exclusive aim: that is, to achieve the best possible authenticity related to three-dimensional perception. A “neutral” reproduction related to linear distortions is required anyway.



Fig. 17.1 Concert hall model in a CAVE-like environment. (CAVE® Automatic Virtual Environment)

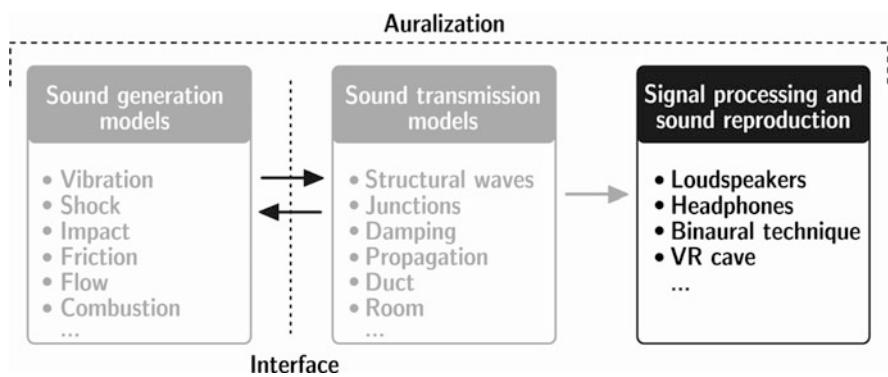


Fig. 17.2 The component of sound reproduction

17.2 Headphone Systems

Headphones or other audio systems integrated in head-mounted displays are well qualified to serve as reproduction transducers, and, thus, they are widely used (Fig. 17.3). Unfortunately, some disadvantages must be discussed which are caused by physical effects in the sound field between the active element of the headphone and the ear canal of the listener. Wearing discomfort and unnatural ear occlusion are

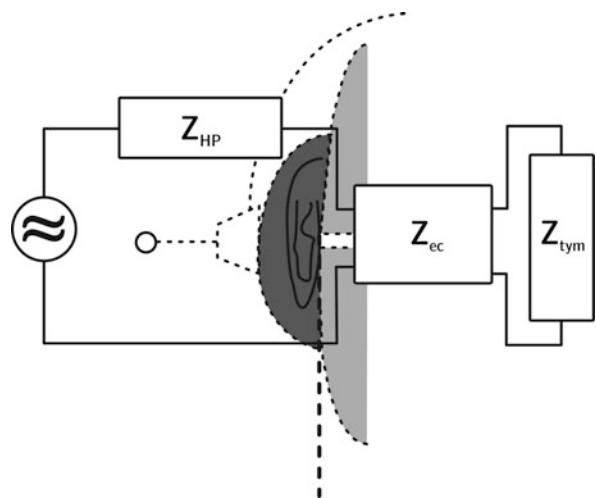


Fig. 17.3 Headphones of various types, open, closed, semi-open

additional factors affecting the quality of the hearing sensation. The so-called in-head localization is one example for such unwanted effects. Externalization of sound sources is one of the main issues that are addressed when discussing headphones. In case of insufficient externalization, the immersion in VR system is drastically reduced. With a proper equalization and special attention to high-frequency radiation into the ear canal, this problem can be partly solved. Adaptive filtering (head tracking) to take head movements into account is also a very important tool when it comes to creating realistic localization and externalization.

Headphone equalization is by far more difficult than loudspeaker equalization. The radiation impedance acting on the transducer cannot be approximated by using elementary field conditions like “piston in free half space”. Instead, the radiation impedance into the ear canal is relevant, which brings us to the first difficulty. Properties of ear canals of listeners differ tremendously among a population of test subjects. When it comes to the input impedance, resonances which can be related to individual physiological features are known only in principle. They can be modelled rather easily, but the model parameters depend on the individual anatomy.

Fig. 17.4 Model of headphone mounted at the ear



Artificial ears were developed as a kind of “average ear”, but their applications are restricted to special headphone types.¹ Even under ideal measurement conditions for digital equalization and calibration, there remains the uncertainty introduced by mounting at the real ear. Particularly for closed headphones, uncertainties are observed that are caused by leakages. Interindividual differences are also important in this context.

In Fig. 17.4 the headphone source impedance, Z_{HP} , is coupled with the ear canal impedance, Z_{ec} , and finally the termination impedance of the eardrum. Headphones connected to ear canal and eardrum impedances have been subject to basic research for many years (Hammershøi 1995). The eardrum sound pressure represents in a purely technical sense the complete excitation signal of the auditory system. For this assumption, other paths of sound transmission to the cochlea, like bone conduction, are neglected. With a good model of the ear canal impedance and definition of the sound pressure at ear canal entrance as driving signal, the eardrum pressure is given unambiguously. The problem, however, is that many factors of uncertainties, as described above, have to be taken into account. Occlusion by a headphone affects the standing wave pattern as well and thus the transfer impedance of the ear canal. Only in specific test scenario, for instance, a blocked ear canal, the driving pressure is measured with the clear and accurate reference (Fig. 17.5).

¹Typically audiometric headphones

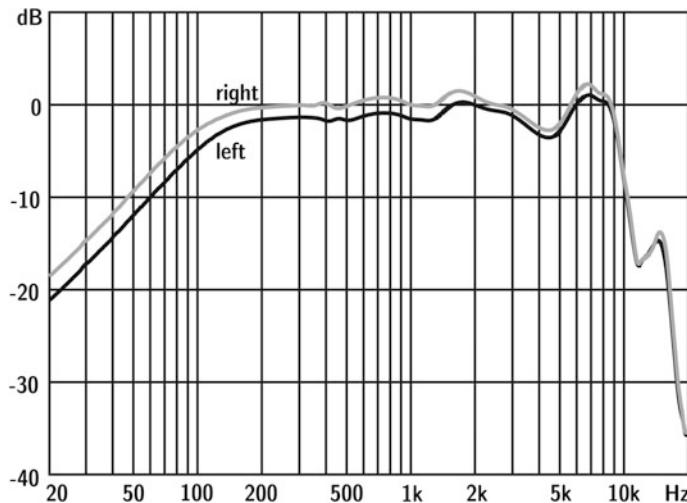


Fig. 17.5 Example of an in-ear headphone frequency response measured with an ear simulator (IEC 60711)

17.2.1 Head-Mounted Displays

Head-mounted displays (HMDs) are equipped with small LCD or OLED² (or other) displays which create stereoscopic images for both eyes. The resolution is rather small compared with full screen CAVE-like solutions, and the main disadvantage is the limited view angle, so that one's own hands and fingers must be rendered. The acoustic path is usually covered by integrated headphones. Due to the uncomfortable wear of the device, the immersion is affected. The fact that components for haptic or tactile stimuli which do not interfere with vision and sound can be easily added is, however, an advantage. Concerning the sound reproduction, all aspects of headphone technology apply. It will be very interesting to see how HMD technology evolves after having entered the consumer market.

Headphone Equalization for Binaural Signals

Headphones used for binaural reproduction should reproduce the binaural signals implemented by binaural synthesis or binaural (dummy head) recording without affecting the binaural cues (Fig. 17.6).

The correction (equalization) filters serve two purposes. First of all, they must provide that the sound pressure at the eardrum is identical in the recording and replay situation. If the sound pressure at the eardrum was perfectly recorded, the sound path

²Liquid crystal device, organic light-emitting diode

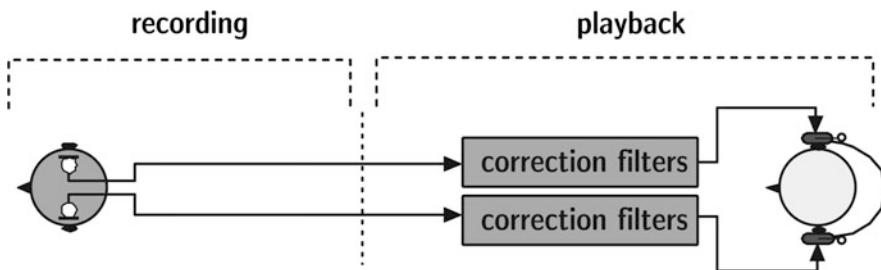


Fig. 17.6 Arrangement for the reproduction of binaural signals. (After Blauert 1996)

between headphone and eardrum was included twice in the replay situation. Therefore, a correction filter must be used to extract the ear canal (see Fig. 17.4). Other filters are also used, but they are used to ensure the compatibility between headphone and loudspeaker reproduction of binaural signals.

A plane wave irradiating a human head at frontal incidence is usually the reference. Perfectly equalized loudspeakers create a quasi-plane wave (p and v in phase). When this wave hits a head, however, the eardrum sound pressure is distorted. In fact, the HRTF for frontal incidence is multiplied with the loudspeaker (plane wave) flat spectrum. Binaural recording, replayed by loudspeakers, carries HRTF twice, once from the dummy head in the recording situation and once from the listener's head. Correction filters for compatibility refer to plane wave at frontal incidence, diffuse field incidence or independence of direction (Genuit 1984). When headphones are used, these filters are obsolete for binaural signals.

However, as the compatibility with loudspeakers is required, commercial headphones are equalized with respect to reference sound fields. Free-field headphones deliver by definition the same hearing impression as a sound event incident in a plane wave at frontal incidence and diffuse field headphones accordingly for random incidence of incoherent sound waves. A reproduction of a recorded diffuse field³ with a diffuse field headphone yields the same hearing sensation as in the real (recorded) situation (Fig. 17.7).

Individual Filters

All definitions described above hold true for standard HRTF and accordingly for average listeners. It was shown by several authors that localization errors can be significantly reduced by using individual filters. The main reason is in the individual features in the HRTF, particularly above about 6 kHz.

When individual HRTF is obtained by measurement at the listener using probe microphones, all filters can also be obtained by measurement data processing and feeding into digital filters. The reference situation in this case may be the quotient of

³For instance, created in a reverberant room far outside the reverberation distance

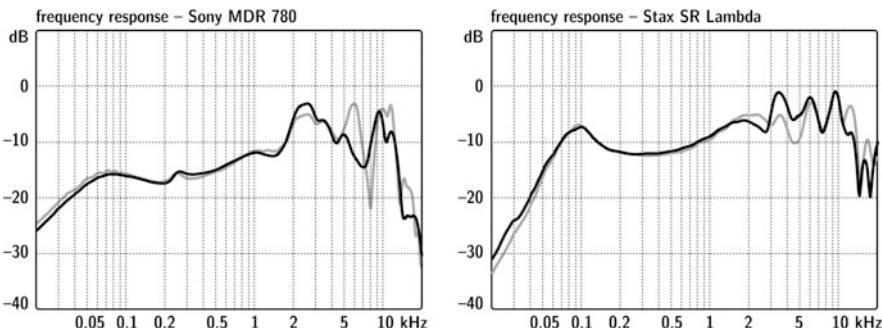


Fig. 17.7 Examples of typical free-field sensitivities of headphones

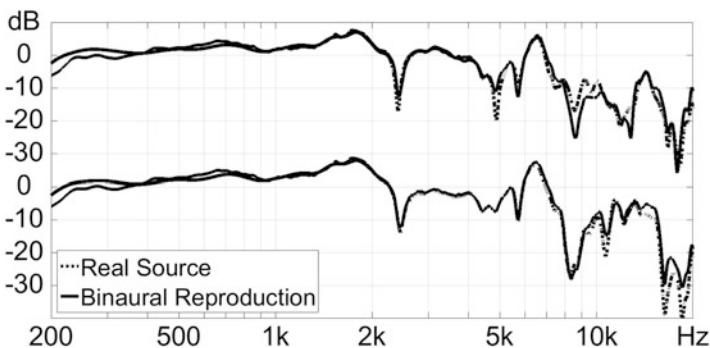


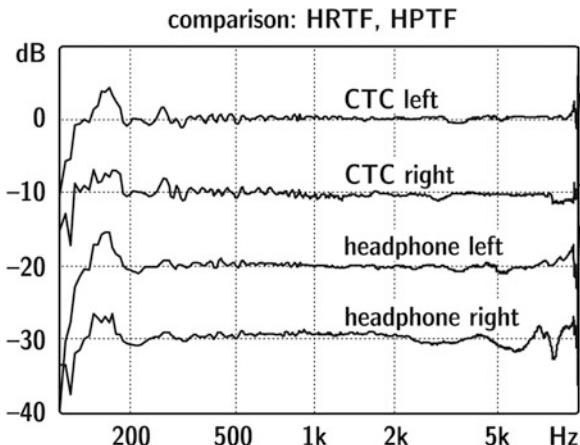
Fig. 17.8 Four equalized binaural transfer functions, two for a loudspeaker and two for headphones (HPTF: headphone transfer function) into the ear canal (individual test subject). The two curves corresponding to left (top) and right ear (bottom) are perceptually identical

the electric voltage fed to a loudspeaker in certain direction and a large distance (>2 m) and the sound pressure measured in the ear canal of a test subject.

This electroacoustic transfer function can be interpreted as individual HRTF.⁴ The same transfer function should be present when the headphone is applied. Figure 17.8 shows a comparison of this experiment. In this figure four curves are shown, two for the right and two for the left ear. They are identical with respect to the resolution of the plot. The deviations are smaller than 0.5 dB. This means that no audible effects should be noticeable in a listening comparison. In fact, in most cases no significant differences were found. And those small differences found are more related to the test conditions of placing and replacing the headphone than to technical aspects. Furthermore, the psychological component in such tests which aim at very small perceptual effects is to be considered. After all it is found that with exact equalization of input signals into the hearing system, the auditory sensation is exactly the same (Blauert 1996).

⁴With exact calibration of the loudspeaker, it is actually the HRTF.

Fig. 17.9 Individually equalized transfer functions of loudspeakers (top, HRTF CTC) and headphones (HPTF, bottom) from the free field to the reference point in the ear canal, divided by the individual transfer function between an electric voltage fed into a loudspeaker in free field and the ear canal sound pressure. Curves shifted in steps of 10 dB for better visibility



After an individual equalization, the remaining differences between loudspeaker or headphone binaural reproduction systems are below the just noticeable differences; see Figs. 17.9 and 17.10. More detailed results were published by Møller (1992) including an analysis of the accurateness of binaural cues and the final performance of headphone reproduction in listening tests. All in all, one can state that headphones can be used for 3D sound reproduction and that they yield good results, but their equalization must be carried out carefully.

Individualization of Custom HRTF

HRTF datasets can be adapted to better match individual binaural cues. Several methods exist, based on anthropometric data, subject selection and interpolation after feature extraction. For feature extraction and data compression, principal component analysis (PCA) or neural network models are commonly in use (Xie 2013). The procedure follows the steps of reduction of complexity of the centred HRTF data, decomposition into orthonormal basis functions by solving a matrix/eigenvector problem, determination of the scores and reconstruction of the data by using the weighted sum on the based functions. Reconstruction of the phase is usually realized by a minimum phase plus all-pass filters. This way, HRTFs for individuals with their own anthropometric data can be created as well as HRTF for directions of sound incidence which were not included in the originally measured HRTF dataset (Xu et al. 2008; Bomhardt et al. 2016).



Fig. 17.10 Probe microphone in the ear canal of a test subject

17.3 Loudspeaker Systems

Spatial sound fields can be created with loudspeakers by using one of two general concepts. One can try to reproduce head-related signals, taking advantage of the fact that the hearing sensation only depends on the two input signals to the eardrums. Loudspeakers arranged around a listening point (“sweet spot”) may also serve for a spatially distributed incident sound field. Furthermore, one can try to create a complete wave field incident on the listening area. The potential to involve more than one listener in the second approach illustrates the conceptual difference between the two methods (Fig. 17.11).

The binaural technology is described in Sect. 17.3.4 (see below). Sound field technology can be defined in various ways. Another basic form of sound source imaging in the horizontal plane is the well-known stereo setup or a surround sound system.⁵ These approaches make use of the psychoacoustic effect of phantom sources. The basis is a multichannel microphone separating the incident field at one listener point (sweet spot) into spatial components.

More accurate for sound wave field reproduction are the methods of higher-order ambisonics, HOA (see 17.3.2), and wave field synthesis, WFS (see 17.3.3). WFS is

⁵Mostly in use are “Dolby Surround[®],” or successor systems such as Dolby Atmos[®].

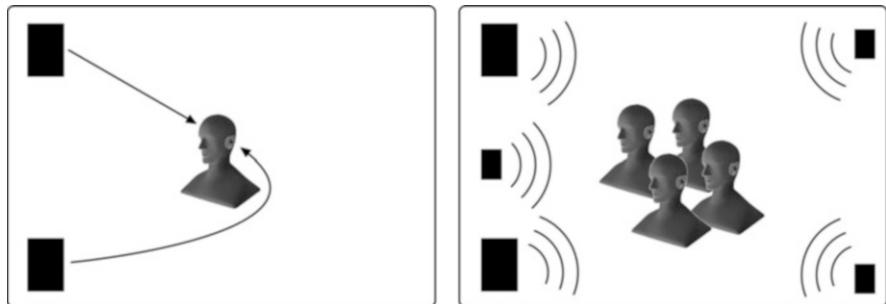


Fig. 17.11 Binaural technology (left) and sound field technology (right)

used to create an approximation to the spatial sound field incident by using a microphone array too, but not at a listener point. Instead, the microphone arrangement is larger, and it is located on elementary geometric figures like straight lines or circles around the listening area. The artificial sound in a larger area can thus be created, and more than one listener can be provided with spatial sound.

17.3.1 VBAP Surround Sound

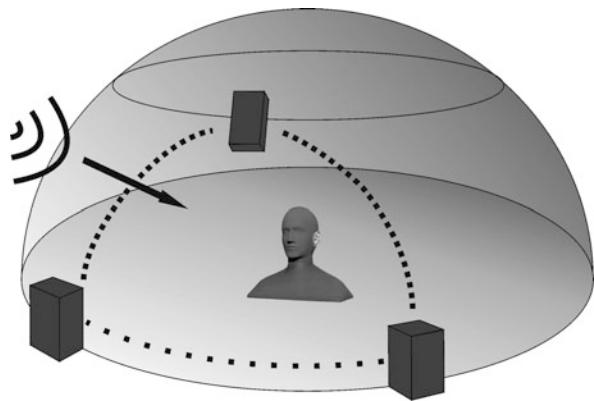
Starting from a home stereo setup, where the balance control between right and left channel is used for image shift additional pairs of loudspeakers can be placed around the sweet spot. The stereo balance controls the relative amplitude between right and left. The same concept is used for other directions.

This perceptual effect stems from the interaural level differences competing with interaural time differences. The overall perception is created from the influences of both cues, which can well be proven by using models of binaural signal processing (Braasch 2005).

This technique was in the early 1970s applied in home entertainment applications referred to as “quadraphony” (with no commercial success however). Nowadays, some virtual reality systems include surround sound systems called vector-base amplitude panning (VBAP). The pioneer in VBAP technology is Pulkki (1997). VBAP is used to position virtual sources in a 3D surround loudspeaker setup. By panning, the same sound signal is applied to a number of loudspeakers with appropriate non-zero amplitudes, typically to one, two or three loudspeakers simultaneously.

The directional coordinates used are the angle between the position vector to a virtual sound source and the median plane and the angle of a projection of this position vector in the frontal direction. Thus, the basic design rules for creating virtual sources are given by elementary matrix formulation of the loudspeaker gains by using the vector base (Pulkki 1997). For the encoding part which is introduced to

Fig. 17.12 3D reproduction of a virtual sound source with VBAP. (After Pulkki 1997)



create the virtual sound sources or the spatial impulse response of the environment, it is referred to Sect. 11.9.1.

For ambitious integration of acoustics into VR, however, this method in its basic form has shortcomings such as inherent implementation of binaural processing mechanisms (phantom sources, see Fig. 17.12). Therefore, low-order VBAP is well suited as a 3D audio format for perceptually plausible spatial scenes but not as platform for unbiased studies of the human hearing and neural system as such. At higher order of channels, the gain matrix values of VBAP converge to those of a sound field reconstruction method: higher-order ambisonics (HOA; see next section).

17.3.2 Ambisonics

Based on research by Gerzon (1976), the so-called B-format or ambisonics technology was developed. The B-format is a four-channel recording standard which uses a sound field microphone. The four channels carry the signals denoted X, Y, Z, and W; see below (Fig. 17.13).

Usually the channels represent the front-back (X), up-down (Z), left-right (Y) and mono (W) signals. X, Y and Z signals stem from figure-of-eight microphones in each specific orientation. The W channel is fed from an omnidirectional microphone. The four channels represent a decomposition into spherical harmonics (see 2.6) including the orders 0 of the monopole component and 1 for three dipole components.

Thus it is possible to achieve an isotropic sound incidence. In a replay situation, a certain geometric approximation of spherical shape by polyhedra or similar arrangements is required. Mixing strategies of the multichannel recording can be adapted to match the recording format to the replay situation.⁶ In terms of spatial sound field

⁶Usually performed by decoding matrices

Fig. 17.13 Sound field microphone for recordings in B-format. (Courtesy of Lab. of Acoustical Imaging and Sound Control, TU Delft)



processing, the decoding process is nothing but a reconstruction by linear combination of zero (W) and first-order (X, Y, Z) spherical harmonics.

Higher-Order Ambisonics, HOA

In rendering of spatial scenes by computer simulation, encoding rules are closely related to the coordinate system of the receiver, where directional factors are applied according to spherical harmonic coefficients of the orders involved (Fig. 17.14).

In general, the encoded signals used are the spatial representations of the sound field in terms of a spherical harmonic decomposition of any order; see 11.9.1. Thus, they are independent of surround sound system in the reproduction setup. The higher the order, the more directional details and higher spatial resolution are achieved. Using higher orders is referred to as “HOA” (Daniel 2000). For complete reconstruction it is required to use a minimum number of L loudspeakers for N ambisonic channels.

$$L \geq N \quad (17.1)$$

The encoding part was introduced in Sect. 11.9.1. For decoding, i.e. sound field reproduction, the control equations for the loudspeaker array are usually expressed



Fig. 17.14 Custom-made 121-channel spherical microphone array (38 cm diameter, spherical harmonics order 10. (Berzborn and Vorländer 2019)

in matrices. y is the column vector of HOA channel information involved. The spherical harmonic coefficients, y , refer to the weight of the corresponding SH coefficient representing the direction of a virtual sound wave in a receiver-centred coordinate system.

$$\mathbf{Y} = [y_0^0 \ y_1^{-1} \ y_1^0 \ y_1^1 \ y_2^{-2} \ y_2^{-1} \ y_2^0 \ y_2^1 \ y_2^2 \dots]^T \quad (17.2)$$

The condition for reconstruction of the correct total sound field from the column vector of sound pressures, p , radiated from the loudspeakers involves the gain matrix, G :

$$\mathbf{Y} = G \cdot p \quad (17.3)$$

G contains the values of the spherical harmonics for the loudspeaker positions of the setup used for replay, with N rows for the chosen truncation order of the HOA signal and L columns for the loudspeakers. The required loudspeaker signals are accordingly

$$p = G^{-1} \cdot \mathbf{Y}. \quad (17.4)$$

G^{-1} is the inverse of the matrix G , which is usually non-quadratic, since usually $L > N$. But the pseudoinverse of G will provide useful values as long as the condition number of G is small. For this, it is advantageous to make the loudspeaker layout as regular as possible (tetrahedron, dodecahedron, etc., see Fig. 17.15). Further effort

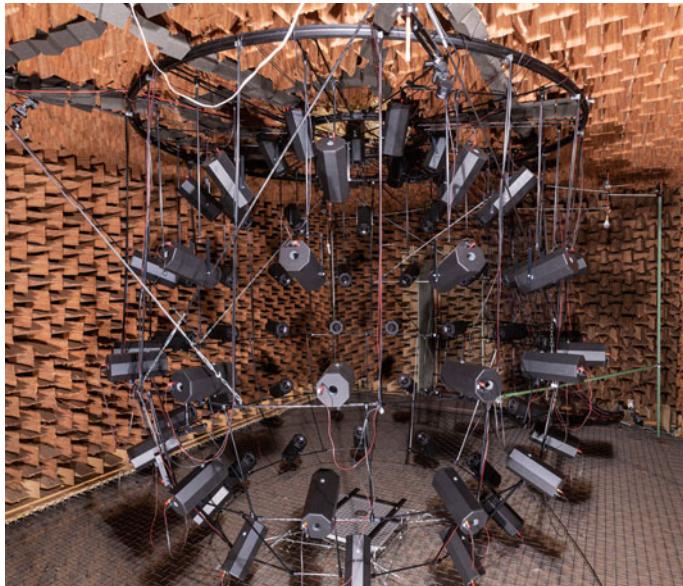


Fig. 17.15 68-channel spherical loudspeaker array for seventh-order HOA. (Institute of Technical Acoustics, RWTH Aachen University)

can be taken to optimize and modify the gain conditions in order to compensate for near-field effects and to get approximations for rendering of virtual sound sources inside the loudspeaker array (Daniel and Moreau 2004).

Apart from the strict physical approach of spherical harmonics, decoding schemes are in use, which benefit from localization and distance perception of human hearing. It must be kept in mind again, however, that virtual sound fields may be used to study the hearing system. A generally applicable reproduction system must not introduce any artificial auditory cue which is not part of the simulated sound.

Transition from HOA into Binaural Formats

Obviously, the difference between spherical array encoding and binaural encoding is the specific spectral filter which leads to the impulse response representation in the signal channels. For simplicity, this is compared with a generalized reflection (directional wavefront, “virtual source”) H_{refl} arriving under the angles ϑ and φ in a defined receiver-oriented coordinate system. In HOA, encoding this reads (see 11.9.1)

$$B_n^m = 4\pi \sqrt{-1}^n Y_n^m(\vartheta, \varphi) H_{\text{refl}}(\vartheta, \varphi) \quad (17.5)$$

$$H_{\text{refl}}(\vartheta, \phi) = \frac{e^{-j\omega t_j}}{ct_j} \cdot \underline{H}_{\text{source}}(\theta, \phi) \cdot \underline{H}_{\text{air}} \cdot \prod_{i=1}^{n_j} \underline{R}_i \quad (17.6)$$

The binaural formulation is

$$H(\vartheta, \phi)|_{\text{left,right}} = H_{\text{refl}}(\vartheta, \phi) \cdot HRTF(\vartheta, \phi)|_{\text{left,right}} \quad (17.7)$$

The transition from HOA formats to binaural signals is, thus, straightforward if the link is made at the spatial encoding of the binaural signals at the two ears (Duraiswami et al. 2005). For this we assume that the signal radiated from the loudspeaker array in the decoding process is perceived by a virtual listener in the centre of the array. Actually the loudspeakers are connected with the listener over the HRTF corresponding to the actual positions of the loudspeakers in the array. Their number is L . This yields

$$S(\omega)|_{\text{left,right}} \approx \sum_{l=1}^L HRTF|_{l,\text{left,right}}(\omega) \cdot S_{l,\text{loudspeaker}}(\omega) \quad (17.8)$$

With HRTF representations encoded in spherical harmonics, the left and right channels at the two ears are given by⁷

$$HRTF(\vartheta, \phi, \omega)|_{\text{left,right}} = \sum_{n=0}^{\infty} \sum_{m=-n}^n H_n^m(\omega) Y_n^m(\vartheta, \phi) \quad (17.9)$$

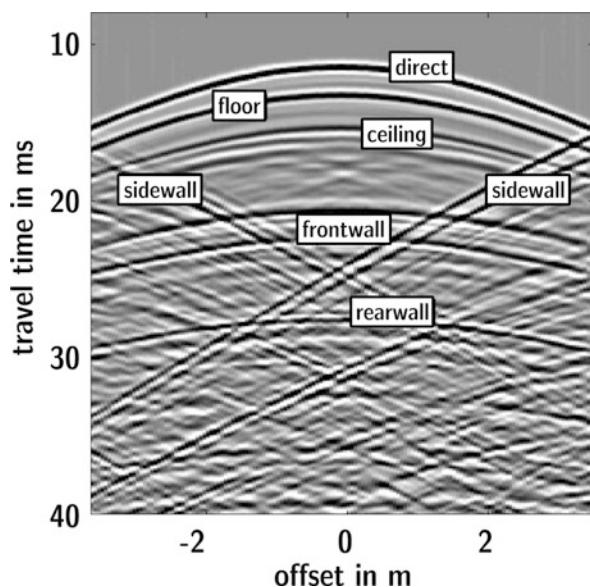
Now, the binaural format can directly be calculated from the HOA format by multiplying in the spherical harmonic domain the 3D impulse response data (H_{refl}) with the HRTF data. As this is done for the left and the right ear, the resulting signal has two channels.

$$S(\omega)|_{\text{left,right}} = \sum_{n=0}^N \sum_{m=-n}^n B_n^m(\omega) H_n^m(\omega) \quad (17.10)$$

Accordingly, a specific HRTF must be defined in order to create the H_n^m coefficients for the transition of HOA to binaural.

⁷In practice, the first sum is approximated by truncation.

Fig. 17.16 Wave field analysis. (Courtesy of Lab. of Acoustical Imaging and Sound Control, TU Delft)



17.3.3 Wave Field Synthesis

The wave field synthesis technology, WFS, is a rigorous approach to wave field reconstruction (Berkhout 1988). Its concept stems from seismic physics and source localization by array techniques. The basic mathematical algorithms are kind of spatial Fourier transformations between the space and the wave number domain. Complex wave fields are decomposed into elementary waves like plane, spherical or cylindrical waves by an appropriate transformation (Fig. 17.16).

In plots of the position along a line versus propagation time, these wavefronts can be studied, either in static condition or in an animated movie in motion, similar to results from FTD simulation.

The wave decomposition is achieved by analysing the signals in microphone arrays. According to the Huygens principle, the points where the sound pressures were recorded at the microphone positions can be interpreted as elementary sources. In replay situation, the wave field is reconstructed by sending waves from these points. This is the step from wave field analysis to synthesis (Fig. 17.17).

As long as the discrete spatial sampling is sufficiently high, any wave field can be reconstructed. Unfortunately, this prerequisite creates apparently severe practical problems. A physically perfect sound field with frequency content up to 10 kHz within a space of $3 \times 3 \times 3 = 30 \text{ m}^3$ requires approximately 50,000 loudspeakers of 3 cm size. However, as discussed in several sections of this book, the physical constraints are not as relevant as the psychoacoustic constraints. The existence of broadband signals and masking in higher-frequency bands allows more relaxed conditions concerning the loudspeaker size and distance. Moreover, the spatial resolution of human hearing is more sensitive in the horizontal plane. Hence, a

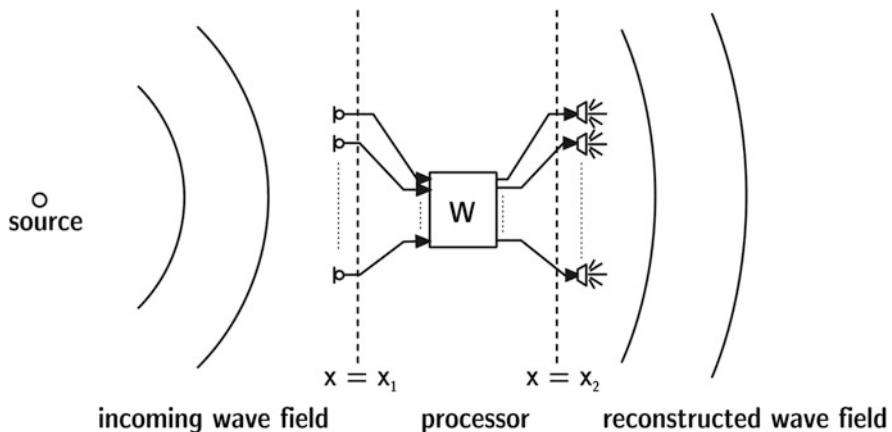


Fig. 17.17 Field reconstruction by loudspeaker arrays. (After Berkhout 1988)

two-dimensional array serves well for the most crucial spatial information. With these simplifications being made, practically installed WFS systems with around 500 loudspeakers are in use. This means that only 1/100 of the theoretically required equipment must be used.

The WFS is based on the Helmholtz-Kirchhoff integral (see also Eq. (10.5)) which is derived from a general form of the Helmholtz-Huygens integral by excluding the source term.

$$p(r) = \iint \left(g(r|r_0) \frac{\partial p(r_0)}{\partial n} - p(r_0) \frac{\partial g(r|r_0)}{\partial n} \right) dS_0 \quad (17.11)$$

r denotes the position of the receiver, and the set of r_0 denotes the source positions in the loudspeaker array.⁸ The integral covers a line or area of sources with the amplitudes $p(r_0)$. The resulting sound pressure at the receiver is thus $p(r)$. The kernel of the integral includes Green's functions of monopole and dipole (pressure gradient) sources. In general, any wave field can be decomposed into the elementary waves from monopoles and dipoles. One effect of the Helmholtz-Kirchhoff formulation is that sources within the surrounded volume cannot be modelled. Solutions for near sources, however, are being developed, but these approaches require head tracking, and this is valid only for one person. For sources from an outside array, however, any arrangement is theoretically possible, and groups of listeners can be fed with 3D sound (see Fig. 17.18).

Cylindrical waves are the functional basis for two-dimensional arrays and reproduction of the sound field in a plane instead of a volume. The corresponding Green's

⁸The source notation by a running index is omitted here.



Fig. 17.18 Installation of WFS. (Courtesy of the Fraunhofer Institute of Digital Media Technology (IDMT), Ilmenau, Germany)

functions do not represent the “correct” distance law in 3D environments. Amplitude correction must therefore be applied.

The process of sound recording and mixing is different from usual techniques applied in audio engineering and music production. The spatial information must be integrated in a flexible way so that position, orientation or movement of the listener is not restricted. The spatial decoding is accordingly transferred to the WFS reproduction system, while each source⁹ as such is recorded separately from other sources in best possible anechoic situation. Furthermore some control parameters like position and relative level of the source are encoded.

The performance environment, for instance, a room, is not included in the recording, but created at the end by the WFS by mapping the parameters and signals into an actual, synthetic or augmented spatial situation (Fig. 17.19).

For a room acoustical simulation, the room impulse responses related to all sources must be calculated for all reference signals (microphone array). The spatial information derived from these calculations can be transformed into the wave-functional basis. Since the impulse responses contain direct sound(s) and reflections, the complete spatial field is reproduced in the complete listening area. The set of room impulse responses can be obtained by any of the methods described in Sect. 11.7, except binaural filters (Fig. 17.20).

⁹Typically a musical instrument or a voice.

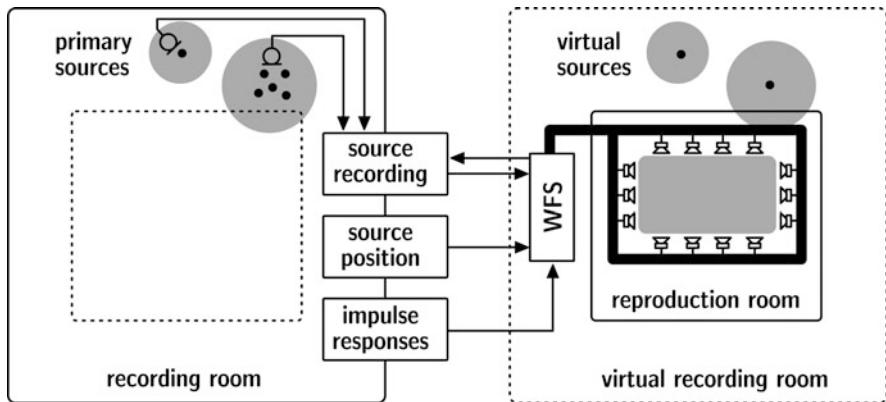


Fig. 17.19 Recording formats of WFS. (After Spors et al. 2004)



Fig. 17.20 CAVE-like environment with WFS. (Courtesy of the Fraunhofer Institute of Digital Media Technology (IDMT) and Technical University Ilmenau, Engineering Design Group)

17.3.4 Binaural Loudspeaker Technology

We shall now go from the approximate creation of wave fields in large audience areas to the one-listener solution. The concept is part of binaural technology. When we use loudspeakers, the binaural loudspeaker arrangement should act as a “virtual headphone”. This condition is already described by the expression “binaural”. Both ears must be fed with binaural signals, as created by binaural synthesis or dummy

Fig. 17.21 The loudspeaker signals Y_{LL} and Y_{RR} should represent the ear signals Z_{LL} and Z_{RR} . Unwanted crosstalk happens over the paths H_{LR} and H_{RL}

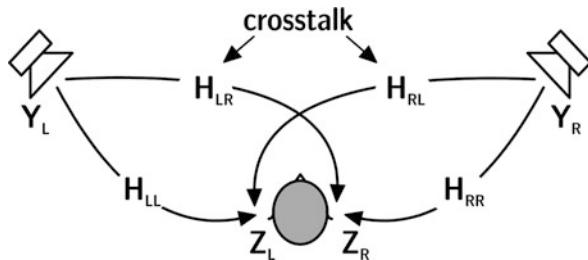
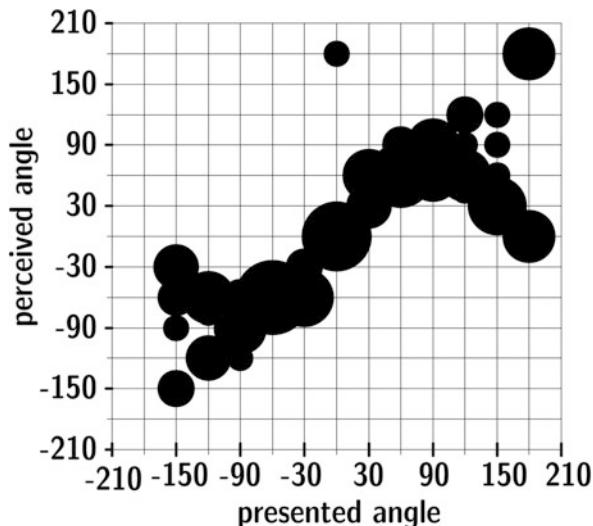


Fig. 17.22 Results of a localization experiment in an anechoic chamber. The diameters of the circles represent the numbers of the perceived direction as response to the synthesized direction of sound incidence



head recordings. Hence a left/right stereo setup should work well. However, it has been well-known for many decades that crosstalk might occur that interferes with the wanted ear-related signals. The virtual headphone and its function are disturbed by the insufficient channel separation (Fig. 17.21).

Without compensation filter, the localization by binaural synthesis is not possible. In an experiment, test subjects were asked to mark the perceived direction of sound incidence. The diameter of the circles represents the number of responses. While expecting all responses to be on the diagonal, it is observed that the uncertainty of the responses is large. Furthermore, front-back confusions occur near the 90° direction indicated by perpendicular groupings of circles (Fig. 17.22).

Crosstalk Cancellation

The required channel separation can be achieved by signal processing. Based on a proposal by Bauer (1963), the first filter structure was formulated by Atal and Schroeder (1963) as a kind of subtraction filter. To discuss the process used for

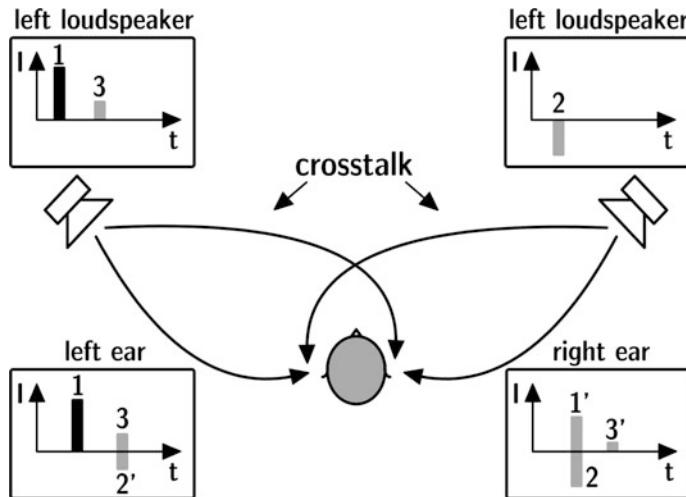


Fig. 17.23 Iterative crosstalk compensation

this filter, we define the so-called ipsilateral ear being oriented towards the direction of sound incidence and the contralateral ear being in the shadow region. Crosstalk occurs at the contralateral ear. For example, if the right ear is the ipsilateral ear for the right loudspeaker, then the left ear is the contralateral ear. The crosstalk signal component from the right loudspeaker into the left ear can be compensated by radiating a respective signal from the left loudspeaker and vice versa.

The compensation signal, however, also creates crosstalk. We must pursue the same strategy for some iterations. These iterations are stable because over the iterations the contralateral signals have less and less energy. Five iteration steps are usually sufficient to achieve a channel separation of 20–25 dB. Simple solutions (low-order iteration), by the way, are implemented in portable radios to create a “wide function”. They suffer, however, from loss in bass performance. The weak bass is in fact an indicator of the quality of the iterations (Fig. 17.23).

Møller (1992) proposed a closed solution for the iterative process. This is possible due to the inherent mathematical structure of a geometric series; see also Schmitz (1994). The validity of the approach and the stability of the filter, however, depend on the level difference between ipsi- and contralateral transfer paths. When the denominators become smaller, the filter produces ringing effects due to large narrowband peaks in the transfer function. The angular zone which provides a natural damping of the contralateral paths is thus limited and depends on the geometric situation between the loudspeaker pair and the listener.

In notation according to Fig. 17.21 yields

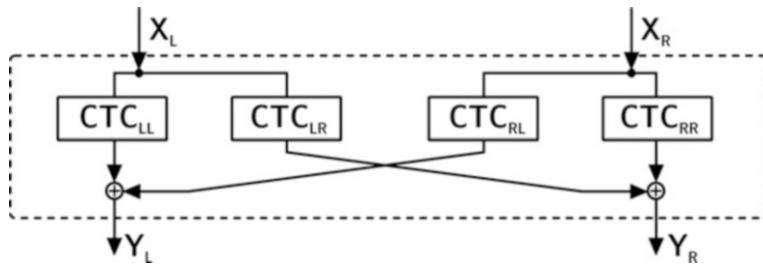


Fig. 17.24 Crosstalk filter design

$$Y_L = \frac{1}{L} \left[\frac{H_{RR}}{H_{LL} \cdot H_{RR} - H_{LR} \cdot H_{RL}} \cdot X_L - \frac{H_{RL}}{H_{LL} \cdot H_{RR} - H_{LR} \cdot H_{RL}} \cdot X_R \right] \quad (17.12)$$

and Y_R accordingly.

The ear signals X_L and X_R are constructed from four filters as represented in the quotients in Eq. (17.12) and the corresponding equation for the right channel (Fig. 17.24).

Stereo Dipole

This technique is similar to the crosstalk cancellation system. The loudspeaker arrangement, however, is a left/right dipole source placed in the front of the listener with at most 10° angle span. An enhanced crosstalk damping is achieved by adjusting the null axis of the dipole into the front plane. Both ears receive signals from one lobe of the dipole only. The result of the stereo dipole source is similar to theoretical monopole and dipole combination.

The stereo dipole system works very well in symmetric geometry between loudspeaker and listener, and it shows best performance in close distances, like in the situation of PC displays. More information can be found in Kirkeby et al. (1998).

Dynamic Crosstalk Cancellation

The compensation method of crosstalk cancellation and of the stereo dipole is based on interferences and shading effects. They are thus sensitive to phase errors caused by geometric variations, particularly in translational or rotational right/left movement. If the compensation filters are calibrated for one specific position (sweet spot), off-side positions in the order of magnitude of 2 cm may lead to audible effects of distorted timbre and localization (Fig. 17.25).

The possible range for these movements is larger for the stereo dipole. All static techniques fail, however, if the listener interacts with the environment or moves freely in the order of magnitude of metres.

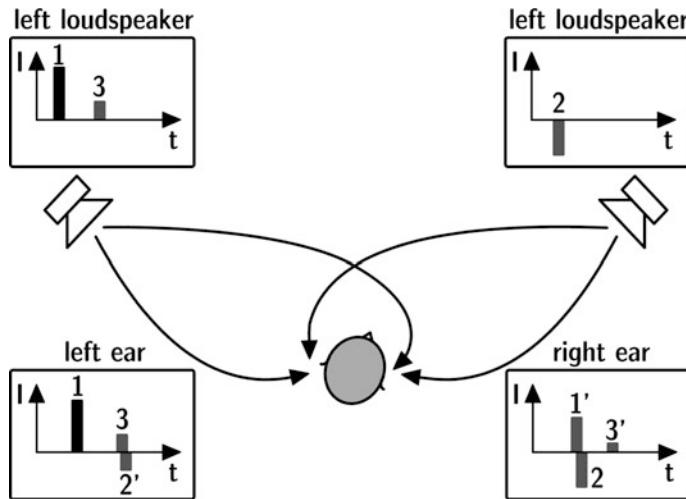


Fig. 17.25 Mismatch by head movement

Crosstalk compensation needs not to be installed in a symmetric arrangement of loudspeakers and listener. Hence there is no need to fix the sweet spot at one arbitrary location. The sweet spot can also be fixed at the listener's position, even when he or she is moving. This can be achieved with adaptive filtering (Gardner 1997).

First of all, the position of the head must be determined by using a head tracker.

Four-Speaker System

Dynamic crosstalk cancellation requires a valid HRTF filter set for each position. Further research by Gardner (1997) and Lentz and Behler (2004) has shown that a dynamic CTC using two speakers is only stable within the angle spanned by the loudspeakers. To provide a free rotation for the listener, the two-speaker solution is expanded to four channels, two of which are running at the same time.

Between the areas a cross-fade algorithm is implemented using two parallel working CTC systems, which are then partly superimposed, so that the current binaural audio signal is filtered with the correct crosstalk cancelling filter for the listener's present position. More details on the implementation and evaluation are found in Lentz et al. (2007).

One example of an integrated system including room acoustic rendering based on physically consistent models and headphone-free 3D sound reproduction is illustrated in Fig. 17.26. It consists of auralization software with high precision in the early part (accurate in scales in microseconds) of the impulse response and smooth updates of the late decay (accurate in scales of milliseconds) which only take place after the listener moved by some metres (Figs. 17.26 and 17.27).

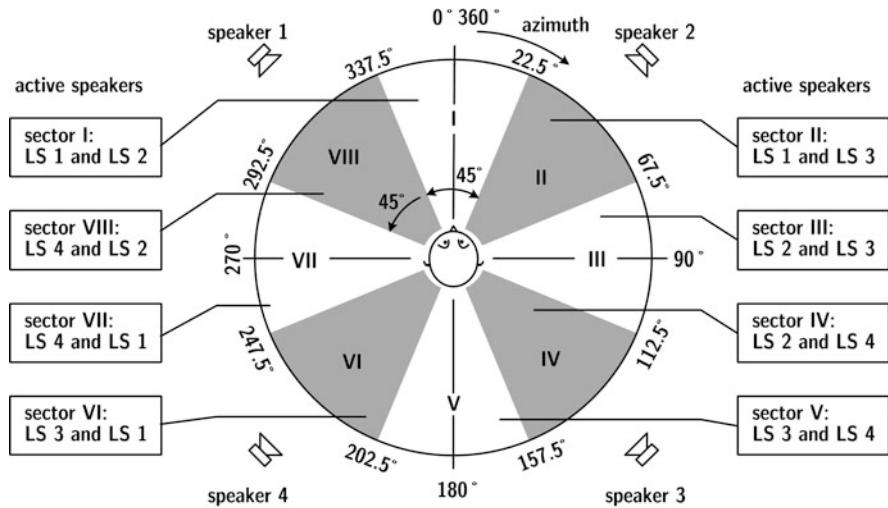


Fig. 17.26 Eliminating the limits in listener rotation: sectors of activation of loudspeaker pairs in a four-speaker dynamic CTC system

17.3.5 Final Remarks: The Human Factor

The signal chain from simulated external sound wave fields to the listeners' eardrum via headphones or spatial audio includes various components which can be associated with linear distortions. The HRTF and the headphone equalization can either be custom or individual. The recording/reproduction system can be controlled by using adaptive filters and head tracking, or it can be based on static filters. There are numerous approaches of HRTF and headphone filters in dynamic and static realization.

Listening tests with all those options show different and hardly reproducible results, when it comes to high-end quality (Brandenburg et al. 2016). The final performance of spatial sound reproduction can be optimized to a certain extent, but the absolutely authentic auditory illusion is not possible at the time being. The reason for this is the listeners' experience and training, which has a significant role. Solutions of filter design which were developed for one listener panel with a certain degree of training are not valid for other listeners. Non-trained listeners may still experience in-head localization or front-back confusion.

Accordingly, all steps of creating a perfect spatial audio technology with HOA solutions or HRTF databases (from custom to individual) are necessary but not sufficient conditions for the listener.

Also for source modelling, natural behaviour of humans is a critical factor when it comes to sensation of authenticity. Dynamic cues in musical sound generation were studied by Ackermann et al. (2019). Dynamic sound sources include the expressive intentions of a performer to the audience. Ackermann's work aimed at assessing the acoustical effect and the perceptual relevance of musicians' movements. In listening

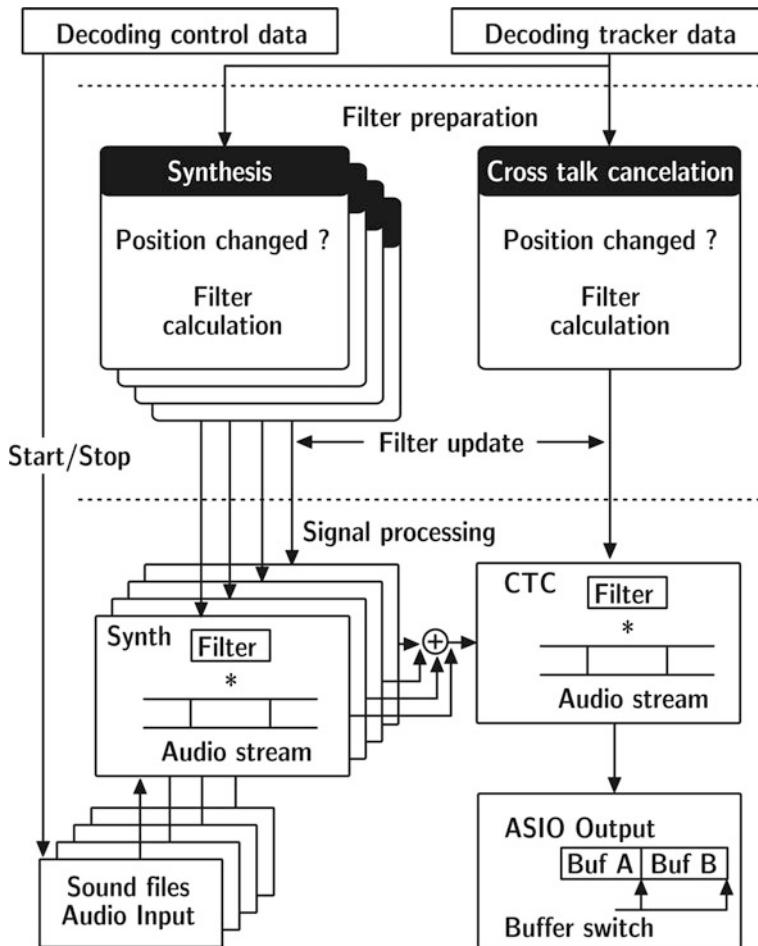


Fig. 17.27 Block diagram of dynamic CTC processing with head-tracker data

tests, they auralized static and dynamic musical performances for comparison. It was shown that the signal-related fluctuations are clearly audible both in various room situations.

Of course, the quality of surround sound technology with loudspeaker systems is influenced by the human factor as well. In this case the signal chain does not include the explicit representation of static or dynamic binaural cues by using HRTF data since the listening situation refers to 3D wave fields impinging on the own head. The HRTF of the actual listener is then included implicitly.

Chapter 18

Acoustic Virtual Reality Systems



18.1 Introduction

Virtual reality (VR) is a computer-generated environment for interaction in real time. One important feature of VR is the multimodality of the human-computer interface. Most VR systems were initially developed for 3D vision. In order to achieve the presence and immersion of the user, VR is not complete without the acoustic and haptic dimensions (and more). The driving forces for establishing VR applications are task-specific interaction scenarios and their acceptance by the user and user feedback. VR technology includes head-mounted displays, screen displays, and full immersive CAVE sytems, and audio reproduction technology via headphones or surround sound systems. Examples are presented and discussed in this chapter, including references to open-source software for acoustics VR and to benchmarks.

User interfaces are well-established for several kinds of applications, in computer operating systems, in specific application software, or in computer games. Any kind of control display for machinery, vehicles, or any other technical object or system can be interpreted as human-machine interface. Acoustic information may replace visual information displayed by system controls or monitoring (Begault 1994).

The purpose of a so-called auditory display is to generate attention and to provide information. It is like an “icon” for the visual sense. Therefore acoustic icons are also referred to as “earcons”. When presented spatially, a large variety of situations and application is available, either related to the actual location and direction of an event or augmented to the specific spatial situation. Examples can be found in cockpits (see, e.g. (Bronkhorst et al. 1996)) or systems for orientation and navigation for visually impaired persons. Earcons can hence represent information of function (Fig. 18.1).

In the field of visual VR, the recent development of graphic processors enabled the simulation and reproduction of rather complex scenarios with a high degree of realism. Furthermore, the quality of technology for VR subsystems, such as motion trackers and projections units, has improved significantly and is available at

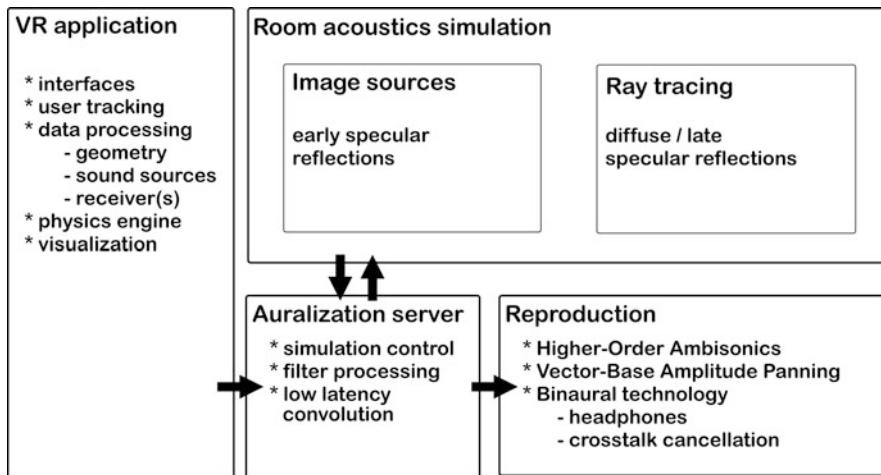


Fig. 18.1 VR system with integration of 3D room acoustics

reasonable costs. As soon as application software is available, it might be possible that VR technology as such is no longer only used in highly sophisticated laboratory arrangements but also for applications of daily use and for the consumer market. Particularly the visual dimension of VR is used today by many groups dealing with visualization of complex numerical or experimental data, for instance, in fluid dynamics or molecular physics.

For the acoustic rendering, several methods have been published (systems integrated in WFS or CTC are examples which were already mentioned), for instance (Huopaniemi et al. 1996; Gröhn et al. 2007). Systems on special hardware are controlled by software plug-ins (Lake Huron 2005; AuSim06). Spatial acoustic parameters like in (OpenAL 2006) are useful for programming acoustic sources.¹ This approach is typically used to create simple acoustic images for computer games or other fields of consumer electronics which do not require the same physical correctness. Other software solutions are given by (Storms 1995; Naef et al. 2002), while (Savioja 1999; Savioja et al. 1999; Hahn et al. 1995) focus on the component of sound reproduction.

The rapid development of CPU and memory technology will also influence VR. PCs will be equipped with multi core technology and parallel concepts. Ray tracing, image source audibility check or other software components can well be implemented on a parallel CPU architecture. Apart from constraints due to data traffic, the gain in simulation speed depends only on the number of processors involved.

This computational power is required for physically based interaction models of haptic, acoustic and visual feedback. In such an environment, the scenes are not

¹But not for realistic spatial scenes in rooms

simple effects,² but in real-time calculated physical reactions. The main conceptual difference is that no database containing a number of fixed reactions is used, like in computer games, but all reactions are created based on physical model equations. The variety of scenes, the plausibility of the virtual environment is thus much larger, so that there is no practical limit for memorized action and reaction. The user can operate more intuitively in the virtual scene. Interesting studies can focus, for instance, on modified physical parameters such as objects falling and bouncing in gravitation on the Moon or the Mars or alternative behaviour of physical properties of objects.

To obtain best possible immersion and presence, the human senses must be addressed in virtual environments in a most realistic way. The user of a VR system should act as freely and naturally as possible. For better immersion, spatial acoustics is necessarily been included, and this is required more and more in large projection systems like L-bench or CAVE-like environment.

The same holds for training in driving and flight simulators, where the projection unit is combined with a large mechanical platform including all human machine interfaces required for the task. The aircraft cabin, for instance, in this respect must be equipped with all components of the virtual environment, and it must reproduce all cues of vision, sound, forces, etc., which represent the actual training situation as best as possible.

Head-mounted displays (HMD) were quite common in the 1990s, but in scientific, VR applications there seems to be a tendency towards setups where the user does not have to wear too much hardware. Nevertheless HMDs became very popular for gaming, and devices at rather low cost arrived in the consumer market.

18.2 Head-Mounted Displays

Head-mounted displays (HMDs) have two small video displays integrated into eyeglasses or a helmet. Two slightly shifted images for the left and right eyes are created for the illusion of depth in the video, similar as with old stereoscopic devices.

When looking at nearby objects, binocular vision provides an essential means of correctly estimating distances. With the right eye, we see a nearby object projected onto a different part of the retina than the left, and this difference becomes more significant as the object gets closer. If we focus both eyes on one point, the two axes of the eye form an angle, which increases the closer the object is. Close objects are seen a little more from one side with the right eye and a little more from the other side with the left eye. The different angles of both eyes produce two different pictures. The lens curvature of the eye then adapts to the distance of the viewed object to produce a sharp image on the retina. The size of the viewing angle and the amount of

²Like in a movie, created once and for one specific intention of effect or feedback



Fig. 18.2 VR user with HMD. The person sees the image of the virtual scene of an entrance foyer as shown on the screen

accommodation provide a measure of the distance of the objects. Thus it is created an illusion of 3D vision.

Methods such as frame multiplexing expand two separate video signals from one signal by alternating the left and right images in successive frames. This preserves the full resolution per each image, but it reduces the frame rate by half. Not all HMDs, however, provide depth perception. Some lower-end modules are essentially bi-ocular devices where both eyes are presented with the same image. 3D video players sometimes allow maximum compatibility with HMDs by providing the user with a choice of the 3D format to be used.

As concerns the auditory component, HMDs are mostly equipped with headphones. Accordingly, the technical details of headphone technology explained in [17.2](#) are relevant and fully sufficient to describe the auditory component of HMDs. One example for application of psychophysical tests in an immersive virtual acoustic environment is (Imran et al. 2019b); see also Fig. [18.2](#).

18.3 CAVE Technology

The CAVE Automatic Virtual Environment, in short “CAVE”, was invented by (Cruz-Neira et al. 1992). A CAVE is a surrounding video display arrangement. The walls and possibly the floor, too, of a CAVE are projection screens. The user wears 3D glasses in order experience a stereoscopic image, similar to 3D movies. In contrast to 3D cinema, the user can freely move in the scene and interact with the scene’s object. This requires tracking devices usually electromagnetic or infrared

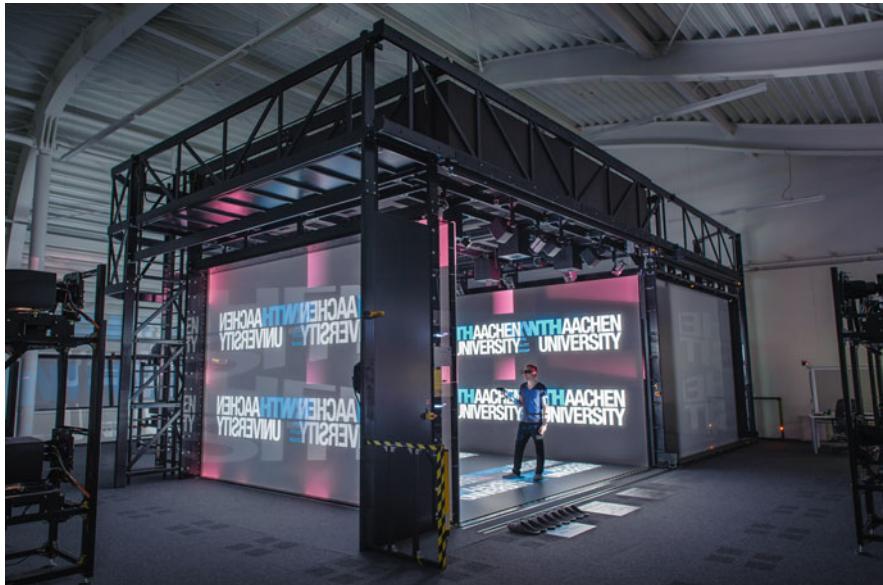


Fig. 18.3 The CAVE-like environment at RWTH Aachen University (aixCAVE). (Kuhlen and Hentschel 2014)

camera systems. The tracker signal can then be used for adjustment of the binocular and the binaural signals for vision and hearing, respectively.

From acoustic point of view, implementations of loudspeaker systems are difficult in combination with rigid video screens. Acoustically transparent screens, on the other hand, do not guarantee sufficient sharpness, contrast and polarization information of the video images. Therefore compromises are unavoidable, see examples in Figs. 18.3, 18.4 and 18.5.

18.4 Open-Source Virtual Acoustics Software

With the rapid progress in science and technology, one of the main challenges is reproducible research. This means that numerical or experimental data, its generation and the background conditions, equipment and calibration shall be documented in order to let scientists reproduce the results from published work. Accordingly, data are more and more published along with the written journal articles or conference papers. These data are often required for “open access”. Libraries, thus, don’t register and archive only text publications but data publications, too.

In software development, there is a similar trend towards open-source algorithms, which are documented and made available to the scientific community in order to get feedback to the authors of the algorithms and to get progress from what was created



Fig. 18.4 The “Virtual Concert Hall” at Technical University Berlin. The test subject (left) is audiovisually embedded in a hemicylindrical 3D video screen with tracked headphone audio reproduction. (Maempel and Horn 2017)

already. Particularly in academic environments in projects with public funding this is even a requirement to get the research projects accepted.

In acoustics, this trend resulted in very powerful open-source tools for creation of virtual auditory environments, a list of which is shown in the table below. Examples of open-source software are known on various levels in research and education (Table 18.1).

EVERTims (Noisternig et al. 2008; Poirier-Quinot et al. 2017) is an auralization framework for virtual acoustics and real-time room acoustic simulation. It consists of a C++ library for real-time room acoustic simulation, Blender and Unity 3D add-ons, Max externals and a spatial renderer application using JUCE.³

With Explauralisation (Krijnen et al. 2012), renderings of auditory walkthroughs and static impressions of buildings and other three-dimensional scenes can be created. It is based on Ray-tracing in a stand-alone application but tightly integrated into Blender. It is connected to the openPSTD project (Krijnen and Hornikx 2014) which is freely accessible for academic research purposes, which enables to efficiently compute sound propagation in the built environment.

³Blender: <https://www.blender.org>, Unity: <https://unity.com>Max: <https://cycling74.com>, JUCE: <https://juce.com>



Fig. 18.5 The CAVE-like environment “CRAIVE” at the Rensselaer Polytechnic Institute. (Courtesy of Jonas Braasch)

i-simpa (Picaut and Fortin 2012) is a graphical user interface for three-dimensional numerical codes for the modelling of sound propagation in complex geometrical domains. It is a tool for acousticians, for teachers and students as well as for researchers.

Pachyderm (van der Harten 2013) is an open source collection of numerical and geometrical simulation techniques curated by open research in acoustical science and education. It includes geometrical acoustics (ray-tracing and image source with enhancements, and visualization), finite volume method, and transfer matrix techniques for calculation of acoustic absorption which can be applied in the simulation techniques.

VA, virtual acoustics (Wefers and Vorländer 2018), creates dynamic auditory worlds that can be interactively experienced. It uses real-time simulation backends or simplified models to create a physics-based auditory impression. VBA (virtual building acoustics) is using VA as rendering engine with an extension of filter design and processing for simulation of sound transmission paths in building structures. This way, immersive environments can be created for sound propagation and sound effects’ simulation for listeners in buildings even if the sound excitation is happening other rooms or outdoors.

Table 18.1 Links to open-source virtual acoustics software

Name	URL
EVERTims	https://evertims.github.io
Explauralisation	http://explauralisation.org
i-simpa	https://i-simpa.ifsttar.fr
openPSTD	http://www.openpstd.org
Pachyderm	http://www.orase.org
VA	http://www.virtualacoustics.org
VBA	http://www.virtualbuildingacoustics.org

18.5 Open Data and Benchmarks for Virtual Acoustics

18.5.1 Room Acoustics

Reference models and reference measurement data can serve as basis for comparison between implementations of simulation and auralization. Simulation software round robins in the 1990s were used to study the accuracy and the limitations of room acoustic simulation (Vorländer 1995; Bork 2000, 2005a, b; Brinkmann et al. 2019); see also 11.6. More such test frameworks are now available in open access of data, also including auralization filters, binaural impulse responses and complete documentation of the scenarios, sources and receivers (Fig. 18.6). These scenes cover a variety of basic and complex environments, starting with a simple reflection plane, finite size reflectors, diffusors and finally including small to large rooms. For all scenes, measurements were preformed (omnidirectional and binaural). The database is available as a data publication⁴ under the licence Creative Commons BY-SA 4.0. It may be used by developers and users of software for comparison and validation.

18.5.2 Computational Acoustics Benchmarks

Sound field models which are based on solutions of the wave equation were introduced in Sect. 10.2. They can best be evaluated in comparison with results from analytic calculations. This way, deviations due to measurement uncertainties are avoided. Well-known examples are the “Stanford Bunny” (Turk and Levoy 1994) for computer images, or the “Cat’s Eye model” (Mechel 2005) for scattering problems in light and sound wave propagation.

More benchmark problems are relevant when it comes to indoor problems and outdoor problems in various settings of object dimensions in comparison with the wavelength.

In computational acoustics research, the benchmarks were defined and agreed upon in technical expert groups. They contain linear acoustic wave propagation as a

⁴<https://doi.org/10.14279/depositonce-6726>

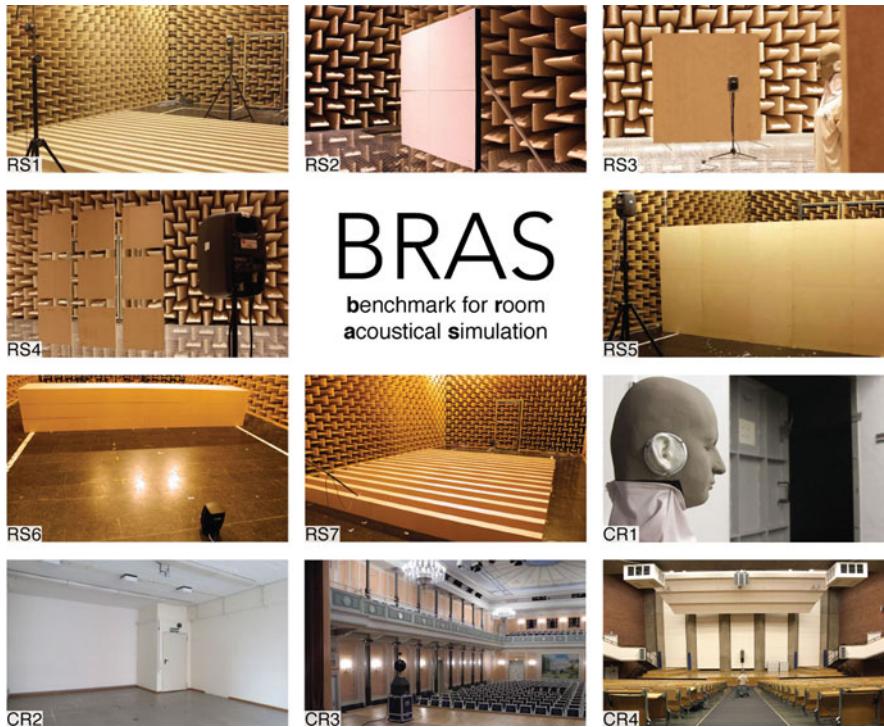


Fig. 18.6 Eleven scenes of the reference database: Scene 1–7 represent isolated acoustic phenomena, scene 8 includes coupled rooms, scenes 9, 10 and 11 include rooms in three categories of volume and reverberation time

solution of the wave equation, including scattering and radiation problems, interior and exterior problems, time domain and frequency domain problems, higher frequency problems for which consideration of energy terms might be more interesting than deterministic evaluation of acoustic quantities such as pressure or particle velocity, fluid-structure interaction problems, including transmission loss through (layered) plates and shells, structure, porous media, flow- and aeroacoustics (with coupling to structural vibrations), and underwater acoustics.

One benchmark database was published by Otsuru et al. (2005). The most extensive project in this respect is the European Acoustics Association “Benchmark Cases for Computational Acoustics” (Hornikx et al. 2015).⁵

⁵<https://eaa-bench.mec.tuwien.ac.at/benchmarks/>

Annex

Material Data

Basic parameters of acoustic materials are the impedance and the surface shape. The equations given in Sect. 3.1 are applicable as first-order approximation of the impedance and reflection factor of acoustic materials. To discuss this in more detail, other parameters such as angle-dependent impedance, porosity, tortuosity, etc. are required. These material data include all necessary information required for calculations of the reflected and the transmitted field. In many cases of sound prediction, however, the absorbed or transmitted energy is a sufficient quantity. When predicting the exterior sound field for the purpose of auralization, the absorption coefficient and the scattering coefficient for random incidence are important input data used in room acoustics simulation software. For predicting sound insulation, the sound reduction index or other standardized sound transmission data are required.

Most data in the following tables are extracted from the most recently established and widely used database of absorption coefficients. The database was developed by Ingolf Bork as part of the project of the “round robin” on room acoustical computer simulations. Other data were obtained from product data provided by manufacturers.

The reference methods used to determine these data are standardized methods for material testing, ISO354 for absorption coefficients, ISO17497 part 1 for scattering coefficients and ISO140 for sound insulation quantities. All standards describe measurement methods obtained in reverberation chambers. For more details see (Bork 2005b)¹ and the listed ISO standards.

These data are applicable for geometric or other energetic prediction models such as ray tracing or SEA. They are, however, not precise enough and do not provide information required for numerical wave models.

¹<https://www.ptb.de/cms/ptb/fachabteilungen/abt1/fb-16/ag-163/absorption-coefficient-database.html>

Tables of Random-Incidence Absorption Coefficients, α

Unless not explicitly specified differently, the data given are random-incidence absorption coefficients, α_s (see [Sect. 3.1](#)).

Massive Constructions and Hard Surfaces

Material	Octave band frequency in Hz						
	125	250	500	1 k	2 k	4 k	8 k
Walls, hard surfaces average (brick walls, plaster, hard floors, etc.)	0.02	0.02	0.03	0.03	0.04	0.05	0.05
Walls, rendered brickwork	0.01	0.02	0.02	0.03	0.03	0.04	0.04
Rough concrete	0.02	0.03	0.03	0.03	0.04	0.07	0.07
Smooth unpainted concrete	0.01	0.01	0.02	0.02	0.02	0.05	0.05
Rough lime wash	0.02	0.03	0.04	0.05	0.04	0.03	0.02
Smooth brickwork with flush pointing, painted	0.01	0.01	0.02	0.02	0.02	0.02	0.02
Smooth brickwork, 10-mm-deep pointing, pit sand mortar	0.08	0.09	0.12	0.16	0.22	0.24	0.24
Brick wall, stuccoed with a rough finish	0.03	0.03	0.03	0.04	0.05	0.07	0.07
Ceramic tiles with smooth surface	0.01	0.01	0.01	0.02	0.02	0.02	0.02
Limestone wall	0.02	0.02	0.03	0.04	0.05	0.05	0.05
Reverberation chamber walls	0.01	0.01	0.01	0.02	0.02	0.04	0.04
Concrete floor	0.01	0.03	0.05	0.02	0.02	0.02	0.02
Marble floor	0.01	0.01	0.01	0.02	0.02	0.02	0.02

Lightweight Constructions and Linings

Material	Octave band frequency in Hz						
	125	250	500	1 k	2 k	4 k	8 k
2 × 13 mm plasterboard on steel frame, 50 mm mineral wool in cavity, surface painted	0.15	0.10	0.06	0.04	0.04	0.05	0.05
Wooden lining, 12 mm fixed on frame	0.27	0.23	0.22	0.15	0.10	0.07	0.06

Glazing

Material	Octave band frequency in Hz						
	125	250	500	1 k	2 k	4 k	8 k
Single pane of glass, 3 mm	0.08	0.04	0.03	0.03	0.02	0.02	0.02
Glass window 0.68 kg/m ²	0.10	0.05	0.04	0.03	0.03	0.03	0.03
Lead glazing	0.30	0.20	0.14	0.10	0.05	0.05	–
Double glazing, 2–3 mm glass, >30 mm gap	0.15	0.05	0.03	0.03	0.02	0.02	0.02
Double glazing, 2–3 mm glass, 10 mm gap	0.10	0.07	0.05	0.03	0.02	0.02	0.02
Double glazing, lead on the inside	0.15	0.30	0.18	0.10	0.05	0.05	–

Wood

Material	Octave band frequency in Hz						
	125	250	500	1 k	2 k	4 k	8 k
Wood, 1.6 cm thick, on 4 cm wooden planks	0.18	0.12	0.10	0.09	0.08	0.07	0.07
Thin plywood panelling	0.42	0.21	0.10	0.08	0.06	0.06	–
16 mm wood on 40 mm studs	0.18	0.12	0.10	0.09	0.08	0.07	0.07
Audience floor, 2 layers, 33 mm on sleepers over concrete	0.09	0.06	0.05	0.05	0.05	0.04	–
Wood, stage floor, 2 layers, 27 mm over airspace	0.10	0.07	0.06	0.06	0.06	0.06	–
Solid wooden door	0.14	0.10	0.06	0.08	0.10	0.10	0.10

Floor Coverings

Material	Octave band frequency in Hz						
	125	250	500	1 k	2 k	4 k	8 k
Linoleum, asphalt, rubber, or cork tile on concrete	0.02	0.03	0.03	0.03	0.03	0.02	–
Cotton carpet	0.07	0.31	0.49	0.81	0.66	0.54	0.48
Loop pile tufted carpet, 1.4 kg/m ² , 9.5 mm pile height: on hair pad, 3.0 kg/m ²	0.10	0.40	0.62	0.70	0.63	0.88	–
Thin carpet, cemented to concrete	0.02	0.04	0.08	0.20	0.35	0.40	–

(continued)

Material	Octave band frequency in Hz						
	125	250	500	1 k	2 k	4 k	8 k
6 mm pile carpet bonded to closed-cell foam underlay	0.03	0.09	0.25	0.31	0.33	0.44	0.44
6 mm pile carpet bonded to open-cell foam underlay	0.03	0.09	0.20	0.54	0.70	0.72	0.72
9 mm tufted pile carpet on felt underlay	0.08	0.08	0.30	0.60	0.75	0.80	0.80
Needle felt 5 mm stuck to concrete	0.02	0.02	0.05	0.15	0.30	0.40	0.40
10 mm soft carpet on concrete	0.09	0.08	0.21	0.26	0.27	0.37	–
Hairy carpet on 3 mm felt	0.11	0.14	0.37	0.43	0.27	0.25	0.25
5 mm rubber carpet on concrete	0.04	0.04	0.08	0.12	0.10	0.10	–
Carpet 1.35 kg/m ² , on hair felt or foam rubber	0.08	0.24	0.57	0.69	0.71	0.73	–
Cocos fibre roll felt, 29 mm thick (unstressed), reverse side clad with paper, 2.2 kg/m ² , 2 Rayl	0.10	0.13	0.22	0.35	0.47	0.57	–

Curtains

Material	Octave band frequency in Hz						
	125	250	500	1 k	2 k	4 k	8 k
Cotton curtains (0.5 kg/m ²) draped to 3/4 area approx. 130 mm from wall	0.30	0.45	0.65	0.56	0.59	0.71	0.71
Curtains (0.2 kg/m ²) hung 90 mm from wall	0.05	0.06	0.39	0.63	0.70	0.73	0.73
Cotton cloth (0.33 kg/m ²) folded to 7/8 area	0.03	0.12	0.15	0.27	0.37	0.42	–
Densely woven window curtains 90 mm from wall	0.06	0.10	0.38	0.63	0.70	0.73	–
Vertical blinds, 15 cm from wall, half opened (45°)	0.03	0.09	0.24	0.46	0.79	0.76	–
Vertical blinds, 15 cm from wall, open (90°)	0.03	0.06	0.13	0.28	0.49	0.56	–
Tight velvet curtains	0.05	0.12	0.35	0.45	0.38	0.36	0.36
Curtain fabric, 15 cm from wall	0.10	0.38	0.63	0.52	0.55	0.65	–
Curtain fabric, folded, 15 cm from wall	0.12	0.60	0.98	1.0	1.0	1.0	1.0
Curtains of close-woven glass mat hung 50 mm from wall	0.03	0.03	0.15	0.40	0.50	0.50	0.50
Studio curtains, 22 cm from wall	0.36	0.26	0.51	0.45	0.62	0.76	–

Seating (Two Seats per m²)

Material	Octave band frequency in Hz						
	125	250	500	1 k	2 k	4 k	8 k
Wooden chairs without cushion	0.05	0.08	0.10	0.12	0.12	0.12	—
Unoccupied plastic chairs	0.06	0.10	0.10	0.20	0.30	0.20	0.20
Medium upholstered concert chairs, empty	0.49	0.66	0.80	0.88	0.82	0.70	—
Heavily upholstered seats, unoccupied	0.70	0.76	0.81	0.84	0.84	0.81	—
Empty chairs, upholstered with cloth cover	0.44	0.60	0.77	0.89	0.82	0.70	0.70
Empty chairs, upholstered with leather cover	0.40	0.50	0.58	0.61	0.58	0.50	0.50
Unoccupied moderately upholstered chairs (0.90 m × 0.55 m)	0.44	0.56	0.67	0.74	0.83	0.87	—

Audience (Unless Not Specified Explicitly, Two Persons per m²)

	Octave band frequency in Hz						
	125	250	500	1 k	2 k	4 k	8 k
Areas with audience, orchestra or choir including narrow aisles	0.60	0.74	0.88	0.96	0.93	0.85	0.85
Audience on wooden chairs, 1 per m ²	0.16	0.24	0.56	0.69	0.81	0.78	0.78
Audience on wooden chairs, 2 per m ²	0.24	0.40	0.78	0.98	0.96	0.87	0.87
Orchestra with instruments on podium, 1.5 m ² per person	0.27	0.53	0.67	0.93	0.87	0.80	0.80
Audience area, 0.72 persons/m ²	0.10	0.21	0.41	0.65	0.75	0.71	—
Audience area, 1 person/m ²	0.16	0.29	0.55	0.80	0.92	0.90	—
Audience area, 1.5 persons/m ²	0.22	0.38	0.71	0.95	0.99	0.99	—
Audience area, 2 persons/m ²	0.26	0.46	0.87	0.99	0.99	0.99	—
Audience in moderately upholstered chairs 0,85 m × 0,63 m	0.72	0.82	0.91	0.93	0.94	0.87	—
Audience in moderately upholstered chairs 0,90 m × 0,55 m	0.55	0.86	0.83	0.87	0.90	0.87	—

Wall Absorbers

Material	Octave band frequency in Hz						
	125	250	500	1 k	2 k	4 k	8 k
Fabric-covered panel, 6 pcf rock wool core	0.46	0.93	1.0	1.0	1.0	1.0	1.0
Fabric-covered panel, 8 pcf rock wool core	0.21	0.66	1.0	1.0	0.97	0.98	0.98
Facing-brick brickwork, open butt joins, brick dimensions $230 \times 50 \times 55$ mm	0.04	0.14	0.49	0.35	0.31	0.36	–
Acoustical plaster, approx. 25 mm thick, $3.5 \text{ kg/m}^2/\text{cm}$	0.17	0.36	0.66	0.65	0.62	0.68	–
Rock wool thickness = 50 mm, 80 kg/m^3	0.22	0.6	0.92	0.90	0.88	0.88	0.88
Rockwool thickness = 50 mm, 40 kg/m^3	0.23	0.59	0.86	0.86	0.86	0.86	0.86
50 mm mineral wool (40 kg/m^3), glued to wall, untreated surface	0.15	0.70	0.60	0.60	0.85	0.90	0.90
50 mm mineral wool (70 kg/m^3) 300 mm in front of wall	0.70	0.45	0.65	0.60	0.75	0.65	0.65
Gypsum board, perforation 19.6%, hole diameter 15 mm, backed by fibrous web 12 Rayl, 100 mm cavity filled with mineral fibre mat 1.05 kg/m^2 , 7.5 Rayl	0.30	0.69	1.0	0.81	0.66	0.62	–
Perforated veneered chipboard, 50 mm, 1 mm holes, 3 mm spacing, 9% hole surface ratio, 150 mm cavity filled with 30 mm mineral wool	0.41	0.67	0.58	0.59	0.68	0.35	–
Fibre absorber, mineral fibre, 20 mm thick, 3.4 kg/m^2 , 50 mm cavity	0.20	0.56	0.82	0.87	0.70	0.53	–
Fibre absorber, mats of porous flexible fibrous web fabric, self-extinguishing	0.07	0.07	0.2	0.41	0.75	0.97	–

Ceiling Absorbers

Material	Octave band frequency in Hz						
	125	250	500	1 k	2 k	4 k	8 k
Plasterboard ceiling on battens with large air-space above	0.20	0.15	0.10	0.08	0.04	0.02	–
Fibre absorber on perforated sheet metal cartridge, 0.5 mm zinc-plated steel, 1.5 mm hole diameter, 200 mm cavity filled with 20 mm mineral wool (20 kg/m^3), inflammable	0.48	0.97	1.0	0.97	1.0	1.0	1.0
Fissured ceiling tile	0.49	0.53	0.53	0.75	0.92	0.99	–
Perforated 27 mm gypsum board (16%), d = 4.5 mm, 300 mm from ceiling	0.45	0.55	0.60	0.90	0.86	0.75	–
Wedge-shaped, melamine foam, ceiling tile	0.12	0.33	0.83	0.97	0.98	0.95	–
Metal panel ceiling, backed by 20 mm Sillan acoustic tiles, panel width 85 mm, panel spacing 15 mm, cavity 35 cm	0.59	0.80	0.82	0.65	0.27	0.23	–

Special Absorbers

Material	Octave band frequency in Hz						
	125	250	500	1 k	2 k	4 k	8 k
Micro-perforated foil "Microsorber" (Kaefer)	0.06	0.28	0.70	0.68	0.74	0.53	—
Micro-perforated glass sheets/5 cm cavity	0.10	0.45	0.85	0.30	0.10	0.05	—
Hanging absorber panels (foam), 400 mm depth, 400 mm distance	0.25	0.45	0.80	0.90	0.85	0.80	—
Hanging absorber panels (foam), 400 mm depth, 700 mm distance	0.20	0.30	0.60	0.75	0.70	0.70	—

Equivalent Absorption Area, A, in m², of Single Objects

Material	Octave band frequency in Hz						
	125	250	500	1 k	2 k	4 k	8 k
Single chair, wood	0.02	0.02	0.03	0.04	0.04	0.04	—
Single chair, upholstered	0.10	0.20	0.25	0.30	0.35	0.35	—
Individual person in a group, sitting or standing, 1 per 6 m ² area; typical minimum	0.05	0.10	0.20	0.35	0.50	0.65	—
Individual person in a group, sitting, 1 per 6 m ² area; typical maximum	0.12	0.45	0.80	0.90	0.95	1.0	1.1
Individual person in a group, standing, 1 per 6 m ² area; typical maximum	0.12	0.45	0.80	1.20	1.30	1.40	1.45

Air Attenuation Coefficient, in 10⁻³ m⁻¹

Temperature in °C and relative humidity	Octave band frequency in Hz						
	125	250	500	1 k	2 k	4 k	8 k
10°, 30–50%	0.1	0.2	0.5	1.1	2.7	9.4	29.0
10°, 50–70%	0.1	0.2	0.5	0.8	1.8	5.9	21.1
10°, 70–90%	0.1	0.2	0.5	0.7	1.4	4.4	15.8
20°, 30–50%	0.1	0.3	0.6	1.0	1.9	5.8	20.3
20°, 50–70%	0.1	0.3	0.6	1.0	1.7	4.1	13.5
20°, 70–90%	0.1	0.3	0.6	1.1	1.7	3.5	10.6

Tables of Random-Incidence Scattering Coefficients, s

Random-incidence scattering coefficients according to the definition in ISO 17497-1 (see also 3.4.2) are related to the surface shape and size. Relevant surface parameters are the length and depth of surface corrugations. It is assumed that the total surface area is large compared with the corrugations and the wavelengths. Please note that this concept is not applicable for single scattering objects.²

To determine the amount of scattering, it is essential that the shape is described by its dimensions, the average structural depth, h , and the average structural length, a , see Fig. A1. The random-incidence scattering coefficients, s , are given in relation to the normalized frequency $a/\lambda = f a/c$. Below $a/\lambda = 0.125$, the random-incidence scattering coefficient, s , is generally smaller than 0.05.

The data listed are rounded values from publications by Vorländer and Mommertz (2000), Jeon et al. (2003, 2004) and Embrechts et al. (2004) and results from other measurements. More information on diffusers and more detailed data are given in Cox and D'Antonio (2004).

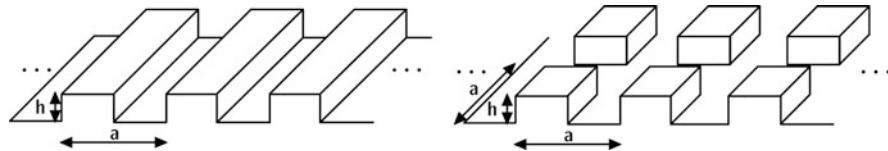


Fig. A.1 Definition of surface dimensions of 1D and 2D corrugations

²See (Cox and D'Antonio 2004) for more information on the single-object diffusion coefficient.

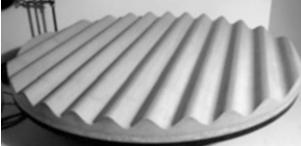
2D Surfaces

Shape of corrugation	a/λ						
	0.125	0.25	0.5	1	2	4	8
Hemispheres of average radius h , randomly distributed, coverage 40% ($h/a \approx 0.25$)	0.1	0.2	0.5	0.6	0.6	0.7	0.8
							
Densely placed identical hemispheres of radius h , $h/a = 0.5$ in regular pattern	0.05	0.05	0.1	0.6	0.6	0.6	—
Hemispheres of average radius h , randomly distributed, coverage 25% ($h/a \approx 0.15$)	0.1	0.1	0.2	0.3	0.4	0.4	0.4
Wooden cubes, regular pattern, $h/a = 0.5$	0.05	0.05	0.25	0.3	0.7	0.9	—
Wooden cubes, random distance and orientation $h/a = 0.5$	0.05	0.05	0.2	0.3	0.6	0.7	—
Ceramic tiles, densely packed, heights h distributed in a range between 1 and 10, average $h/a \approx 1$	0.1	0.4	0.9	0.7	0.7	0.7	—
							
Wooden boxes of various sizes, random pattern, average $h/a = 0.5$	0.05	0.05	0.15	0.4	0.7	0.9	—
							

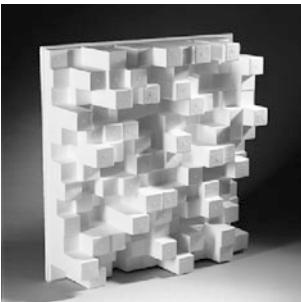
(continued)

Shape of corrugation	a/λ						
	0.125	0.25	0.5	1	2	4	8
Trapezoidal grating $h/a \approx 0.5$	0.05	0.05	0.1	0.9	0.8	0.9	0.9
							

1D Surfaces

Shape of corrugation	a/λ						
	0.125	0.25	0.5	1	2	4	8
Stairs (sawtooth) $h/a = 0.3$	0.05	0.05	0.2	0.3	0.4	0.45	—
Aperiodically distributed rectangular battens, $h/a = 0.5$	0.1	0.6	0.5	0.4	0.3	0.4	—
Periodically distributed rectangular battens, $h/a = 0.5$	0.1	0.6	0.6	0.5	0.5	0.5	—
Periodically distributed hemi-cylinders $h/a = 0.25$	0.1	0.1	0.3	0.7	0.8	0.8	—
Sinusoidal, $h/a = 0.31$	0.05	0.05	0.2	0.7	0.8	0.85	—
							

Diffusers

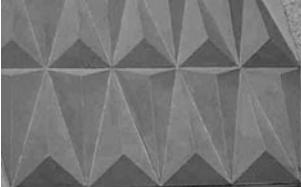
Type	Octave band frequency in Hz						
	125	250	500	1 k	2 k	4 k	8 k
RPG “Skyline”	0.01	0.08	0.45	0.82	1.0	—	—
							
RPG “QRD”	0.06	0.15	0.45	0.95	0.88	0.91	—
							

(Courtesy of RPG Diffusor Systems, Inc.; www.rpginc.com)

Seating and Audience

	Octave band frequency in Hz						
	125	250	500	1 k	2 k	4 k	8 k
Shape of corrugation							
Theatre audience	0.3	0.5	0.6	0.6	0.7	0.7	0.7
Amphitheatre steps, length 82 cm, height 30 cm (Farnetani 2005)	0.05	0.45	0.75	0.9	0.9	—	—
Rows of classroom tables and persons sitting on chairs	0.2	0.3	0.4	0.5	0.5	0.6	0.6

Round Robin III – Wall and Ceiling

Shape of corrugation	Octave band frequency in Hz						
	125	250	500	1 k	2 k	4 k	8 k
Rectangular and prism boxes (studio wall), “Round Robin III” (after (Bork 2005a))	0.50	0.90	0.95	0.95	0.95	0.95	–
							
Trapezoidal boxes (studio ceiling), “Round Robin III” (after (Bork 2005a))	0.13	0.56	0.95	0.95	0.95	0.95	–
							

Tables of Sound Reduction Indices, *R*

Masonry

Material	Octave band frequency in Hz						
	125	250	500	1 k	2 k	4 k	<i>R_w</i>
Lightweight concrete (1200 kg/m ³) 140 mm	30.9	32.1	42.0	49.1	54.4	59.9	45
Concrete (2300 kg/m ³) 240 mm	45.6	51.9	58.7	66.0	70.7	72.3	63
Aerated concrete (400 kg/m ³) 150 mm, render 2 × 10 mm	24.1	25.9	35.6	42.4	47.7	53.2	39
Brick (1400 kg/m ³) 175 mm, render 2 × 15 mm	28.6	43.0	50.3	58.1	63.2	68.1	52
Calcium silicate (1200 kg/m ³) 115 mm	29.1	33.0	40.1	47.8	54.9	60.4	44
Calcium silicate (2000 kg/m ³) 175 mm, render 2 × 15 mm	39.6	45.6	52.7	60.3	65.4	70.0	56

Lightweight Constructions

Material	Octave band frequency in Hz						
	125	250	500	1 k	2 k	4 k	R_w
Gypsum board 2 × 12.5 mm with 25 mm filled gap	30.0	43.0	53.0	60.0	65.0	50.0	51
2 × 15 mm WallBoard 146 mm, "C" studs, 25 mm Isowool APR 1200	33.8	35.6	51.7	56.2	59.5	49.8	51

Doors

Material	Octave band frequency in Hz						
	125	250	500	1 k	2 k	4 k	R_w
Chipboard. 13 mm, P 3 20 mm, chipboard. 13 mm	19.2	35.0	37.9	3.9	35.6	42.0	38
Veneer 6 mm, TW 1 40 mm, gypsum board. 12.5 mm, veneer 10 mm	22.0	33.0	40.0	40.0	44.0	40.0	40
Veneer 6 mm, TW 1 50 mm, veneer 10 mm	21.0	21.0	36.0	37.0	41.0	40.0	35
Wood fibre 3.5 mm, slats, wood fibre 3.5 mm	27.0	27.0	29.0	28.0	30.0	35.0	30

Glazing

Material	Octave band frequency in Hz						
	125	250	500	1 k	2 k	4 k	R_w
Single pane 3 mm	18.7	22.0	24.2	28.6	34.7	29.4	29
6 mm	23.4	27.4	31.8	35.2	26.8	35.5	32
10 mm	26.6	30.1	32.2	30.6	34.9	45.3	33
12 mm	31.3	33.1	31.5	32.3	39.4	45.7	34
Double glazing 4-6-4	25.7	25.0	23.4	34.1	40.4	36.5	31
Double glazing 8-12-8	29.2	27.3	31.1	36.8	35.0	46.7	34

References

- Ackermann D, Böhm C, Brinkmann F, Weinzierl S (2019) The acoustical effect of musicians' movements during musical performances. *Acust Acta Acust* 105:356
- Ahnert W, Feistel R (1993) EARS auralization software. *J Audio Eng Soc* 41:894
- Alarcão D, Bento Coelho JL (2003) Lambertian enclosures – a first step towards fast room acoustics simulation. *Build Acoust* 10:33
- Allen JB, Berkley DA (1979) Image method for efficiently simulating small-room acoustics. *J Acoust Soc Am* 65:943
- Aretz M (2007) Modellierung der wechselseitigen Kopplung von Luftschall mit Plattenstrukturen für raumakustische FEM Simulationen (Modelling of the mutual coupling of airborne sound with plate structures for room acoustic FEM simulations). Diploma (MSc) thesis RWTH Aachen University, Aachen, Germany
- Aretz M, Vorländer M (2014a) Combined wave and ray based room acoustic simulations of audio systems in car passenger compartments, part I: boundary and source data. *Appl Acoust* 76:82
- Aretz M, Vorländer M (2014b) Combined wave and ray based room acoustic simulations of audio systems in car passenger compartments, part II: comparison of simulations and measurements. *Appl Acoust* 76:52
- Arntzen M, Rizzi SA, Visser HG, Simons DG (2014) Framework for simulating aircraft flyover noise through non-standard atmospheres. *AIAA J Aircraft* 51:956
- Atal BS, Schroeder MR (1963) Apparent sound source translator, Tech. report. US Patent 3,236,949, February 23, 1963
- Atal B, Schroeder MR, Sessler GM (1965) Subjective reverberation time and its relation to sound decay. In: Proceedings of the ICA, Liège, Belgium, G32
- AUSIM3D (2006). <http://www.ausim3d.com>
- Barron M (1971) The subjective effects of first reflections in concert halls – the need for lateral reflections. *J Sound Vib* 15:475
- Barron M (2000) Auditorium acoustics and architectural design. E&FN Spon, London
- Bass HE, Sutherland LC, Zuckerwar AJ, Blackstock DT, Hester DM (1995) Atmospheric absorption of sound: further developments. *J Acoust Soc Am* 97:680
- Bauer BB (1963) Stereophonic earphones and binaural loudspeakers. *J Audio Eng Soc* 9:148
- Begault D (1991) Challenges to the successful implementation of 3-D sound. *J Audio Eng Soc* 39:864
- Begault D (1994) 3-D sound for virtual reality and multimedia. Academic Press Professional, Cambridge, MA
- Behler G, Müller S (2000) Technique for the derivation of wide band room impulse response. In: Proceedings of Tecniacustica/FIA Madrid, paper aaq11

- Behler G, Genuit K, Sottek R, Vorländer M (2006) Description of broadband structure-borne and airborne noise transmission from the powertrain. In: Proceedings of Fisita, world automotive congress, Yokohama, Japan
- Behler GK, Pollow M, Vorländer M (2012) Measurements of musical instruments with surrounding spherical arrays. In: Proceedings of IoA/CFA Nantes
- Bento Coelho JL, Alarcão D, Almeida AM, Abreu T, Fonseca N (2001) Contributions on room acoustical computer simulation – room acoustics design by a sound energy transition approach. *Acust Acta Acust* 86:903
- Berkhout AJ (1988) A holographic approach to acoustic control. *J Audio Eng Soc* 36:977
- Berzborn M, Vorländer M (2019) A high order rigid spherical microphone array design using MEMS microphones. In: Proceedings of 23rd ICA Aachen 2019, Germany
- Biot MA, Tolstoy I (1957) Formulation of wave propagation in infinite media by normal coordinates with an application to diffraction. *J Acoust Soc Am* 29:381
- Blanc-Benon P, Dallois L, Juvé D (2001) Long range sound propagation in a turbulent atmosphere within the parabolic approximation. *Acta Acust Acust* 87:659
- Blauert J (1996) Spatial hearing: the psychophysics of human sound localization, 2nd edn. MIT Press, Cambridge, MA
- Blauert J (ed) (2005) Communication acoustics. Springer, Berlin Heidelberg/New York
- Blesser B (2001) An interdisciplinary synthesis of reverberation viewpoints. *J Audio Eng Soc* 49:867
- Bomhardt R, Braren H, Fels J (2016) Individualization of head-related transfer functions using principal component analysis and anthropometric dimensions. *POMA* 29:050007
- Börger G, Blauert J, Laws P (1977) Stereophone Kopfhörerwiedergabe mit Steuerung bestimmter Übertragungsfaktoren durch Kopfdrehbewegungen. *Acustica* 39:22
- Borish J (1984) Extension of the image model to arbitrary polyhedra. *J Acoust Soc Am* 75:1827
- Bork I (2000) A comparison of room simulation software – the 2nd round robin on room acoustical computer simulation. *Acust Acta Acust* 84:943
- Bork I (2005a) Report on the 3rd round robin on room acoustical computer simulation, part I: measurements. *Acta Acust Acust* 91:740
- Bork I (2005b) Report on the 3rd round robin on room acoustical computer simulation, part II: calculations. *Acta Acust Acust* 91:753
- Bösing M (2013) Acoustic modeling of electrical drives: noise and vibration synthesis based on force response superposition. Doctoral thesis, RWTH Aachen University
- Braasch J (2005) In: Blauert J (ed) Modelling of binaural hearing. Chapter 4 in communication acoustics. Springer, Berlin Heidelberg/New York
- Brandenburg K, Werner S, Klein F, Sladeczek C (2016) Auditory illusions through headphones: history, challenges, and new solutions. In: Proceedings of international congress on acoustics, Buenos Aires 2016
- Brinkmann F, Aspöck A, Ackermann D, Lepa S, Vorländer M, Weinzierl S (2019) A round robin on room acoustical simulation and auralization. *J Acoust Soc Am* 145:2746
- Brinkmann F, Aspöck L, Ackermann D, Opdam R, Vorländer M, Weinzierl S (2020) A benchmark for room acoustical simulation. Concept and database. *Appl Acoust*
- Bronkhorst AW, Veltman JAH, van Breda L (1996) Application of a three-dimensional auditory display in a flight task. *Hum Factors* 38:23
- Brungart DS, Rabinowitz WM, Durlach NI (1996) Auditory localization of a nearby point source. *J Acoust Soc Am* 100:2593
- Brungart DS, Simpson BD, McKinley RL, Kordik AJ, Dallman RC, Ovenshire DA (2004) The interaction between head-tracker latency, source duration, and response time in the localization of virtual sound sources. In: Proceedings of ICAD '04 – 10th meeting of the international conference on auditory display, Sydney
- Brunskog J, Hammer P (2003) The interaction between the ISO tapping machine and lightweight floors. *Acta Acust Acust* 89:296

- Burkhard MD, Sachs RM (1975) Anthropometric manikin for acoustic research. *J Acoust Soc Am* 58:214
- Burton AJ, Miller GF (1971) The application of integral methods to the solution of some exterior boundary value problems. *Proc R Soc Lond A* 323:201
- Calamia P, Svensson, UP (2007) Fast time-domain edge-diffraction calculations for interactive acoustic simulations. *EURASIP J Adv Signal Process*, Special Issue on Spatial Sound and Virtual Acoustics, article ID 63560
- Camilo TS, Medrado LO, Tenenbaum RA (2002) New software for room acoustics simulation: a study of its performance and validation by an international comparison. *J Acoust Soc Am* 112:2396
- Christensen F, Jensen CB, Møller H (2000) The design of VALDEMAR – an artificial head for binaural recording purposes. In: Proceedings of 109th AES Convention, Los Angeles, Preprint 4404
- Cops A, Myncke H (1973) Determination of sound absorption coefficients using a tone-burst technique. *Acustica* 29:287
- Cox TJ, D'Antonio P (2004) Acoustic absorbers and diffusers: theory, design and application. Spon Press, London/New York
- Cox TJ, Dalenbäck B-I, D'Antonio P, Embrechts JJ, Jeon JY, Mommertz E, Vorländer M (2006) A tutorial on scattering and diffusion coefficients for room acoustic surfaces. *Acta Acust Acust* 92:1
- Cremer L (1948) Die wissenschaftlichen Grundlagen der Raumakustik. Band I, Geometrische Raumakustik. Hirzel-Verlag Stuttgart
- Cremer L, Heckl M, Ungar EE (1973) Structure-borne sound. Springer, Berlin/Heidelberg/New York
- Cruz-Neira C, Sandin DJ, DeFanti TA, Kenyon RV, Hart JC (1992) The CAVE: audio visual experience automatic virtual environment. *Commun ACM* 35(6):64
- Dalenbäck B-I (1995) A new model for room acoustic prediction and auralization. Doctoral thesis, Chalmers University, Gothenburg, Sweden
- Dalenbäck B-I (1996) Room acoustic prediction based on a unified treatment of diffuse and specular reflection. *J Acoust Soc Am* 100:899
- Dalenbäck B-I, McGrath D (1995) Narrowing the gap between virtual reality and Auralization. In: Proceedings of the 15th ICA, Trondheim, p 429
- Dalenbäck B-I, Strömberg M (2006) Real time walkthrough auralization – the first year. In: Proceedings of the IoA spring conference, Copenhagen
- Daniel J (2000) Acoustic field representation, application to the transmission and the reproduction of complex sound environments in a multimedia context. Doctoral thesis, Université Paris 6, France
- Daniel, J, Moreau S (2004) Further study of sound field coding with higher order ambisonics. In: Proceedings of 118th AES Convention Berlin
- Denon – Anechoic Orchestral Music Recording (1995) Audio CD, Denon Records, ASIN: B0000034M9
- Dohm M (2004) Untersuchungen zum Übersprechen bei der Körperschallübertragung in Kraftfahrzeugen (Investigation on the crosstalk in structure-borne sound transmission in vehicles). Diploma (MSc) thesis RWTH Aachen University and HEAD acoustics, Germany
- Dross P (2006) Real-time capable reverberation estimator for virtual environments. Diploma (MSc) thesis, RWTH Aachen University, Aachen, Germany
- Duraiswami R, Zotkin DN, Li Z, Grassi E, Gumerov NA, Davis L (2005) High order spatial audio capture and its binaural head-tracked playback over headphones with HRTF cues. In: Proceedings of 119th AES Convention New York, NY, USA
- EBU – Sound Quality Assessment Material (1988) Recordings for subjective tests. SQAM Compact Disk, Tech. 3253-E
- Embrechts JJ, De Geeter L, Vermeir G, Vorländer M, Sakuma T (2004) Calculation of the random-incidence scattering coefficients of a sine-shaped surface. *Acta Acust Acust* 92:593

- EN12354 – Building acoustics: Estimation of acoustic performance of buildings from the performance of elements, 6 Parts
- Erradj A (2019) Sound path algorithm with higher order reflection and diffraction components for the auralization of urban environments. MSc thesis RWTH Aachen University, Germany
- Fahy FJ (1995) The vibro-acoustic reciprocity principle and applications to noise control. *Acustica* 81:544
- Fant G (1970) Acoustic theory of speech reproduction, 2nd edn. Mouton, Den Haag, Paris
- Farina A (1995) RAMSETE – a new pyramid tracer for medium and large scale acoustic problems. In: Proceedings of EURONOISE, Lyon, p 55
- Farnetani A (2005) Investigation on the acoustics of ancient theatres by means of modern technologies. PhD thesis, University of Ferrara, Italy
- Fastl H (1998) Pitch strength and frequency discrimination for noise bands or complex tones. In: Palmer AR et al (eds) *Psychophysical and physiological advances in hearing*. Whurr Publishers, London
- Fels J, Buthmann VM (2004) Head-related transfer functions of children. *Acta Acust Acust* 90:918
- Fletcher NH, Rossing TD (1998) *The physics of musical instruments*. Springer, New York
- Freiheit R (2005) Historic recording gives choir “alien” feeling: in anechoic space, no one can hear you sing – experiment benefits acoustical research and design. In: Proceedings of the ASA/NOISE-CON 05, Minneapolis, MN
- Fruhmann M (2004a) On the pitch strength of bandpass noises. In: Proceedings of ICA 2004, Kyoto, Japan
- Fruhmann M (2004b) On the pitch strength of harmonic complex tones and combfilter noises. In: Proceedings of the CFA/DAGA ‘04, Strasbourg, France
- Funkhouser TA, Carlbom I, Elko G, Pingali G, Sondhi M, West J (1998) A beam tracing approach to acoustic modelling for interactive virtual environments. In: Computer Graphics, SIGGRAPH ‘98, p 21
- Funkhouser TA, Min P, Carlbom I (1999) Real-time acoustic modelling for distributed virtual environments. In: ACM Computer Graphics, Proceedings of the SIGGRAPH ‘99, p 365
- Gardner WG (1997) 3-D audio using loudspeakers. Ph.D thesis, Massachusetts Institute of Technology
- Genuit K (1984) Ein Modell zur Beschreibung von Assenohrübertragungseigenschaften. Doctoral thesis RWTH Aachen University, Germany
- Genuit K (2000) Application of binaural transfer path analysis to sound quality tasks. In: Proceedings of the European conference on vehicle noise and vibration, IMechE HQ, London
- Genuit K, Xiang N (1997) Binaural “hybrid” model for simulation of engine and wind noise in the interior of vehicles. In: Proceedings of the SAE noise & vibration conference, Traverse City
- Gerretsen E (1979) Calculation of the sound transmission between dwellings by partitions and flanking structures. *Appl Acoust* 12:413
- Gerretsen E (1986) Calculation of airborne and impact sound insulation between dwellings. *Appl Acoust* 19:245
- Gerzon MA (1976) Multidirectional sound reproduction systems. UK-Patent no. 3 997 725
- Gibbs BM, Qi N, Moorhouse AT (2007) A practical characterisation for vibro-acoustic sources in buildings. *Acta Acust Acust* 93:84
- Giron F (1996) Investigations about the directivity of sound sources. Doctoral thesis, Ruhr-Universität Bochum, Germany
- Gottlob D (1973) Vergleich objektiver Parameter mit Ergebnissen subjektiver Untersuchungen an Konzertsälen. Doctoral Thesis, Göttingen, Germany
- Griesinger D (1989) Practical processors and programs for digital reverberation. In: Proceedings of AES 7th international conference, Toronto
- Gröhn M, Lokki T, Takala T (2007) Localizing sound sources in a cave-like virtual environment with loudspeaker array reproduction. *Presence Teleop Virt Environ* 16:157
- Hahn JK, Geigel J, Lee JW, Gritz L, Takala T, Mishra S (1995) An integrated approach to motion and sound. *J Vis Comput Animat* 6:109

- Hamilton B, Bilbao S (2017) FDTD methods for 3-D room acoustics simulation with high-order accuracy in space and time. *IEEE/ACM Trans Audio, Speech, Lang Process* 25:11
- Hammershøi D (1995) Binaural technique – a method of true 3D sound reproduction. Doctoral thesis, Aalborg University, Denmark
- Hammershøi D, Møller H (2002) Methods for binaural recording and reproduction. *Acust Acta Acust* 88:303
- Hammershøi D, Møller H (2005) Binaural technique – basic methods for recording, synthesis and reproduction. In: Blauert (ed) *Communication acoustics*. Springer, Berlin/Heidelberg/New York
- Hansen V, Munch G (1991) Making records for simulation tests in the Archimedes project. *J Audio Eng Soc* 39:768
- Hargreaves J, Rendell LR, Lam YW (2019) A framework for auralization of boundary element method simulations including source and receiver directivity. *J Acoust Soc Am* 145:2625
- Heimes, A, Imran M, Vorländer M (2019) A real-time virtual reality building acoustic auralization framework for psychoacoustic experiments with contextual and interactive features. In: Proceedings of the 23rd ICA Aachen 2019, Germany
- Heinz R (1993) Binaural room simulation based on the image source model with addition of statistical methods to include the diffuse sound scattering of walls and to predict the reverberant tail. *Appl Acoust* 38:145
- Hodgson M (1983) Theoretical and physical models as tools for the study of factory sound fields. Doctoral thesis, University of Southampton, UK
- Hornikx M, Kaltenbacher M, Marburg S (2015) A platform for benchmark cases in computational acoustics. *Acta Acust Acust* 101:811
- Houtgast T, Steeneken H (1973) The modulation transfer function in room acoustics as a predictor of speech intelligibility. *Acustica* 28:66
- Huopaniemi J, Savioja L, Takala T (1996) DIVA virtual audio reality system. In: Proceedings of the international conference auditory display (ICAD '96), Palo Alto, California, p 111
- Huopaniemi J, Zacharov N, Karjalainen M (1997) Objective and subjective evaluation of head-related transfer function filter design. *J Audio Eng Soc* 47:218
- Imran M, Heimes A., Vorländer M. (2019a) Sound insulation auralization filters design for outdoor moving sources. In: Proceedings of the 23rd ICA Aachen 2019, Germany
- Imran M, Vorländer M, Schlittmeier SJ (2019b) Audio-video virtual reality environments in building acoustics: an exemplary study reproducing performance results and subjective ratings of a laboratory listening experiment. *J Acoust Soc Am* 146(3):EL310
- ISO140 – acoustics – measurement of sound insulation of buildings and of building elements (various parts)
- ISO17497 – acoustics – sound-scattering properties of surfaces – part 1: measurement of the random-incidence scattering coefficient in a reverberation room (ISO 17497-1:2004); part 2: Measurement of the directional diffusion coefficient in a free field (DIS 2007)
- ISO354 – acoustics – measurement of sound absorption in a reverberation room (ISO 354:2003)
- ISO9613 – acoustics – attenuation of sound during propagation outdoors – part 1: calculation of the absorption of sound by the atmosphere (ISO 9613-1:1993)
- ITU.p58 (1996) Head and torso simulator for telephonometry. International Telecommunication Union, Geneva
- Jedrzejewski M, Marasek K (2004) Computation of room acoustics using programmable video hardware. In: Proceedings of the ICCVG '04, Warsaw, Poland
- Jeon JY, Zhu L, Yoo K (2003) Miral concert hall: “ceramic palace” for sound scattering. In: Proceedings of the internoise Jeju, p 545
- Jeon JY, Lee SC, Vorländer M (2004) Development of scattering surfaces for concert halls. *Appl Acoust* 65:341
- Jeon JY, Ryu JK, Jeong JH, Tachibana H (2006) Review of the impact ball in evaluating floor impact sound. *Acta Acust Acust* 92:777

- Karjalainen M, Huang, O, Smith JO (2005) Digital Waveguide Networks for Room Response Modeling and Synthesis. Proc. AES Convention Barcelona, paper 6394
- Keller JB (1962) Geometrical theory of diffraction. J Opt Soc Am 52:116
- Kephalopoulos S, Paviotti M, Anfosso-Lédée F (2012) Common Noise Assessment Methods in Europe (CNOSSOS-EU). Reference report. Publications Office of the European Union, ISBN: 978-92-79-25281-5 (PDF), <https://doi.org/10.2788/31776> (online)
- Kirkeby O, Nelson PA (1999) Digital filter design for inversion problems in sound reproduction. J Audio Eng Soc 47:583
- Kirkeby O, Nelson PA, Hamada H (1998) The “stereo dipole” – a virtual source imaging system using two closely spaced loudspeakers. J Audio Eng Soc 46:387
- Kistler DJ, Wightman FL (1992) A model of head-related transfer functions based on principal components analysis and minimum phase reconstruction. J Acoust Soc Am 91:1637
- Kleiner M, Dalenbäck B-I, Svensson P (1993) Auralization – an overview. J Audio Eng Soc 41:861
- Kleiner M, Granier E, Svensson P (1995) Coupling of low and high frequency models in auralization. In: Proceedings of the 15th ICA, Trondheim, Norway, p 533
- Klemenz M (2005) Sound synthesis of starting electric railbound vehicles and the influence of consonance on sound quality. Acta Acust Acust 91:779
- Kob M (2002) Physical modelling of the singing voice. Doctoral thesis, RWTH Aachen University, Germany
- Kohlrausch AG, van de Par S, van Eijk R, Juola JF (2006) Human performance in detecting audio-visual asynchrony. J Acoust Soc Am 120:3084
- Kouyoumjian RG, Pathak PH (1974) A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface. Proc IEEE 62:1448
- Krämer P (1995) Künstliche räumliche Verhallung auf der Grundlage eines physikalischen Simulationsmodells. Doctoral thesis, Technical University Berlin, Germany
- Krijnen, T, Hornikx M (2014) open PSTD: the open source implementation of the pseudo spectral time-domain method. In: Proceedings of the forum Acusticum Krakow, Poland
- Krijnen, TF, Beetz J, Voorthuis JCT, de Vries B (2012). Explauralisation: the experience of exploring architecture made audible. In: Proceedings of the 30th international conference on education and research in computer aided architectural design in Europe, Prague, Czech Republic, p 593
- Krokstad A, Strøm S, Sørdsdal S (1968) Calculating the acoustical room response by the use of a ray-tracing technique. J Sound Vib 8:118
- Kropp W, Sabiniarz P, Brick H, Beckenbauer T (2012) On the sound radiation of a rolling tyre. J Sound Vib 331(8):1789
- Kropp W, Forssén J, Estévez-Mauriz L (2016) Urban sound planning – the SONORUS project. Chalmers University, Gothenburg, Sweden. http://publications.lib.chalmers.se/records/fulltext/242257/local_242257.pdf
- Kuhlen TW, Hentschel B (2014) Quo vadis CAVE – does immersive visualization still matter? IEEE Comput Graph Appl J 34(5):14
- Kulkarni A, Colburn HS (1995) Efficient finite-impulse-response filter models of head-related transfer functions. J Acoust Soc Am 97:3278
- Kulkarni A, Colburn HS (1998) Role of spectral detail in sound-source localization. Nature 396:747
- Kulowski A (1985) Algorithmic representation of the ray tracing technique. Appl Acoust 18:449
- Kuttruff H (1971) Simulierte Nachhallkurven in Rechteckräumen mit diffusem Schallfeld. Acustica 25:333
- Kuttruff H (1985) Stationäre Schallausbreitung in Flachräumen. Acustica 57:62
- Kuttruff H (1993) Some remarks on the simulation of sound reflection from curved walls. Acustica 77:176
- Kuttruff H (1995) A simple iteration scheme for the computation of decay constants in enclosures with diffusely reflection boundaries. J Acoust Soc Am 98:288
- Kuttruff H (2007) Acoustics – an introduction. Taylor & Francis, London
- Kuttruff H (2016) Room acoustics, 6th edn. CRC Press, London

- Kuttruff H, Thiele R (1954) Über die Frequenzabhängigkeit des Schalldrucks in Räumen. *Acustica* 4:614
- Lake Huron.htm (2005). <http://www.lake.com.au>. Version 2005
- Lam YW, Hargreaves J (2012) Time domain modelling of room acoustics. In: Proceedings of the acoustics 2012, April 2012, Nantes, France, 2012
- Langer S, Schanz M (2008) Time domain boundary element method. In: Marburg S, Nolte B (eds) Computational acoustics of noise propagation in fluids – finite and boundary element methods. Springer, Berlin, Heidelberg
- LaViola JJ (2003) A testbed for studying and choosing predictive tracking algorithms in virtual environments. In: Proceedings of the 7th international immersive technology workshop
- Lehnert H (1992) Binaurale Raumsimulation: Ein Computermodell zur Erzeugung virtueller auditiver Umgebungen. Doctoral thesis, Ruhr-University Bochum, Germany
- Lehnert H, Blauert J (1989) A concept for binaural room simulation. In: Proceedings of IEEE workshop on application of signal processing to audio & acoustics, New Paltz, NY
- Lehnert H, Blauert J (1992) Principles of binaural room simulation. *Appl Acoust* 36:259
- Lentz T (2007) Binaural technology for virtual reality. Doctoral thesis, RWTH Aachen University, Germany
- Lentz T, Behler G (2004) Dynamic cross-talk cancellation for binaural synthesis in virtual reality environments. In: Proceedings of 117th convention audio engineering society, San Francisco
- Lentz T, Schröder D, Vorländer M, Assenmacher I (2007) Virtual reality system with integrated sound field simulation and reproduction. In: EURASIP J Appl Sig Process, Special Issue on Spatial Sound and Virtual Acoustics
- Lewers T (1993) A combined beam tracing and radiant exchange computer model of room acoustics. *Appl Acoust* 38:161
- Lievens M Brunskog J (2007) Model of a person walking as a structure borne sound source. In: Proceedings of 19th ICA Madrid
- Lindau A, Hohn T, Weinzierl S (2007) Binaural resynthesis for comparative studies of acoustical environments. In: Proceedings of the 122th AES convention, Vienna
- Lokki T (2002) Physically-based auralization -- design, implementation, and evaluation. Doctoral thesis, Helsinki University of Technology
- Lüke H-D (1999) Signalübertragung, 7th edn. Springer, Berlin
- Lyamshev LM (1959) A question in connection with the principle of reciprocity in acoustics. Sov Phys Dokl 4:406
- Lyon R (1975) Statistical enegy analysis of dynamical systems. MIT press, Cambridge, MA
- Mackensen P (2003) Auditive localization. Head movements, an additional cue in localization. Doctoral thesis, Technical University Berlin, Germany
- Maekawa Z (1968) Noise reduction by screens. *Appl Acoust* 1:157
- Maempel H-J, Horn M (2017) The Virtual Concert Hall – a research tool for the experimental investigation of audiovisual room perception. *Int J Stereo Immersive Media* 1(1):78
- Magalhães MBS, Tenenbaum RA (2004) Sound sources reconstruction techniques: a review of their evolution and new trends. *Acta Acust Acust* 90:199
- Maillard J, Kacem A, Martin N, Faure B (2019) Physically-based auralization of rolling noise. In: Proceedings of the 23rd ICA Aachen 2019, Germany, p 1667
- Majdak P, Balazs P, Laback B (2007) Multiple exponential sweep method for fast measurement of head-related transfer functions. *J Audio Eng Soc* 55:623
- Marburg S, Schneider S (2003) Performance of iterative solvers for acoustic problems. Part I. Solvers and effect of diagonal preconditioning. *Eng Anal Bound Elem* 27:727
- Martin V, Guignard T (2006) Image-source method and truncation of a series expansion of the integral solution – case of an angular sector in two dimensions. *J Acoust Soc Am* 120:597
- Masiero B (2013) Individualized binaural technology. Doctoral thesis RWTH Aachen University, Germany
- Masiero B, Pollow M, Fels J (2011) Design of a fast broadband individual head-related transfer function measurement system. In: Proceedings of the forum acusticum, Aalborg, Denmark

- McGrath D (1996) Real-time auralization with the Huron digital audio convolution workstation. *J Acoust Soc Am* 100:2579
- Mechel FP (2002) Improved mirror source method in roomacoustics. *J Sound Vib* 256:873
- Mechel FP (2005) The cat's eye model. *Acta Acust Acust* 91:653
- Mechel FP et al (2002) Formulas of acoustics. Springer, Berlin Heidelberg/New York
- Meesawat K, Hammershøi D (2003) The time when the reverberant tail in binaural room impulse response begins. 115th AES Convention, New York, Preprint 5859
- Medwin H (1981) Shadowing by finite noise barriers. *J Acoust Soc Am* 69:1060
- Meier A (2000) Die Bedeutung des Verlustfaktors bei der Bestimmung der Schalldämmung im Prüfstand. Doctoral thesis, RWTH Aachen University, Germany
- Meyer J (1995) Akustik und musikalische Aufführungspraxis. Verlag Erwin Bochinsky, 2. Auflage, Frankfurt
- Minaar P, Plogsties J, Christensen F (2005) Directional resolution of head-related transfer functions required in binaural synthesis. *J Audio Eng Soc* 53:919
- Miwa T (1975) Mechanical impedance of the human body in various postures. *Ind Health* 13:1
- Moldrzyk C, Ahnert W, Feistel S, Lentz T, Weinzierl S (2004) Head-tracked auralization of acoustical simulation. In: 117th AES Convention, San Francisco, Preprint 6275
- Möller H (1992) Fundamentals of binaural technology. *Appl Acoust* 36:171
- Mommertz E (1996) Untersuchung akustischer Wandeigenschaften und Modellierung der Schallrückwürfe in der binauralen Raumsimulation. Doctoral thesis RWTH Aachen University, Germany
- Moore BCJ, Glasberg BR (1996) A revision of Zwicker's loudness model. *Acust Acta Acust* 82:335
- Morse PM, Ingard KU (1968) Theoretical acoustics. McGraw-Hill, New York
- Möser M (2004) Engineering acoustics – an introduction to noise control. Springer, Berlin Heidelberg/New York
- Mourjopoulos J (1994) Digital equalization of room acoustics. *J Audio Eng Soc* 42:884
- Naef M, Staadt O, Gross M (2002) Spatialized audio rendering for immersive virtual environments. In: Proceedings of the ACM symposium on virtual reality software and technology, Hong Kong, China, p 65
- Naylor GM (1993) ODEON – another hybrid room acoustical model. *Appl Acoust* 38:131
- Noisternig M, Katz B, Siltanen S, Savioja L (2008) Framework for real-time auralization in architectural acoustics. *Acta Acust Acust* 94:1000
- Ochmann M (1990) Die Multipolstrahlersynthese – ein effektives Verfahren zur Berechnung der Schallabstrahlung von schwingenden Strukturen beliebiger Oberflächengestalt. *Acustica* 72:233
- Ondet AM, Barby JL (1988) Sound propagation in fitted rooms – comparison of different models. *J Sound Vib* 125:137
- Ondet AM, Barby JL (1989) Modelling of sound propagation in fitted workshops using ray tracing. *J Acoust Soc Am* 85:787
- OPENAL (2006). <http://www.openal.org>
- Oppenheim AV, Schafer RW (1989) Discrete-time signal processing, Prentice Hall signal processing series. Prentice Hall, Upper Saddle River
- Ostashev VE (1997) Acoustics in moving inhomogenous media. E&FN SPON, London
- Otondo F, Rindel JH (2005) A new method for the radiation representation of musical instruments in auralizations. *Acta Acust Acust* 91:902
- Otsuru T, Sakuma T, Sakamoto S (2005) Constructing a database of computational methods for environmental acoustics. *Acoustical Sci. & Tech.* 26(2):221
- Papoulis A (1981) Signal Analysis. McGraw-Hill, New York
- Pätynen J, Lokki T (2010) Directivities of symphony orchestra instruments. *Acta Acust Acust* 96:138
- Pelzer S, Aspöck L, Schröder D, Vorländer M (2014) Integrating real-time room acoustics simulation into a CAD modeling software to enhance the architectural design process. *Buildings* 2:113

- Petersson BAT, Gibbs BM (2000) Towards a structure-borne sound source characterization. *Appl Acoust* 61:325
- Petersson BAT, Plunt J (1982) On effective mobilities in the prediction of structure-borne sound transmission between a source and a receiving structure, part 1: theoretical background and basic experimental studies procedures. *J. Sound Vib.* 82, 517; part 2: estimation of mobilities. *J Sound Vib* 82:531
- Picaut J, Fortin N (2012) I-Simpa, a graphical user interface devoted to host 3D sound propagation numerical codes. In: Proceedings of the CFA/IOA acoustics 2012, Nantes, France
- Piercy JE, Embleton TFW, Sutherland LC (1977) Review of sound propagation in the atmosphere. *J Acoust Soc Am* 61:1403
- Pieren R (2018) Auralization of environmental acoustical sceneries – synthesis of road traffic, railway and wind turbine noise. Doctoral thesis, TU Delft, The Netherlands
- Pieren R, Heutschi K, Wunderli JM, Snellen M, Simons DG (2017) Auralization of railway noise: emission synthesis of rolling and impact noise. *Appl Acoust* 127:34
- Pohl A (2014) Simulation of diffraction based on the uncertainty relation. Doctoral thesis, HafenCity University Hamburg, Germany
- Poirier-Quinot D, Katz B, Noisternig M (2017) EVERTims: open source framework for real-time auralization in architectural acoustics and virtual reality. In: Proceedings of the 20th international conference on digital audio effects (DAFx-17), Edinburgh, UK
- Pösselt C, Schroeter J, Opitz H, Divenyi P, Blauert J (1986) Generation of binaural signals and home entertainment. In: Proceedings of the 12th ICA, Toronto
- Pulkki V (1997) Virtual sound source positioning using vector base amplitude panning. *J Audio Eng Soc* 45:456
- Reichardt W, Abdel Alim O, Schmidt W (1974) Abhängigkeit der Grenzen zwischen brauchbarer und unbrauchbarer Durchsichtigkeit von der Art des Musikmotives, der Nachhallzeit und der Nachhalleinsatzzeit. *Appl Acoust* 7:243
- Richter J-G, Fels J (2019) On the influence of continuous subject rotation during high-resolution head-related transfer function measurements. *IEEE/ACM Trans Audio, Speech Lang Process* 27 (4):730
- Rindel JH (1993) Modelling the angle-dependent pressure reflection factor. *Appl Acoust* 38:223
- Rindel JH (2006) Auralization of airborne sound insulation including the influence of source room. In: Proceedings of the EURONOISE '07, Tampere, Finland
- Rindel JH, Christensen CL (2003) Room acoustic simulation and auralization – how close can we get to the real room? In: Proceedings of the WESPAC8, Melbourne, Australia
- Rindel JH, Otondo F, Christensen F (2004) Sound source representation for auralization. In: Proceedings of the international symposium on room acoustics – design and science, Awaji, Japan
- Römer B (2004) Simulation der Schallabstrahlung von Kfz-Motoren (Simulation of sound radiation from combustion engines). Diploma (MSc) thesis RWTH Aachen University, Germany
- Saarela J, Savioja L (2019) Audibility of dispersion error in room acoustic finite difference time-domain simulation in the presence of a single early reflection. *J Acoust Soc Am* 145:2761
- Sahai A, Wefers F, Pick S, Stumpf E, Vorländer M, Kuhlen T (2016) Interactive simulation of aircraft noise in aural and visual virtual environments. *Appl Acoust* 101:24
- Sakuma T, Yasuda Y (2002) Fast multipole boundary element method for large-scale steady-state sound field analysis. Part I: setup and validation. *Acta Acust Acust* 88:513
- Salomons E, van Maercke D, Defrance J, de Roo F (2011) The Harmonoise Sound Propagation Model. *Acta Acust Acust* 97:62
- Salomons EM (2001) Computational atmospheric acoustics. Springer, Dordrecht
- Savioja L (1999) Modelling techniques for virtual acoustics. Doctoral thesis, Helsinki University of Technology, Finland
- Savioja L (2010) Real-time 3D finite-difference time-domain simulation of mid-frequency room acoustics. In: Proceedings of the 13th DAFX Graz, Austria, p 43

- Savioja L, Svensson UP (2015) Overview of geometrical room acoustic modeling techniques. *J Acoust Soc Am* 138:708
- Savioja L, Välimäki V (2000) Reducing the dispersion error in the digital waveguide mesh using interpolation and frequency-warping techniques. *IEEE Trans Speech Audio Process* 8:184
- Savioja S, Huopaniemi J, Lokki T, Väänänen R (1999) Creating interactive virtual acoustic environments. *J Audio Eng Soc* 47:675
- Schmitz A (1994) Naturgetreue Wiedergabe kopfbezogener Schallaufnahmen über zwei Lautsprecher mit Hilfe eines Übersprechkompensators. Doctoral thesis, RWTH Aachen University, Germany
- Schmitz A (1995) Ein neues digitales Kunstkopfmesssystem. *Acustica* 81:416
- Schmitz O, Feistel S, Ahnert W, Vorländer M (2001) Grundlagen raumakustischer Rechenverfahren und ihre Validierung. In: Proceedings of the DAGA '01, Hamburg-Harburg, p 24
- Scholl W (2001) Impact sound insulation: the standard tapping machine shall learn to walk! *Build Acoust* 8:245
- Schröder D (2011) Physically based real-time auralization of interactive virtual environments. Doctoral thesis, RWTH Aachen University, Germany
- Schröder D, Lentz T (2006) Real-time processing of image sources using binary space partitioning. *J Audio Eng Soc* 54:604
- Schröder D, Dross P, Vorländer M (2007) A fast reverberation estimator for virtual environments. In: Proceedings of the 30th AES conference, Saariselkä, Finland
- Schroeder MR (1954) Die statistischen Parameter der Frequenzkurven von großen Räumen. *Acustica* 4:594
- Schroeder MR (1961) Improved quasi-stereophony and colourless artificial reverberation. *J Acoust Soc Am* 33:1061
- Schroeder MR (1965) New method for measuring reverberation time. *J Acoust Soc Am* 37:409
- Schroeder MR (1973) Computer models for concert hall acoustics. *Am J Phys* 41:461
- Schroeder MR, Kuttruff KH (1962) 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
- Schroeder MR, Atal BS, Bird C (1962) Digital computers in room acoustics. In: Proceedings of 4th ICA, Copenhagen, M21
- Sellerbeck P (2003) Untersuchungen zur reziproken Transferfadanalyse im Fahrzeug mit Hilfe eines binauralen Schallsenders. Diploma (MSc thesis) RWTH Aachen University and HEAD acoustics, Aachen, Germany
- Shabtai N, Vorländer M (2015) Acoustic centering of sources with high-order radiation patterns. *J Acoust Soc Am* 137:1947
- Shabtai N, Behler G, Vorländer M, Weinzierl S (2017) Generation and analysis of an acoustic radiation pattern database for forty-one musical instruments. *J Acoust Soc Am* 141:1246
- Shaw EAG (1982) External ear response and sound localization. In: Gatehouse W (ed) Localization of sound: theory and applications. Amphora, Groton, p 30
- Shumacker R, Brand R, Gilliland M, Sharp W (1969) Study for applying computer-generated images to visual simulations. Report AFHRL-TR-69-14, U.S. Air Force Human Resources Laboratory
- Siebrasse KF (1973) Vergleichende subjektive Untersuchungen zur Akustik von Konzertsälen. Doctoral Thesis, Göttingen, Germany
- Skudrzyk E (1971) The foundations of acoustics. Springer-Verlag, Wien New York
- Sottek R (2003) Virtual binaural auralisation of product sound quality: importance and application in practice. In: Proceedings of the EURONOISE Naples
- Sottek R (2004) Design of vehicle interior sound using a noise synthesis technology. In: Proceedings of the internoise, Prague

- Sottek R, Müller-Held B (2007) Binaural transfer path analysis and synthesis (BTPA/BTPS) using substructuring techniques based on finite element analysis (FEA) and measurements. In: Proceedings of the SAE noise & vibration conference, Detroit, paper 2007-01-2226
- Sottek R, Riemann D, Sellerbeck P (2004) Virtual binaural auralisation of vehicle interior sounds. In: Proceedings of the joint congress CFA/DAGA'04, Strasbourg
- Sottek R, Behler G, Kellert T (2005) Beschreibung der breitbandigen Körper- und Luftschallausbreitung aus dem Powertrain – Binaurale Transferpfadanalyse und –synthese. FFV Frankfurt – Final report, www.head-acoustics.de
- Spandöck F (1934) Akustische Modellversuche. Annalen der Physik V 20:345
- Spandöck F (1965) Die Vorausbestimmung der Akustik eines Raumes mit Hilfe von Modellversuchen. In: Proceedings of 5th ICA Vol II, Liège, p 313
- Spors S, Teutsch H, Kuntz A, Rabenstein R (2004) Sound field synthesis. In: Huang Y, Benesty J (eds) Audio signal processing for next-generation multimedia communication systems. Kluwer Academic Publishers, New York
- Stephenson UM (1985) Eine Schallteilchen-Computersimulation zur Berechnung für die Hörsamkeit in Konzertsälen massgebenden Parameter. *Acustica* 59:1
- Stephenson UM (1996) Quantized beam tracing – a new algorithm for room acoustics and noise immission prognosis. *Acust Acta Acust* 82:517
- Stephenson UM (2004a) Derivation of the reduction of computation time by the voxel crossing technique. In: Proceedings of the CFA/DAGA '04, Strasbourg
- Stephenson UM (2004b) Beugungssimulation ohne Rechenzeitexplosion: Die Methode der quantisierten Pyramidenstrahlen – ein neues Berechnungsverfahren für Raumakustik und Lärmimmissionsprognose. Doctoral thesis, RWTH Aachen University, Germany
- Stephenson UM (2006) Analytical derivation of a formula for the reduction of computation time by the voxel crossing technique used in room acoustical simulation. *Appl Acoust* 76:959
- Stephenson UM, Svensson UP (2007) Can also sound be handled as stream of particles? – an improved energetic approach to diffraction-based uncertainty principle – from ray to beam tracing. In: Proceedings of DAGA '07, Stuttgart
- Stienen J, Vorländer M (2019) Progressive region-of-interest filtering for urban sound auralization applications with multiple reflected and diffracted propagation paths. In: Proceedings of the 23rd ICA 2019 Aachen, Germany
- Storms R (1995) Npsnet-3d sound server: an effective use of the auditory channel, Naval Post-graduate School, Monterey CA, MSc thesis
- Strauss H (1998) H. Strauss. implementing Doppler Shifts for virtual auditory environments. 104th convention audio engineering society, Amsterdam, The Netherlands
- Suh JS, Nelson PA (1999) Measurement of transient response of rooms and comparison with geometrical acoustic models. *J Acoust Soc Am* 105:2304
- Summers J, Takahashi K, Shimizu Y, Yamakawa T (2004) Assessing the accuracy of auralizations computed using a hybrid geometrical acoustics and wave-acoustics method. *J Acoust Soc Am* 115:2514
- Svensson P, Fred RI, Vanderkooy J (1999) Analytic secondary source model of edge diffraction impulse responses. *J Acoust Soc Am* 106:2331
- Tachibana H, Tanaka H, Kimura S (1998) Development of new heavy and soft impact source for the assessment of floor impact sound of building. In: Proceedings of the internoise '98, Christchurch, New Zealand
- Terhardt E (1974) Pitch, consonance, and harmony. *J Acoust Soc Am* 55:1061
- Thaden R (2005) Auralisation in building acoustics. Doctoral thesis, RWTH Aachen University, Germany
- Torres RR, Kleiner M, Dalenbäck B-I (2002) Audibility of “diffusion” in room acoustics auralization: an initial investigation. *Acust Acta Acust* 86:919
- Torres JCB, Petraglia MR, Tenenbaum RA (2004) An efficient wavelet-based HRTF model for auralization. *Acta Acust Acust* 90:108

- Tsingos N, Funkhouser T, Ngan A, Carlbom I (2001) Modelling acoustics in virtual environments using the uniform theory of diffraction. In: Proceedings of the SIGGRAPH, Computer Graphics
- Turk G, Levoy M (1994) Zippered polygon meshes from range images. In: Proceedings of the SIGGRAPH, Computer Graphics, p 311
- Välimäki V (1995) Discrete-time modelling of acoustic tubes using fractional delay filters. Doctoral thesis, Helsinki University of Technology, Finland
- Välimäki V, Karjalainen M, Laakso T (1993) Fractional delay digital filters. In: Proceedings of the IEEE international symposium on circuits and systems, Chicago, Illinois, p 355
- van der Giet M (2011) Analysis of electromagnetic acoustic noise excitations. Doctoral thesis, RWTH Aachen University
- van der Harten A (2013) Pachyderm acoustical simulation: towards open-source sound analysis. Archit Des 83:138
- van Maercke D (1986) Simulation of sound fields in time and frequency domain using a geometrical model. In: Proceedings of 12th ICA Toronto, E11-7
- VDI3760 – VDI directive (1996) Berechnung und Messung der Schallausbreitung in Arbeitsräumen (Computation and measurement of sound propagation in workrooms) Association of German engineers
- Vercammen M (2010) Sound concentration caused by curves surfaces. Doctoral thesis, TU Eindhoven, The Netherlands
- Vian J-P, Martin J (1992) Binaural room acoustics simulation: practical uses and applications. Appl Acoust 36:293
- Vian J-P, van Maercke D (1986) Calculation of the room impulse response using a ray-tracing method. In: Proceedings of ICA symposium on acoustics and theatre planning for the performing arts, Vancouver, Canada, p 74
- Vigeant MC, Wang LM, Rindel JH (2007) Investigations of multi-channel auralization technique for solo instruments and orchestra. In: Proceedings of the 19th ICA, Madrid 2007, Spain
- Vorländer M (1988) Ein Strahlverfolgungsverfahren zur Berechnung von Schallfeldern in Räumen. Acustica 65:138
- Vorländer M (1989) 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
- Vorländer M (1995) International round robin on room acoustical computer simulations. In: Proceedings of the 15th ICA Trondheim 1995, Norway, p 689
- Vorländer M (2000) A fast room acoustical simulation algorithm based on the free path distribution. J Sound Vib 232:129
- Vorländer M (2013) Computer simulations in room acoustics: concepts and uncertainties. J Acoust Soc Am 133:1203
- Vorländer M, Mommentz E (2000) Definition and measurement of random-incidence scattering coefficients. Appl Acoust 60:187
- Vorländer M, Thaden R (2000) Auralisation of airborne sound insulation in buildings. Acust Acta Acust 86:70
- Vorländer M, Schröder D, Pelzer S, Wefers F (2015) Virtual reality for architectural acoustics. J Build Perform Simul 8:15
- Wallace CE (1972) Radiation resistance of a rectangular plate. J Acoust Soc Am 51:946
- Watters BG (1965) Impact-noise characteristics of female hard-heeled foot traffic. J Acoust Soc Am 37:619
- Wefers F (2014) Partitioned convolution algorithms for real-time auralization. Doctoral thesis, RWTH Aachen University, Germany
- Wefers F, Vorländer M (2011) Optimal filter partitions for real-time FIR filtering using uniformly-partitioned FFT-based convolution in the frequency-domain. In: Proceedings of the 14th international conference on digital audio effects, DAFX-11, Paris
- Wefers F, Vorländer M (2014) Uniformly-partitioned convolution with independent partitions in signal and filter. In: Proceedings of AES 137th Convention, Los Angeles

- Wefers F, Vorländer M (2015) Strategies for the real-time auralization of fast moving sound sources in interactive virtual environments. In: Proceedings of the internoise 2015, San Francisco
- Wefers F, Vorländer M (2018) Flexible data structures for dynamic virtual auditory scenes. Virtual Reality. Springer. <https://doi.org/10.1007/s10055-018-0332-9>
- Whitaker RW, Norris DE (2008) Infrasound propagation, in handbook of signal processing in acoustics. Springer, New York
- Wightman FL, Kistler DJ (1989) Headphone simulation of free-field listening I: stimulus synthesis, II: psychophysical validation. *J Acoust Soc Am* 85:858. (part I) and 868 (part II)
- Wightman FL, Kistler DJ (1999) Resolution of front-back ambiguity in spatial hearing by listener and source movement. *J Acoust Soc Am* 105:2841
- Williams E (1999) Fourier acoustics: sound radiation and near-field acoustical holography. Elsevier, New York
- Xie B (2013) Head-related transfer function and virtual auditory display, 2nd edn. J. Ross Publishing, New York
- Xu S, Li Z, Salvendy G (2008) Individualized head-related transfer functions based on population grouping. *J Acoust Soc Am* 124:2708
- Yasuda Y, Sakuma T (2003) Fast multipole boundary element method for large-scale steady-state sound field analysis. Part II: examination of numerical items. *Acta Acust Acust* 89:28
- Zienkiewicz OC (1977) The finite element method, 3rd edn. McGraw-Hill, London
- Zotter F (2009) Analysis and synthesis of sound-radiation with spherical arrays. Doctoral thesis, Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, Austria
- Zwicker E, Fastl H (1999) Psychoacoustics, fact and models, 2nd edn. Springer, Berlin/Heidelberg/New York

Index

A

A/D converter, 108
Absorption
 in ray tracing, 179
Absorption coefficient, 33, 175, 219
 data, 333
Acoustic materials, 333
Acoustic virtual reality, 297
Air gap, 35
Airborne sound
 modelling, 244–246
Airborne sound transmission, 236–237
Aliasing, 109
Anechoic chamber, 123
Artificial ear, 300
Attenuation coefficient, 47
Auditory source width, 94
Auralization, 100, 141
 airborne sound insulation, 246–253
 BTPS, 273–275
 building acoustics, 235
 impact sound, 259–261
 real-time, 283, 285–289
 room acoustics, 172
 software, 215
 sound insulation, 247
 structure-borne sound, 255
 vehicle acoustics, 267
 vibration source, 261
Automotive engineering, 267
A-weighting, 78

B

Beam tracing, 204
Bending stiffness, 70, 132

Bending wave, 69, 239
Binary space partitioning (BSP), 202
Binaural
 hearing, 76, 82
 rendering, 141
 synthesis, 140, 141, 315
 technology, 315
 transfer path synthesis (BTPS), 265, 268
Boundary condition, 54
Boundary element method, 151
Building acoustics, 235
Building elements, 238–240

C

CAD model, 172
Cavity, 52
Center time, 93, 97
Character, 75
Characteristic impedance, 8
Clarity, 93, 97
Cochlea, 77
Combustion engine, 268
Computer simulation
 room acoustics, 171
Cone tracing, 204
Convolution, 103, 135, 142
 discrete, 136
 FFT, 135, 138
 filter, 135
 segmented, 139
Crosstalk, 317
 cancellation, 317
 cancellation, dynamic, 318, 319
 filter, 318
Curved wall, 184

D

Decay curve, 60
 Decibel, 12
 Definition, 93, 97
 Diffraction, 42, 200
 Diffuse field, 57
 Digital filter, 115
 FIR, 117
 IIR, 117
 Dipole source, 22
 Dirac pulse, 104, 159
 Direct sound, 62, 89
 Directivity, 24
 Directivity factor, 24
 Discretization, 108
 Dispersion, 70
 Displacement, 3
 Doppler effect, 47
 Double wall, 240
 Dummy head, 85
 reciprocal, 272

E

Ear canal, 76
 Eardrum impedance, 300
 Early Decay Time (EDT), 91
 Early reflections, 90
 Eigen frequency, 53
 Elastic constants, 67
 Electrical equivalent, 131
 Energy density, 12, 92
 Energy time curve, 89
 Equal loudness level contours, 78
 Equivalent absorption area, 60
 Equivalent circuit, 164
 Eyring's equation, 59

F

Feedback, 271
 Fermat's principle, 45
 Filter
 adaptive, 281, 319
 impact sound, 259
 sound insulation, 247
 update, 280
 Finite element method, 151
 Floor
 construction, 258
 covering, 256
 impedance, 256, 262
 lightweight, 261, 263

massive, 258
 mobility, 261
 wooden, 256

Fluctuation strength, 81
 Fourier synthesis, 10
 Fourier transformation, 9, 106, 109
 discrete (DFT), 110
 fast (FFT), 112
 Frequency domain, 109
 Frequency resolution, 157
 Frequency spacing, 55
 Front-back confusion, 88, 316

G

Geometrical acoustics, 162, 163, 171
 uncertainties, 207
 Grazing incidence, 199
 Green's function, 146

H

Harmonic wave, 9
 Head rotation, 279
 Head tracker, 281
 Headphone, 280, 298, 301
 binaural reproduction, 301
 diffuse-field, 302
 equalization, 299
 free-field, 302
 Head-related impulse response (HRIR), 140
 Head-related transfer function (HRTF), 82, 83,
 140, 141, 143, 218
 dynamic cues, 281
 individual, 302
 near-field, 143
 Head-tracker, 280
 Hearing threshold, 10
 Helmholtz equation, 53, 151
 Helmholtz-Huygens integral, 147, 153, 313
 Helmholtz-Kirchhoff integral, 153, 313
 Histogram, 183
 Human hearing, 75
 Human voice, 125

I

Image source, 38, 172
 audibility, 196
 construction, 196
 preprocessing, 202
 real-time, 284
 Impact sound, 255–258

- contact force, 263
modelling, 256
Impact sound level, 256
Impedance, 33
Impedance plane, 37
Impulse response, 103, 135
 binaural, 223
 room, 89
 sound insulation, 249
 spatial, 221
In-head localization, 299
Inner ear, 77
Integrated impulse response, 91
Interactive, 277
Interaural cross correlation function, 94
Interaural cross-correlation coefficient, 95
interaural level difference (ILD), 82
Interaural time difference (ITD), 82
Interface problem, 131
- J**
Junction, 73, 245
Just noticeable difference, 95
- L**
Lambert's law, 43, 163, 181
Laplace operator, 6
Lateral reflections, 93
Linear distortion, 105
Linear time-invariant system, 102
Listener envelopment, 94
Localization, 76, 143
Localization blur, 84
Longitudinal wave, 68
Loss factor, 73, 246
Loudness, 78
Loudspeaker
 reproduction, 305
LTI system, 103
- M**
Masking, 79
Mesh, 152
Middle ear, 76
Modal analysis, 155
Modal overlap, 54
Modal response, 156
Modal superposition, 155, 156
Mode, 53
Modulation transfer function (MTF), 93
- Monopole, 17
Moving sources, 292–295
Multimodal, 277
Multipole source, 22
Multipole synthesis, 29
- N**
Numerical model, 152
- O**
Octave band, 15
One-third octave band, 15
Outer ear, 76
- P**
Particle velocity, 3
Partition wall, 238
Peripheral hearing organ, 75
Phantom source, 305
Pitch, 81, 82
Pitch strength, 82
Plane waves, 7, 8
Point contact, 131, 256
 multiple, 132
Point mobility, 132
Point source, 18
Poisson distribution, 187
Poisson ratio, 68
Polygon, 175
Post-masking, 79
Propagation, 145
Psychoacoustics, 75
Pyramid tracing, 204
- Q**
Quality, 75
Quantization, 108
- R**
Radiation efficiency, 72, 258
Radiation impedance, 20, 120
Radiation pattern, 124
Radiosity, 163
Ray, 177
Ray tracing, 172, 177–194
 deterministic, 204
 hybrid, 204
 stochastic, 204

- Ray tracing (*cont.*)
 uncertainties, 186–192
- Reciprocity, 148, 162, 166
 vibro-acoustic, 272
- Recording
 anechoic, 126
 orchestra, 127, 129, 130
- Rectangular plate, 72
- Reflection, 32
- Reflection factor, 32, 195
- Resonance frequency, 36
- Reverberance, 90
- Reverberation, 59
 digital, 288
 real-time, 287–289
- Reverberation time, 90
- Room acoustics
 auralization, 219, 290, 291
 hybrid simulation models, 210–216
 rendering, 319
 software, 211
- Room impulse response, 222
 binaural, 223, 224
- Root mean square, 12
- Rough surface, 40
- Round robin
 computer simulation, 208
- S**
- Sabine's equation, 59
- Sampling, 109
- Sampling frequency, 108
- Sampling theorem, 110
- Scattering, 210
- Scattering coefficient, 41, 177
 data, 333
- Scattering cross section, 39
- Schroeder frequency, 56
- Sharpness, 80
- Signal, 102
- Signal processing, 102, 136, 218
 auralization, 218
 binaural, 284
 real-time, 278
 sound insulation, 247
- Simulation models, 146
- Snell's law, 32
- Sound
 airborne path, 167
 apparent s. reduction index, 242
 direct s. transmission, 243
 energy, 12
- flanking s. reduction index, 244
 flanking s. transmission, 243
 intensity, 13
 power, 19, 20
 pressure, 3
 pressure level, 13
 propagation, 31
 propagation curves, 63
 ray, 56
 reduction index, 237, 240
 reduction index data, 333
 rendering, 278
 structure-borne path, 167
 transmission in buildings, 235
 wave, 3
- Source
 impedance, 262
- Source signal, 119, 135
- Spaciousness, 93
- Spatial sound, 140
- Specular, 195
- Speech, 126
 intelligibility, 92
 transmission index (STI), 93
- Speed of sound, 7
- Spherical harmonics, 24, 29
- Spherical wave, 17, 18, 199–200
- Statistical energy analysis (SEA), 157, 244, 257
- Steady-state transfer function, 105
- Stereo dipole, 318
- Strength, 92, 96
- Structural reverberation time, 73, 246, 258
- Structure-borne sound, 67
- Structure-borne source, 130
- Surround sound, 305
- Sweet spot, 305
- T**
- Tapping machine, 130, 255
 force-time signal, 260
- Time domain, 109
- Tonality, 81
- Transfer functions, 105, 135, 148
- Transfer impedance, 148
- Transfer path
 airborne, 268
 characterization, 271–273
 structural, 271
- Transmission coefficient, 237
- Transversal wave, 68
- Two-port, 164–166, 266
 network, 167

- parameters, 167
transmission, 238
- V**
Vibration level difference, 74
Vibration reduction index, 74, 245
Virtual environment, 277
Virtual headphone, 315
Virtual reality (VR), 281, 323
Volume velocity, 18
- W**
Walking person, 259
Wall impedance, 32
Wave equation, 4–6
Wave field synthesis (WFS), 312, 313
Waveguide, 160
Window technique, 114
- Y**
Young's modulus, 68