RWTHedition

RWTH Aachen

Michael Vorländer

Auralization

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

First edition

Springer

Prof. Dr. Michael Vorländer
RWTH Aachen
Institut für Technische Akustik
Neustraße 50
52066 Aachen
Germany
mvo@akustik.rwth-aachen.de

Foreword

RWTH Aachen University is one of Germany's leading universities of technology with a long tradition of close collaboration with national and international industry and research centres.

The University was founded as a polytechnical institute by an industry initiative in 1870, in a then fringe area of the Prussian heavy industry. Re-established as the Institute of Technology of North Rhine-Westphalia after World War II, it has always shown a strong commitment to contribute to the advance of science and engineering and to respond to the increasing demands made by society, ultimately advancing the development of the economy at a national and international level.

In light of this, RWTH Aachen University and Springer-Verlag Publishing have recently signed a cooperative agreement on the publication of selected works of authors from the University within a special series titled "RWTH edition" and published in English. The present work comprises the first volume of this series, which has evolved from the long-standing, fruitful cooperation between the two partners with the explicit goal of fostering the sustainable dissemination of groundbreaking research results to the international scientific community, thereby initiating and facilitating scientific exchange worldwide. Such a step was deemed particularly meaningful considering the growing importance of interdisciplinary projects worldwide, which are based on innovative solution-finding and sound basic research.

As Rector of RWTH Aachen University, I would like to express my sincere gratitude to Springer-Verlag for its commitment and ongoing co-operation.

May this series arouse the interest of the scientific community and contribute to an outstanding and solid education of our students and graduates.

Sincerely,
Prof. Dr. rer. nat. Burkhard Rauhut

Preface

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 occurs 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 on a model scale so 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 technique can be considered 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 reinvented by Schroeder et al. (1962) at 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 that time, several improvements in modelling algorithms, in binaural processing and in reproduction technique have been made. Today, commercial software for room acoustical simulation is 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 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 a free field and in rooms, in Chaps. 7 to 9, an introduction is given to 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 placing more emphasis on acoustics in enclosures in Chap. 11, including room acoustics auralization in Sect. 11.6. (It is presented as a first example where the technique of auralization is applied.[1]) Chapters 12 and 13 deal with strategies of simulation and auralization for applications in architectural acoustics other than room acoustics: airborne and impact sound insulation and 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. 14. The final Chaps. 15 and 16 are focused on aspects of real-time processing and 3-D 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 third-year university students, readers 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

---

[1]   Historically, it was also first established in room acoustics.

research, the technique of auralization may be useful for creating 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.

Acknowledgments

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

Aachen, June 2007   *Michael Vorländer*

Content

Introduction

Acoustics – the science of sound – is a very broad discipline. It has many links to other natural and human sciences and to electrical, mechanical and civil engineering. The glue 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 in 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 are motivated from another area: psychoacoustics. Physical data are 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 science.

The 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 geometric 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 manifold. Accordingly, the methods and tools of acoustic engineering are, too,

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, AlCons	%	Speech intelligibility
Product Sound Quality	Loudness, roughness, tonality …	sone, asper, …	Acoustic comfort

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 regulations of sound and noise effects. Another aspect might be the historical development of international standards[2] in acoustics, describing acoustics in relation to easily measurable quantities such as 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-world problems of daily life in 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 such as size, colour, and brightness and maybe on

---

2   International Standard Organization, ISO.

information about objects or resolution of details. A painting reproduced according to this verbal or numerical description will never be exactly such as 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! In 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."

> Auralization is the technique of creating audible sound files from numerical (simulated, measured, or synthesized) data.

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 here: type of sound generation, direction of the event, movement of the source, listener movement, environment room (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 simulation 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 vibroacoustic 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 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. The core of this book is accordingly formed by Chaps. 7, "Signal processing for auralization," 8, "Characterization of sources," and 9, "Convolution of sound sysnthesis."

Toward 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 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. 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, sonification) is not present to the same extent and is added often just as an effect and without plausible or authentic reference to the virtual scene.

The process of generating the cues for the respective senses (3-D image, 3-D audio, …) 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). Thus 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 objects of daily use, furniture, tools to finally the dimensions of humans are in the range of metres and fractions of metres. Wave physics is most difficult (and most interesting) when wavelengths are of the same order of magnitude as the objects in the sound path. Neither approximations of small wavelengths (such as in

optics) nor large wavelengths (such as 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, although numerous approximations must be made to reach this goal. But in the end, 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, too.

During the interaction with real scenes several senses are stimulated. Acoustics, besides vision, haptics and tactile cognition, yields 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, etc., and finally reaches the receiver's ears. Humans as recipients evaluate the diverse characteristics of the total sound, segregate it into its individual objects, and furthermore, evaluate the environment itself, 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.

1   Fundamentals of acoustics

"Waves" are well-known to everybody, even when 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 Maxell's theory in 1864 and experimentally discovered by Heinrich Hertz in 1888. Their spectrum from gamma radiation, X-rays, ultraviolet, the visible range, infrared toward spectra for technical communication systems offers a fascinatingly wide area of natural phenomena and technical applications. Waves are to be considered 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.[3]
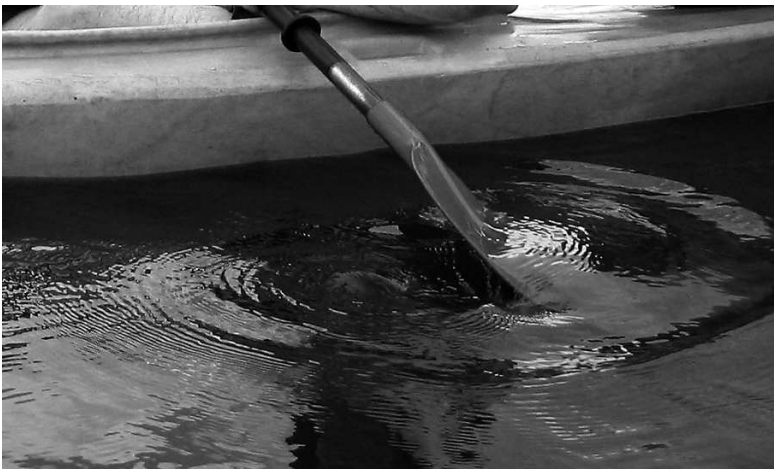
Fig. 1.1. Waves in water

---

[3] A row of dominos falling is an example where a kind of transport wave is observed without energetic equilibrium. Energy is not recycled. This effect is not a wave in the physical sense we discuss here!
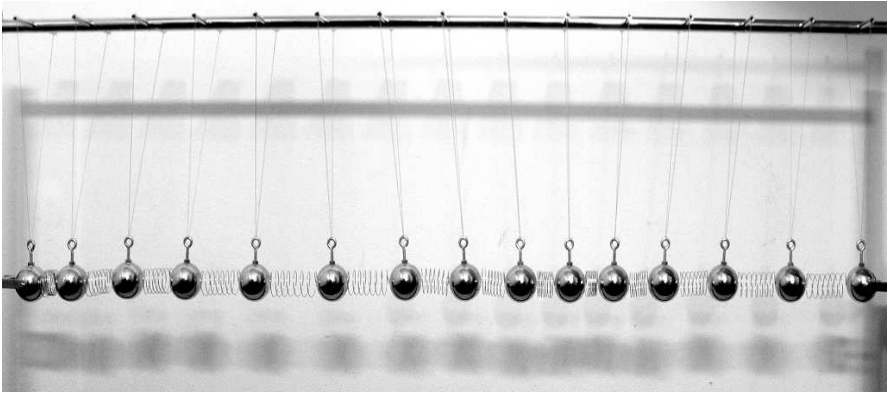
Fig. 1.2. Photograph of metal spheres connected by springs

Although water waves are possibly one of the most enjoyable examples, we now focus on a mechanical system to illustrate of 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 energy), recycles and transfers its energy to the next mass, and so on. By intuition, we can imagine easily that heavy masses with large inertia and soft springs with high flexibility provide this transport effect at 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 next neighbour 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.

1.1   Sound field equations and the 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, taken such as a photograph at a certain time, and we observe that the medium's molecules
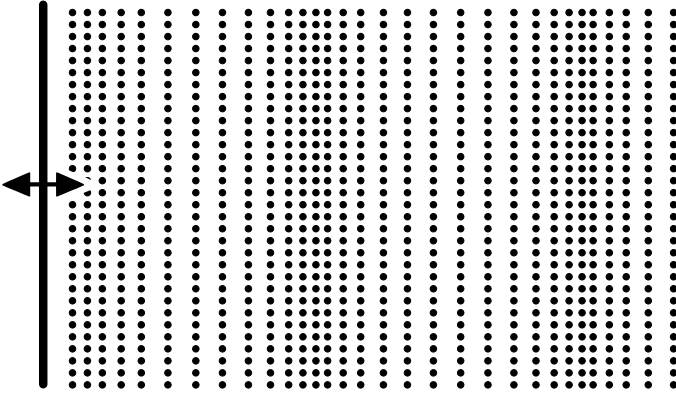
Fig. 1.3. Microscopic view of a medium with sound

or atoms are somehow displaced from their original position, of course, independent of the constant irregular thermal movement.

1.1.1 Sound field quantities

In a sound wave, the particles (gas molecules, crystal lattice atoms, etc.) follow a space- and time-dependent displacement vector, $\bar{s}$. The time derivative of this displacement, accordingly, is the particle velocity,

$$\bar{v} = \frac{\partial \bar{s}}{\partial t},\tag{1.1}$$

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

The displacements are neither homogeneous nor isotropic.[4] 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 at rest. The density fluctuations due to sound are then

$$\rho = \rho_{tot} - \rho_0,\tag{1.2}$$

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

$$p = p_{tot} - p_0.\tag{1.3}$$

---

[4] Homogenous and isotropic means independent of translation or rotation, respectively.

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 typically the leading 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.

In fluid media such as 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

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

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).

1.1.2   Derivation of the wave equation

We consider 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 periodically the volume element with a strength of $qS\mathrm{d}x$ (in units of $[\mathrm{m}^3/\mathrm{s}]$). It excites a small disturbance from the pressure equilibrium. The source can be assumed anywhere in the tube, for instance, 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,
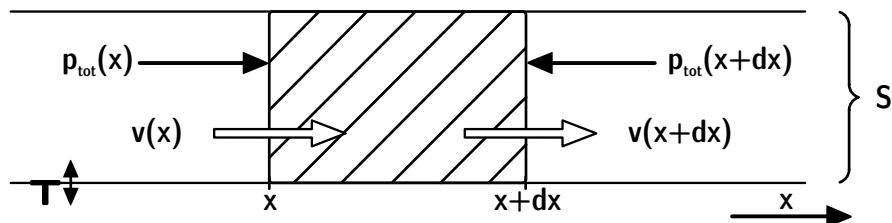
Fig. 1.4. Volume element in a one-dimensional fluid medium. Pressure variations are induced by a "pumping" source on the left side

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.

According to the Euler equation of all forces involved, we obtain a movement of medium mass:

$$\left[p_{\text{tot}}(x) - p_{\text{tot}}(x + dx) \right] \cdot S = \rho_{\text{tot}} S dx \cdot \frac{dv}{dt} . \tag{1.5}$$

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

$$\frac{dv}{dt} = \frac{\partial v}{\partial t} + \frac{\partial x}{\partial t} \frac{\partial v}{\partial x} = \frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} , \tag{1.6}$$

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

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

On the other hand, a movement of mass must comply with 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 injection caused by the source.

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

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

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

Equations (1.7) and (1.9) allow the derivation of an acoustic theory. They are, however, coupled in three variables, pressure, density and particle velocity, and they are nonlinear. Two linear equations can be easily found if the effects of sound are assumed small. We will see later that this is a quite sufficient approximation 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 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.
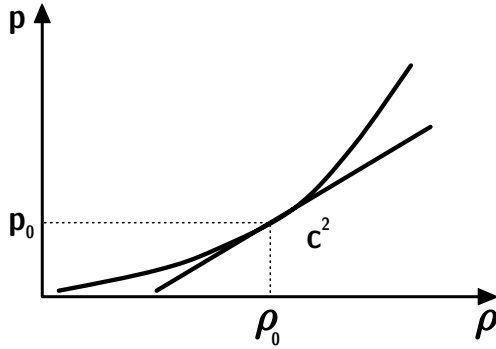
Fig. 1.5. Linearization of small amplitudes of density as a function of pressure

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

$$p = \left(\frac{\mathrm{d}p_{tot}}{\mathrm{d}\rho_{tot}}\right)_{ad} \cdot \rho = c^2 \cdot \rho \,, \tag{1.10}$$

with a constant c^2 as an abbreviation of $\mathrm{d}p/\mathrm{d}\rho$.

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

$$-\frac{\partial p}{\partial x} = \rho_0 \frac{\partial v}{\partial t} \tag{1.11}$$

$$-\rho_0 \frac{\partial v}{\partial x} = \frac{1}{c^2} \frac{\partial p}{\partial t} + \rho_0 q \,. \tag{1.12}$$

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

$$\frac{\partial^2 p}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = -\rho_0 \frac{\partial q}{\partial t} \,, \tag{1.13}$$

or in short,

$$\Delta p - \frac{1}{c^2} \ddot{p} = -\rho_0 \dot{q} \,. \tag{1.14}$$

The equation is well known in mathematics and physics as a wave equation. In the same formal notation, it can be derived for particle velocity, density, or temperature. In three dimensions, it is extended with the spatial

differential operator, the Laplace operator, in Cartesian coordinates (x, y, z) according to common notation used in mathematics and physics

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \,, \tag{1.15}$$

in cylindrical coordinates (r, φ, z),

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

or in polar coordinates (r, ϑ, φ),

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

1.2 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 of wave acoustics such as speed of sound, sound intensity, 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 the 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's solution); see, for instance, (Kuttruff 2007):

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

The first term, f, describes the propagation of the local state of sound pressure $p(x, t)$ in space and time in the positive x direction, the latter term, g, in the 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 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 of ideal gas theory from

$$c^2 = \left(\frac{\mathrm{d}p_{tot}}{\mathrm{d}\rho_{tot}} \right)_{ad} = \frac{\kappa p_{tot}}{\rho_{tot}} \,. \tag{1.19}$$

Taking into account more thermodynamic effects, the humidity, altitude, etc., it can be estimated rather precisely (ISO9613). For most cases of sound propagation in air, the approximation

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

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

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

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

and the speed of sound is

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

Now we come back to the elementary solution $p(x, t)$. Assuming $g = 0$, we now consider a wave in the 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:

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

and

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

This ratio, Z_0, is called the wave impedance or characteristic impedance of 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 driving pressure needed to set the medium's particles into motion.

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	$1{,}48 \cdot 10^6$

1.3 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,

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

$\hat{p}$ 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 the space and time domains 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 rescale the periods of the wave in space and time, respectively, to 2π. The period in space is called wavelength, λ. We obtain it from

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

With the important relation,

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

we can introduce the temporal period, T, of the wave:

$$\omega = \frac{2\pi}{T} = 2\pi f . \qquad (1.28)$$

f denotes the frequency in the unit Hertz.

1.4 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 the 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 a "spectrum."

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

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

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

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 toward aperiodic signals ($T_0 \to \infty$), which have then a continuous spectrum (line spacing $\to 0$). More details will be discussed in Sect. 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.

1.5 Energy and level

Usually the sound pressure is not presented in linear form in its units of pascals ($1\,\mathrm{Pa} = 1\,\mathrm{N/m^2}$). 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 a first approximation; see also Chap. 6). The range of sound pressures should be also discussed in relation to our initial assumption that sound is a small displacement or pressure or density fluctuation compared with static conditions. If we just assume a static pressure of $100\,\mathrm{kPa}$ in the atmosphere at sea level, we have to deal with sound pressures of orders of magnitude lower, between $0.00001\,\mathrm{Pa}$ and $1000\,\mathrm{Pa}$. 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 lower than the static pressure p_0 by orders of magnitude.

type	time domain	frequency domain

violine

flute

piano

drum

vowel (a)

consonant (k)
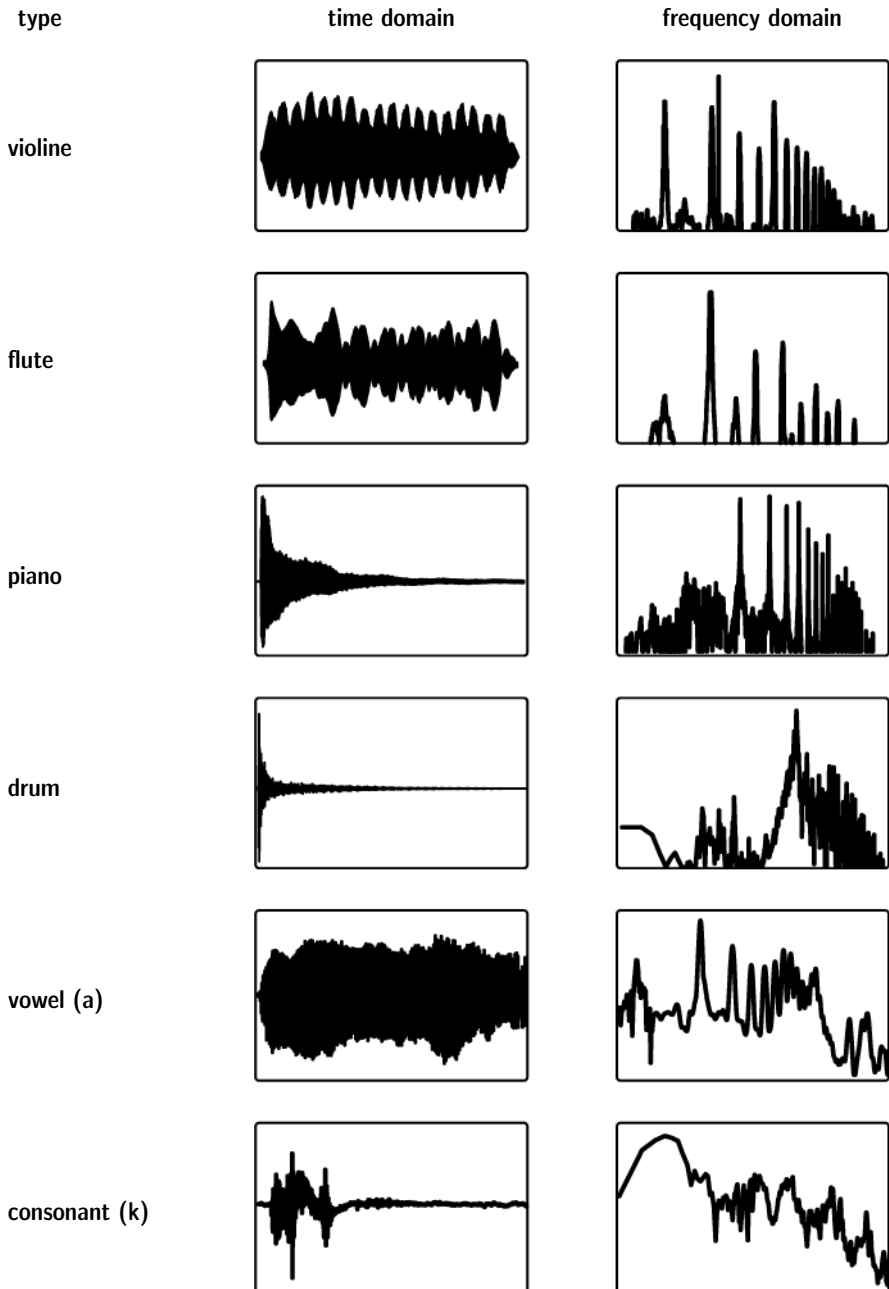
Fig. 1.6. Typical acoustic signals and their spectra

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 with 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 investigation 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, 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 kinetic and potential energy. The volume shall 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:

$$w = \frac{\rho_0 \hat{v}^2}{2} = \frac{\hat{p}^2}{2\rho_0 c^2} \tag{1.31}$$

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

With introduction of the "root mean square" sound pressure (rms) the sound pressure level can finally be written as:

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

with

$$\tilde{p} = \sqrt{\frac{1}{T}\int_0^T p^2(t)\mathrm{d}t} = p_{rms} \tag{1.33}$$

and setting the reference sound pressure $p_0 = 20\,\mu\mathrm{Pa}$ which is approximately the human hearing threshold in midfrequencies.

Table 1.2. Sound pressure levels of typical sound events

Event	Level in dB
Hearing threshold in midfrequencies	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

1.6   Sound intensity

The microscopic energy in a sound wave is not just a static phenomenon. Energy is also transported. Wave propagation should not be mixed up with particle flow, however. At zero mean flow (such as random wind), 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 into sound fields, particularly for sound fields 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 $1 \, \mathrm{m}^2$. It is called sound intensity:

$$\vec{I} = \overline{p \cdot \vec{v}} = \frac{1}{T} \int_0^T p\vec{v}\mathrm{d}t \, . \tag{1.34}$$

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

$$\left|\vec{I}\right| = \frac{\overline{p^2}}{\rho_0 c} = \frac{\tilde{p}^2}{\rho_0 c} \, , \tag{1.35}$$

with the direction of sound intensity in direction of propagation. Sound intensity can also be denoted with a level, the intensity level:
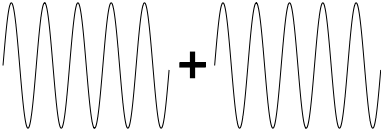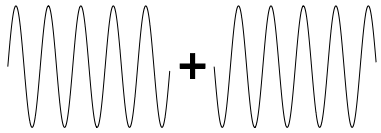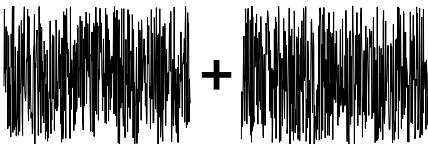
$$L_I = 10 \log \frac{\left| \vec{I} \right|}{I_0} .$$
(1.36)

$I_0 = 10^{-12}$ W/m^2. This choice of reference intensity is made to adjust for the same levels of sound pressure and sound intensity in a plane wave.

1.7 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 binominal formula $(p_1 + p_2)^2 = p_1^2 + p_2^2 + 2 p_1 p_2$, where the latter term of pressure signal multiplication cancels with incoherent p_1 and p_2.

Table 1.3. Example: Level addition of two signals of 50 dB

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

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

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

or in level representation,

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

1.8 Frequency bands

In acoustics, standardized frequency bands are often used, typically 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);

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

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,

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

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 discussion of transmission characteristics or noise frequency content. The band-filtered results, however, are still sound levels. The specific spectrum in interpretation of Fourier transformation theory is discussed in more detail later (see Sect. 7.2).
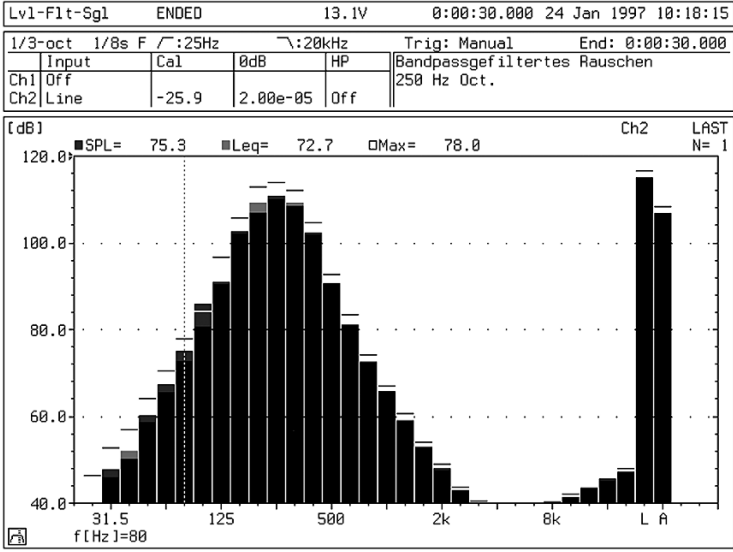
Fig. 1.7. Band-filtered spectrum of a broadband signal (courtesy of Norsonic A/S)

2 Sound sources

For the discussion of plane waves, we concentrated on the sound field at arbitrary observation points. We did not consider sound sources. Characterization of sound sources, however, is one of the most important tasks in simulation and auralization. A first approach to a physical description of a source can be found in solving the wave equation in polar coordinates. At the origin of the coordinate system, at $r = 0$, we assume a small, point-such as 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 into motion. The corresponding strength is denoted as "volume flow" or "volume velocity."

2.1 Spherical waves

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

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r}, \tag{2.1}$$

and the homogeneous wave equation reads

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

By using the solution with the form

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

the differential equation is transformed into the simpler equation

$$\frac{\partial^2 f}{\partial r^2} - \frac{1}{c^2}\ddot{f} = -\rho_0 \dot{Q}. \tag{2.4}$$

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

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

$p(r,t)$ represents a pressure wave propagating from $r=0$ toward larger r. It is called a "spherical wave."[5]

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 yields (after (Kuttruff 2007))

$$p(r,t) = \frac{\rho_0}{4\pi r} \dot{Q}\left(t - \frac{r}{c}\right).$$ (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 for poor low-frequency output of small loudspeakers.

2.2   Harmonic monopole source and sound power

We assume harmonic excitation

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

Then, with Eq. (2.6) we obtain

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

and

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

the latter by using Eq. (1.11).

---

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

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

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

Furthermore, the sound intensity is

$$I = \frac{\rho_0 \hat{Q}^2}{32\pi^2 c} \cdot \frac{\omega^2}{r^2} \approx \frac{\tilde{p}^2}{\rho_0 c} \quad \text{for } kr \rangle\rangle 1 \,.$$
(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 by an observation surface area of $4\pi r^2$. The result is

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

Thus we find the relation

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

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

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

These findings can be applied to 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	$10^3 - 10^4$	150–160

Sound power levels are part of the characterization of sources. These levels of sound emission, however, should never be confused with the sound pressure at the receiver. Without knowing the distance between source and receiver, 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 found in discussion of point sources are not generally present. First of all, a real source will have a limited power capacity. Accordingly, the force applied on 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 consider distributed source arrangements and as a starting point, the pulsating sphere.

2.3 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

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

where $\underline{p}(a, t)$ denotes 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. Thus we can calculate the radiation impedance.

$$\underline{Z}_r = \frac{\rho_0 c S}{1 + \dfrac{1}{jka}}.\tag{2.16}$$

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

$$P = \frac{1}{2}|v(a,t)|^2 \mathrm{Re}\{\underline{Z}_r\} = \frac{1}{2}|v(a,t)|^2 W_r,\tag{2.17}$$

$$W_r = \frac{\rho_0 c S}{1 + \dfrac{1}{k^2 a^2}} \approx \begin{cases} S\rho_0 c k^2 a^2 & \text{for} \quad ka \ll 1 \\[2mm] S\rho_0 c & \text{for} \quad ka \gg 1 \end{cases}.\tag{2.18}$$
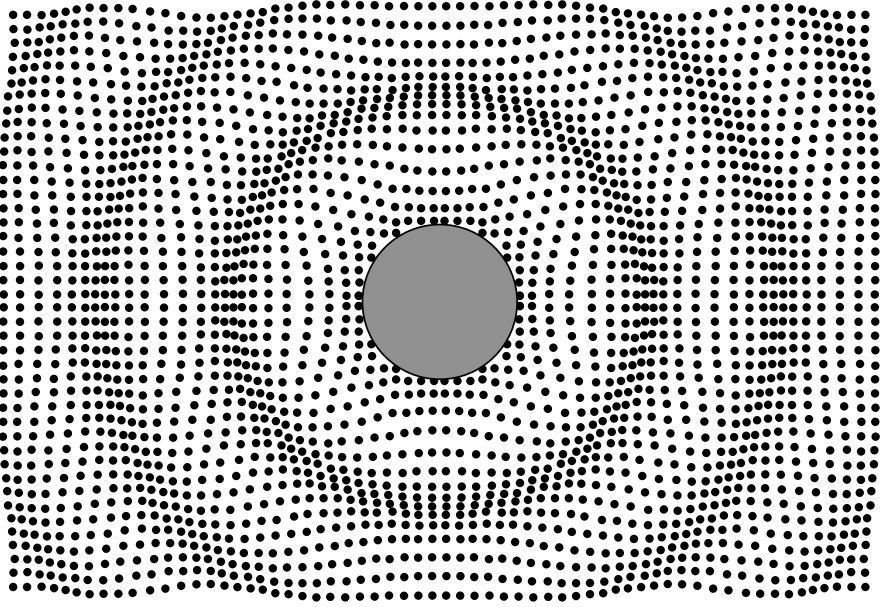
Fig. 2.1. Radiation from a spherical source (after (Möser 2004))

Because the equation for radiation impedance has a mathematical structure like the impedance of a parallel circuit, it can be rewritten in the form of admittance:

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

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

$$m_s = 4\pi a^3 \rho_0 .$$
(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 relationships are very basic, but they apply 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 also has a specific near-field. For more details, refer to (Morse and Ingard 1968).

2.4   Multipoles and extended sources

Multipole sources or extended sources represented by an arrangement of monopoles can be composed by summing or integrating over a distribution of monopoles or infinitesimally small surface elements. 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 toward 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. 2.5.

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, $\hat{Q}_2 = -\hat{Q}_1 = Q$.

The general equation for point source superposition

$$\underline{p} = \frac{j\omega\varrho_0}{4\pi} \sum_n \hat{\underline{Q}}_n \frac{e^{j(\omega(-kr_n)}}{r_n} \qquad , \quad \vec{r}_n = |r_n|, \tag{2.21}$$

reads in this case,

$$\underline{p} = \frac{j\omega\varrho_0 \hat{Q}}{4\pi r} e^{j(\omega t - kr)} \left(e^{j\frac{kd}{2}\cos\vartheta} - e^{-j\frac{kd}{2}\cos\vartheta} \right). \tag{2.22}$$
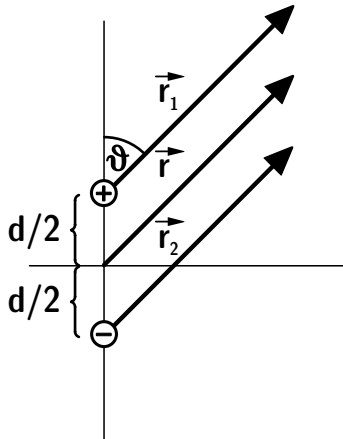
Fig. 2.2. Dipole source
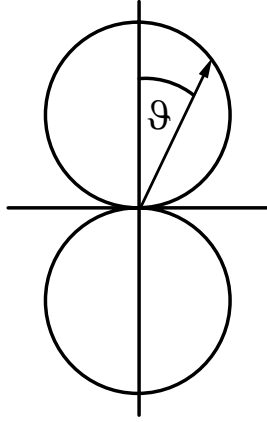
Fig. 2.3. Figure-of-eight directionality of a dipole source. Note the rotational symmetry in the vertical axis

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

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

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.

Directional characteristics are present for any kind of source with a surface velocity distribution differing from spherical uniformity. Generally 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 a characteristic source dimension. Under these assumptions, the summation or integration result can be separated into (r) and (ϑ, φ), thus, into a distance-dependent monopole term and a directional term dependent on the polar and azimuthal angles.

If all monopole sources have identical phases, we can further define

$$\hat{Q}_{\text{tot}} = \sum_n \hat{Q}_n \tag{2.24}$$

and the directionality factor $\underline{G}(\vartheta, \varphi)$, which is usually illustrated in polar diagrams,

$$\underline{G}(\vartheta, \varphi) = \frac{1}{\hat{Q}_{tot}} \sum_n \hat{Q}_n e^{jkr_n' \cos[\bar{r}, \bar{r}_n']} . \tag{2.25}$$

In the end, this yields

$$p = \frac{j\omega\omega_0 \hat{Q}_{tot}}{4\pi r} e^{j(\omega t - kr)} \underline{\Gamma}(\vartheta, \varphi). \tag{2.26}$$

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

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

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

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

$$I_n = \frac{P}{4\pi r^2} |\Gamma|^2_{max=1}, \tag{2.28}$$

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

$$L = L_w - 20\log r - 11 + L_D \tag{2.29}$$

where L_D denotes the directivity factor,

$$L_D = 10\log \iint_\Omega |\Gamma(\vartheta, \varphi)|^2 \, d\Omega. \tag{2.30}$$
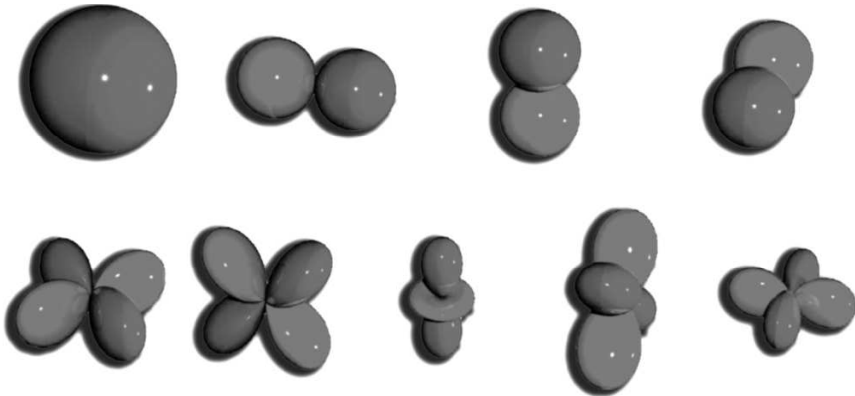
Fig. 2.4. Monopole, dipole and quadropole radiation patterns

2.5   Spherical harmonics

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

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

that leads to the three differential equations, including three constants, n, C and k.

$$\left(\frac{\partial^2}{\partial \varphi^2} + n^2 \right)\Phi = 0 \,,$$

$$\frac{1}{\sin\vartheta}\frac{\partial}{\partial\vartheta}\left(\sin\vartheta \frac{\partial P}{\partial\vartheta} \right) + \left(C - \frac{n^2}{\sin^2\vartheta} \right)P = 0 \,, \tag{2.32}$$

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial R}{\partial r} \right) + \left(k^2 - \frac{C}{r^2} \right)R = 0 \,.$$

The solution for the azimuthal angle is straightforward:

$$\Phi = \cos(n\varphi) + \sin(n\varphi) \,, \tag{2.33}$$

while for the elevation, a functional basis of Legendre polynomials completes the result for the spatial angles,

$$P_m^n(x) = (1-x^2)^{\frac{n}{2}}\frac{d^n}{dx^n}P_m(x)\,, \quad 0 \le n \le m \,, \tag{2.34}$$

and

$$\begin{aligned} Y_{mn}^1(\vartheta,\varphi) &= \cos(n\varphi)P_m^n(\cos\vartheta) \\ Y_{mn}^{-1}(\vartheta,\varphi) &= \sin(n\varphi)P_m^n(\cos\vartheta) \end{aligned} \tag{2.35}$$

These functions are called "modes" or "spherical harmonics." Legendre polynomials are easily found in mathematical textbooks, in mathematical software and, most useful for our purposes, in software libraries for signal processing. Due to the orthogonality of spherical harmonics, arbitrary far-field directivity patterns can be expanded into a set of spherical harmonics. The spatial radiation distribution in general is thus given by linear combinations of Eq. (2.35).

If the near field and the source-receiver distance shall be considered, too, the third part of the differential equation is further investigated (with the constant C derived from the polar angle equation):

$$\frac{\partial}{\partial r}\left(r^2 \frac{\partial R}{\partial r}\right) + k^2 r^2 R = m(m+1)R . \tag{2.36}$$

This differential equation can be converted into

$$\frac{\partial^2 R}{\partial \zeta^2} + \frac{2}{\zeta}\frac{\partial R}{\partial \zeta} + \left(\zeta^2 + m(m+1)\right)\frac{R}{\zeta^2} = 0 , \tag{2.37}$$

the solutions of which are linear combinations of Bessel and Neumann functions, called spherical Hankel functions of the second kind and mth order:

$$R_m(\zeta) = h_m^2(\zeta) = \sqrt{\frac{\pi}{2\zeta}} J_{m+\frac{1}{2}}(\zeta) - j\sqrt{\frac{\pi}{2\zeta}} N_{m+\frac{1}{2}}(\zeta) , \tag{2.38}$$

with the option to develop the Hankel function into a series as a function of the Helmholtz number $\zeta = k$:

$$h_0^2(\zeta) = \frac{je^{-j\zeta}}{\zeta} ,$$

$$h_1^2(\zeta) = -\frac{e^{-j\zeta}}{\zeta}\left(1 - \frac{j}{\zeta}\right) ,$$

$$h_2^2(\zeta) = -\frac{je^{-j\zeta}}{\zeta}\left(1 - \frac{3j}{\zeta} - \frac{3}{\zeta^2}\right) , \tag{2.39}$$

$$\dots$$

In summary, the spherical modes are given by

$$\psi_{mn}^1(k,r,\vartheta,\varphi) = h_m^2(kr)P_m^n(\cos\vartheta)\cos(n\varphi)$$
$$\psi_{mn}^{-1}(k,r,\vartheta,\varphi) = h_m^2(kr)P_m^n(\cos\vartheta)\sin(n\varphi) \tag{2.40}$$

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

$$\int \psi_{mn}\psi_{m'n'}\mathrm{d}V = \begin{cases} 0 \text{ for } n \neq n' \vee m \neq m' \\ 1 \text{ for } n = n' \wedge m = m' \end{cases} . \tag{2.41}$$
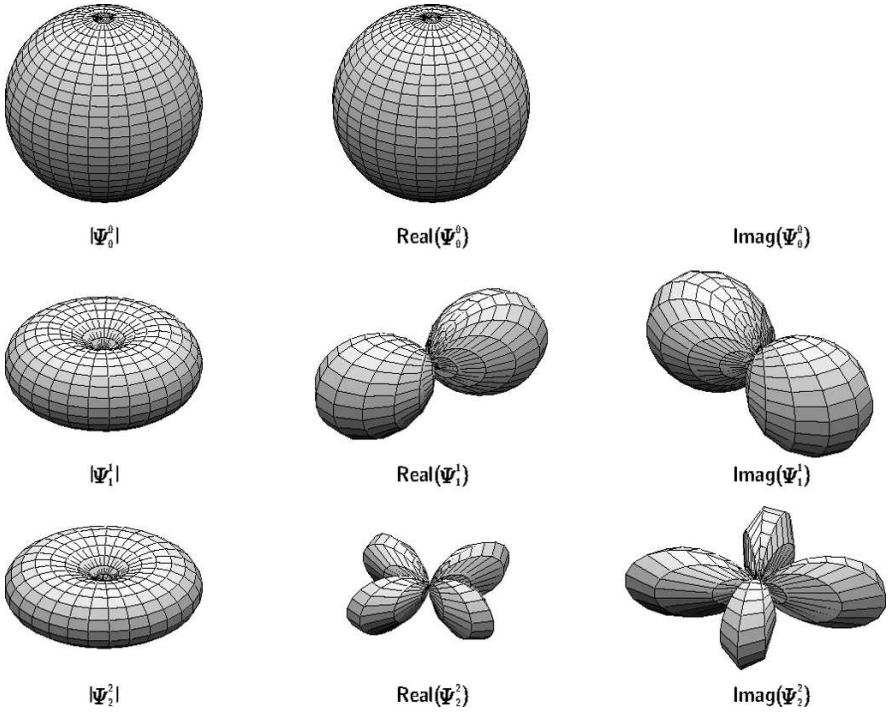
Fig. 2.5. Spherical harmonics

3 Sound propagation

Sound radiation from sources often happens in free space. This applies to the sound that reaches the receiver directly over a free line of sight.[6] The dependence of sound intensity and sound pressure level on the propagation distance is an important law of sound propagation. The following table summarizes some basic distance laws.

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
Circular rigid piston in baffle, on axis	$L = L_w - 20\log r - 11 + L_D$	6
Incoherent line above ground ($P' = P/\Delta x$) $(L'_w = L_w - 10\log\dfrac{\Delta x}{1m})$	$L = L'_w - 10\log r - 3$	3

3.1 Reflection of plane waves at an impedance plane

Undisturbed free-field propagation is a good sound field model for anechoic environments. Any obstacle in the field, however, will interact with the incident sound pressure. Small or large objects or room boundaries influence the total sound pressure through reflection, scattering and diffraction.

---

[6] Obviously this applies typically to outdoor sound propagation or to sound propagation in close distances between souce and receiver, so that no obstacle is in the propagation path.
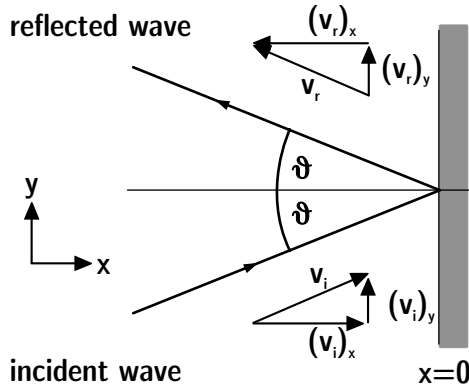
Fig. 3.1. Reflection of plane waves at impedance plane

Furthermore, properties of the medium such as inhomogeneity of sound velocity and viscous effects can introduce influences such as refraction and attenuation, respectively. A plane wave incident on a (infinitely) large smooth wall is reflected according to Snell's law. This can be described with a "specular" reflection.

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

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

and the reflected wave is

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

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

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

$Z_0 = \rho_0 c$ 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. If the impedance is independent of the angle of incidence, we talk about "local reaction." The consequence of local reaction is that 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. 5.2) and for porous absorbers with high flow resistivity.

Particularly important is the absorption coefficient, α:

$$\alpha = \frac{|p_i|^2 - |p_r|^2}{|p_i|^2} = 1 - |\underline{R}|^2 , \tag{3.4}$$

and the specific impedance

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

For locally reacting surfaces, the absorption coefficient is

$$\alpha = \frac{4\,\mathrm{Re}(\underline{\varsigma})\cos\vartheta}{1 + 2\,\mathrm{Re}(\underline{\varsigma})\cos\vartheta + |\underline{\varsigma}|^2\cos^2\vartheta} . \tag{3.6}$$

3.1.1   Examples of wall impedances

For perpendicular incidence, we can define some extreme cases. Real materials approximate these conditions quite well. For example, a heavy concrete wall represents a "hard wall," and the ocean surface in underwater sound represents a "soft wall." Some other examples are listed as follows (after (Kuttruff 2007)):

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

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

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

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 with neglected internal and mounting stiffness and losses. The force $p - p'$ on a surface area of $1\,\mathrm{m}^2$ excites the layer to vibrations with the velocity $v' = v$. p, v denote the pressure and particle velocity on the incident side and p', v' on the transmission side of the mass layer.

One can easily show that

$$\underline{Z} = \left(\frac{\underline{p}}{\underline{v}}\right)_{x=0} = Z_0 + j\omega m'' . \tag{3.7}$$
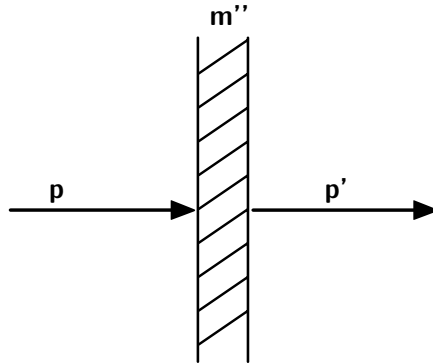
Fig. 3.2. Sound hitting a mass layer

The absorption coefficient is

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

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

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

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

Mass layer in front of a hard wall

We now consider the combined impedance from a mass layer mounted with an air gap to the rigid wall. At first, the air gap alone is considered. It results from choosing a new reference plane for $x = 0$. Shifting the reference plane this way may also be intended to place another object such as porous fabric at $x = 0$ (see below).

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

$$\underline{R} = \left|\underline{R}\right| \mathrm{e}^{-2jkd} \; , \quad \left|\underline{R}\right| = 1 \tag{3.10}$$
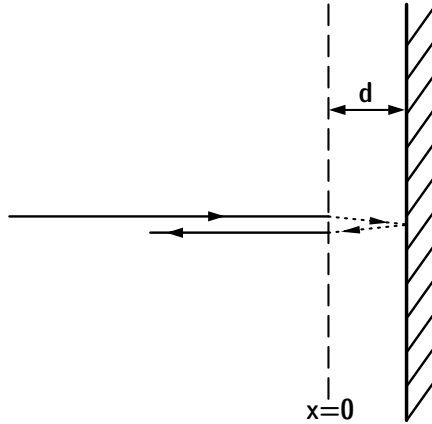
Fig. 3.3. Sound incident on a virtual air gap in front of a hard wall

and

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

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

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

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

$$s'' = \frac{1}{n''} = \frac{Z_0 c}{d}.\tag{3.13}$$

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'', addressed to porous material placed in the air gap.

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

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

$$\omega_0 = c\sqrt{\frac{\rho_0}{m''d}}.\tag{3.15}$$
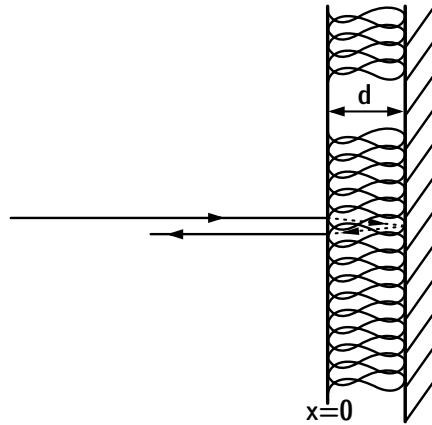
Fig. 3.4. Sound hitting a mass layer with air gap and porous filling

Porous layer in front of a hard wall

The impedance of the air gap is now combined with a purely resistive component, w'':

$$\underline{Z} = w'' - jZ_0 \cot(kd) \,. \tag{3.16}$$
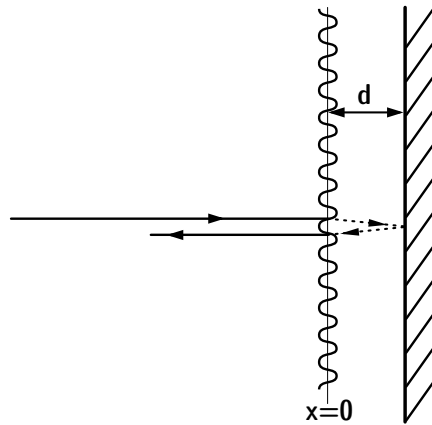
Fig. 3.5. Sound hitting a curtain in front of a wall

3.2   Spherical wave above impedance plane

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. In contrast to a plane wave reflection, for spherical wave incidence, the impedance plane is hit at various angles and the total reflection must be integrated:

$$p = \frac{j\omega\omega_0\hat{Q}e^{-jkr_0}}{4\pi r_0} - \frac{\omega^2\rho_0\hat{Q}}{4\pi c}\int_{\vartheta} J_0(kr\sin\vartheta)e^{-jk(z+z_0)\cos\vartheta}\underline{R}(\vartheta)\sin\vartheta d\vartheta. \quad (3.17)$$

The first term represents the direct sound and the latter the contribution of the reflection. An approximation of the entire expression assuming a constant angle of incidence, ϑ_0, is

$$p = \frac{j\omega\omega_0\hat{Q}e^{-jkr_0}}{4\pi r_0} + \frac{j\omega\omega_0\hat{Q}e^{-jkr_1}}{4\pi r_1}\underline{R}(\vartheta_0) . \quad (3.18)$$

This equation assumes a constant angle and, thus, a plane wave. The contribution of the reflection can be related to another point source, called an "image source" (see also Sect. 11.3), which is apparently located below the surface and radiates a spherical wave whose amplitude is reduced by $\underline{R}(\vartheta_0)$. Although a spherical wave is present, the reflection is calculated for constant angle, ϑ_0. Apparently, for the moment of reflection, the sound
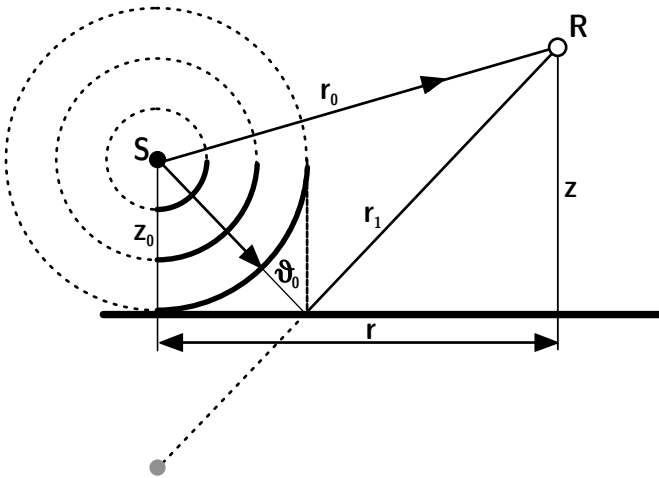
Fig. 3.6. Spherical wave reflection above impedance plane

field model is switched to a plane wave.[7] This "image source model" is even exact when $R = 1$ or $R = -1$. For other reflection factors, the approximation was identified sufficiently accurately for z, z_0 and $r \gg \lambda$ (Suh and Nelson 1999); see also Sect. 11.3.3.

In a reflecting plane, plane waves or spherical waves can be separated into elementary sources (Huygens' principle). These sources together build up an interference field that produces a plane or spherical wave. This model is the background for the simplicity in Eq. (3.18) (which is exact when $|R| = 1$).

If, however, the plane has discontinuities in 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.

3.3   Scattering

3.3.1   Object scattering

Sound waves may hit obstacles. Depending on the size of the objects compared with the wavelength, the scattered field has large amplitudes in the forward direction ("forward scattering"), in the reverse direction ("reflection") or in any other direction following a specific distribution. The exact formulation and solution of the scattered field amplitude is a difficult problem, except in academic cases of objects such as spheres, cylinders, etc.

An efficient strategy for addressing the problem in practical cases is to map the scattered field to an equivalent field created by a spherical scatterer. By this approach, the scattering cross section is defined. For an incidence plane wave with intensity I_0, the scattering cross section is ($\lambda \ll a$)

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

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,

$$p = p_0 + p_s \qquad (3.20)$$

---

[7]   Plane wave approximation.
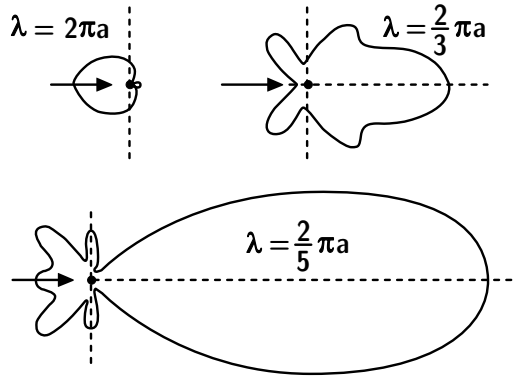
Fig. 3.7. Sound scattering at a sphere with radius *a* (after (Morse and Ingard 1968))

p_s must fulfil 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

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

and for the particle velocity of the scattered wave,

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

The radiated field generated from this velocity distribution is to be calculated by using the models of Sect. 2.4 in an equivalent radiation problem.

3.3.2 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 in a way 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 are significantly smaller than $\lambda/2$. When corrugations are of 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.
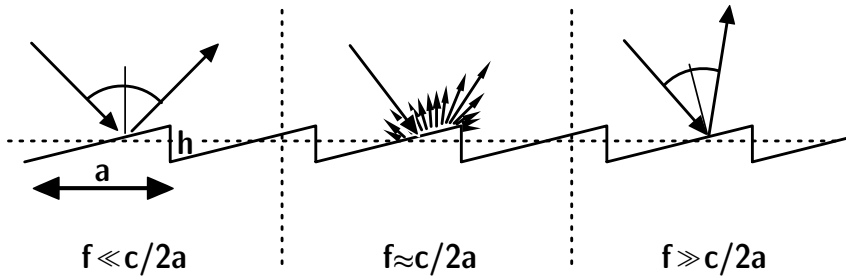
Fig. 3.8. Scattering caused by surface corrugations

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

For some corrugation types, analytic or numerical solutions are available. 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 decomposition; the zero-order component (zero-order lobe) represents the specular reflection component. Higher order lobes direct the sound in nonspecular directions.

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

$$E_{spec} = (1-\alpha)(1-s) \equiv (1-a), \quad E_{total} = (1-\alpha) \tag{3.23}$$
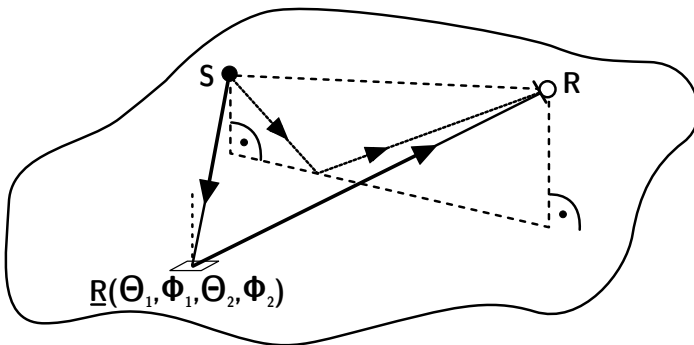
Fig. 3.9. Delayed sound paths from surface scattering
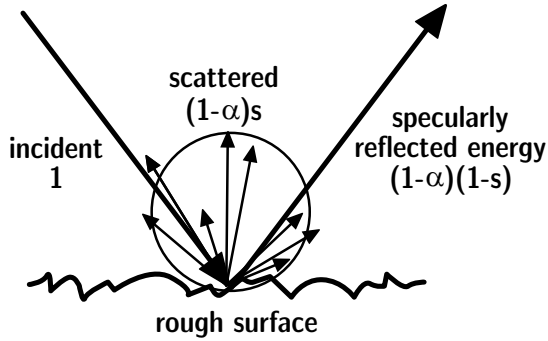
Fig. 3.10. Energy reflected from a corrugated surface into a scattered and specularly reflected portion. Definition of the total reflected energy $(1 - \alpha)$, the scattered energy $(1 - \alpha)s$ and specularly reflected enery $(1 - \alpha)(1 - s)$ (Vorländer and Mommertz 2000)

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

$$s = \frac{a - \alpha}{1 - \alpha} = 1 - \frac{E_{\text{spec}}}{E_{\text{total}}}. \tag{3.24}$$
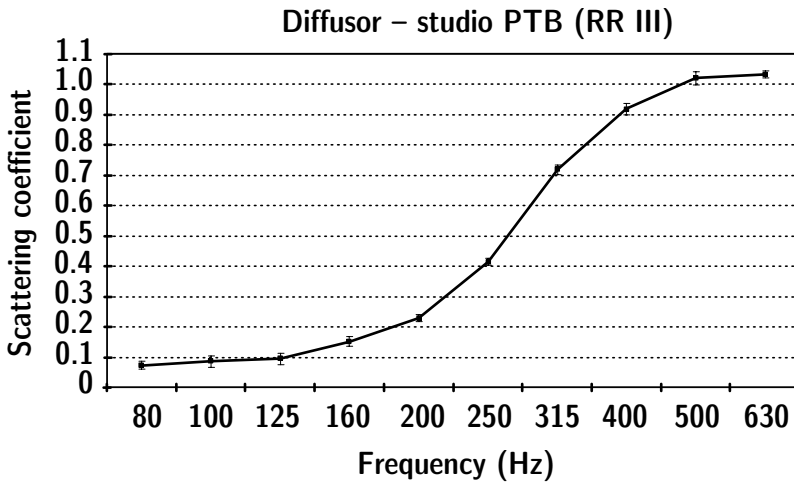
Fig. 3.11. Random-incidence scattering coefficients measured on a scale-model sample of a diffusing ceiling structure (PTB, Round Robin III, used in (Bork 2005a))

Furthermore, measurement data are available to serve as input data for simulation software (see Sect. 11.2 and Annex). These data describe the energetic amount of scattering (scattering coefficient) compared with the zero-order scattering lobe (specular component).

The uniformity of the directional scattering distribution (diffusion coefficient) is also of interest, but this should not be confused with the scattering coefficient (Cox and D'Antonio 2004; Cox et al. 2006). The directional distribution of the sound scattered from the surface of the object is obtained analogously to the way one might test the uniformity of loudspeaker radiation. Thus, a free-field polar response must be calculated or measured. The diffusion coefficient is then a single number 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.

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.

$$\left|\bar{I}(\vartheta)\right| = (1-\alpha)\frac{B dS}{\pi r^2}\cos\vartheta. \qquad (3.25)$$
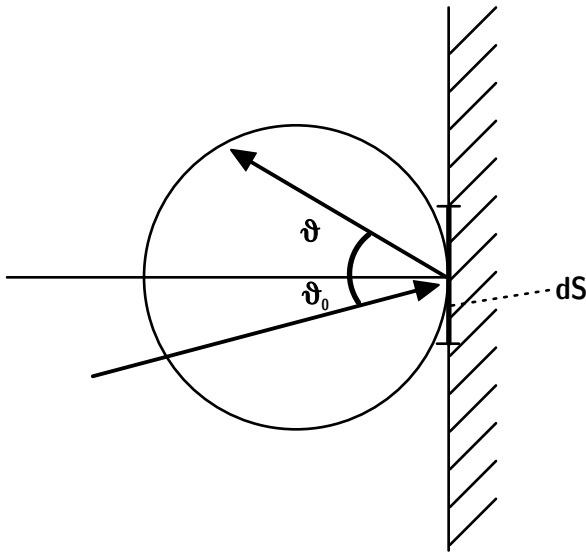
Fig. 3.12. Probability distribution of scattered sound (Lambert's law)

B denotes the irradiation strength on the wall; α the part of the incident energy *B*d*S* which is not reflected from the wall. This kind of scattering distribution creates a constant illumination effect on a detector with fixed sensor area, such as a membrane or, in the optical analogy, such as a camera or an eye with a 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 such as a disc than a sphere, since the light scattered from the surface near the moon's apparent midpoint (at a direction normal to the observer) has the same effective brightness as the light scattered from the apparent circumference, with an observation angle of almost $\pi/2$.

3.4 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 with respect to wavelength, a shadow region first appears and then grows clearer and sharper. The shadow results from a total cancellation of the incident wave by the diffraction wave.

As for scattering, the calculation of the diffracted field is analytically possible for geometrically simple obstacles (sphere, circular disc, cylinders, free edge on a screen, slits or holes in a screen, etc.).

Diffraction models must be taken into account when sound propagation is predicted for large distances in urban areas, for instance, or in open-plan offices, just to give a few examples. Noise barriers are a typical example for the application of engineering models of diffraction. It can be shown (Maekawa 1968) that 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):

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

Because the sizes of many objects encountered in daily life are of the order of magnitude of the wavelength, diffraction is easily noticeable, although hard to predict by calculation. Diffraction influences 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.
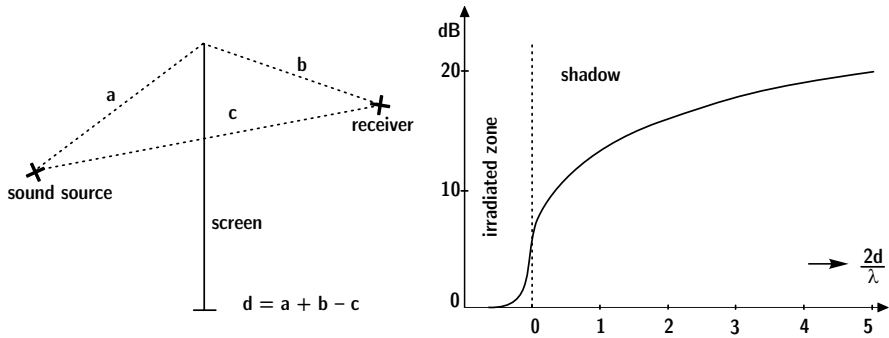
Fig. 3.13. Estimating the insertion loss of a screen

3.5 Refraction

Based on Fermat's principle, 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,

$$n = \frac{c'}{c} = \frac{\sin \vartheta'}{\sin \vartheta} . \tag{3.27}$$

for calculating the geometric conditions.

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

$$T = \frac{2Z'}{Z'+Z_0} , \tag{3.28}$$

while for the reflection factor, Eq. (3.3) still holds.
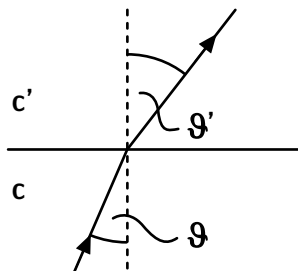
Fig. 3.14. Refraction at the boundary between two media

This approach can also be used to derive curved transmission paths in layered media. If the sound speed changes gradually, for instance, with temperature at various atmospheric elevations or with 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 $c + \mathrm{d}c$,

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

This effect corresponds to a curved sound path with reciprocal radius:

$$\frac{1}{r_k} = \frac{1}{c}\frac{\partial c}{\partial n}, \tag{3.30}$$

where n denotes the normal direction of the sound wave. The curvature is the bigger, the larger the sound speed gradient in the direction normal to the propagation. This effect can lead to curved sound rays in the atmosphere. In some weather conditions (temperature increasing with height), upward radiated sound may be bent down to reach the ground again. Long-distance sound propagation is affected greatly. The same will be observed with wind speed profiles. In outdoor sound propagation, therefore, refraction is an essential aspect of sound field modelling.

3.6 Attenuation

Attenuation is another effect of long-distance sound propagation. It should be noted that long-distance sound propagation may happen outdoors, of course, but also 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 is outdoor or indoor, the latter involves numerous wall reflections.

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

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

where m denotes the energetic attenuation coefficient and

$$\underline{k}' = \frac{\omega}{c} - j\frac{m}{2}. \tag{3.32}$$

The intensity along the x coordinate in a plane wave is

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

or expressed in effective level loss,

$$D = 4{,}34 \, m \quad \text{dB/m}, \tag{3.34}$$

with the level loss after propagation over the distance x

$$L = L_0 - D \cdot x. \tag{3.35}$$

The reason for attenuation is viscosity, heat conduction and thermal re-laxation (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.
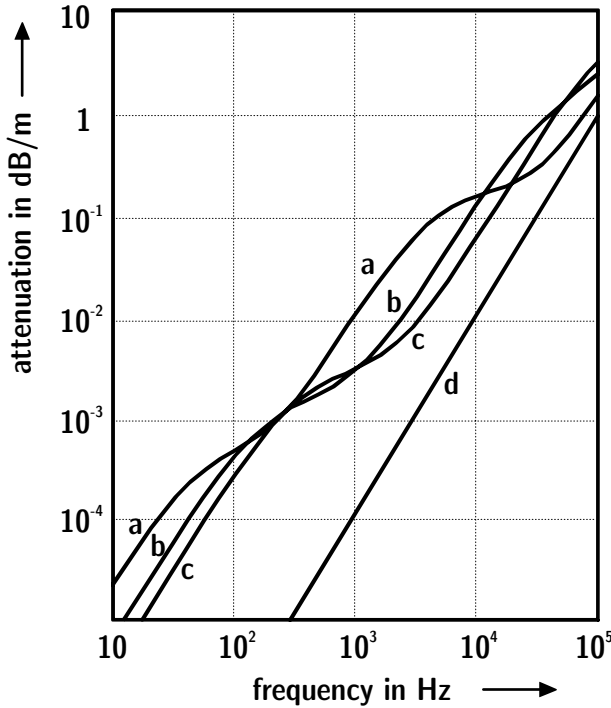
Fig. 3.15. Typical attenuation curves (after (Kuttruff 2007)) of humid air a) 10%, b) 40%, c) 80%, d) classical theory ($\sim \omega^2$)

3.7 Doppler effect

One of the acoustic effects best known to the public is the Doppler effect, the frequency shift perceived when cars, police or fire-brigade horns are passing by at high speeds. Moving sound sources or receivers cause a change in the received rate of sound pressure maxima and minima and, thus of frequency at the receiver. Differing from the analogy in 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 toward the source, the received frequency is higher than the radiated. Inserting $x = x_0 - Vt$ into a harmonic sound pressure equation yields

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

The perceived frequency is thus

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

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

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

with the effective frequency

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

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 distinction from relative movement can be better understood by discussing the extreme cases $(V \to c)$ with respect to a harmonic signal. The rate of received pressure maxima determines the received frequency.

Table 3.2. Four cases of the Doppler effect at relative speed $c0$

1) Moving receiver toward the source

$$f'= f\left(1+\frac{V}{c}\right)\xrightarrow{V \to c} 2f$$
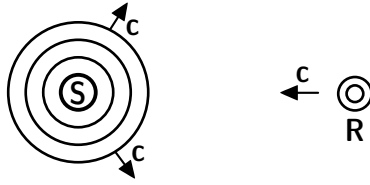
2) Moving receiver away from the source

$$f'= f\left(1-\frac{V}{c}\right)\xrightarrow{V \to c} 0$$
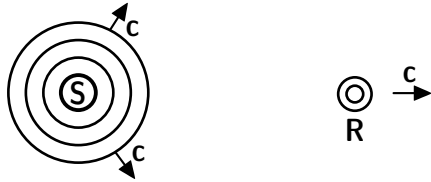
3) Moving source toward the receiver

$$f'= \frac{f}{1-\dfrac{V}{c}}\xrightarrow{V \to c} \infty$$
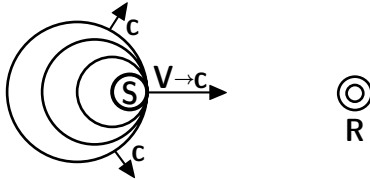
4) Moving source away from the receiver

$$f'= \frac{f}{1+\dfrac{V}{c}}\xrightarrow{V \to c} f/2$$
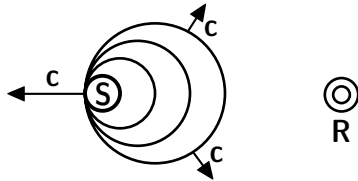
4 Sound fields in cavities and in rooms

In the previous chapters, the basics of sound wave radiation and propagation were discussed. With these mathematical models, free-field propagation can be studied and simulated. Numerous problems, however, are related to sound in cavities, whether small as in appliances or in vehicles, a little larger as in living rooms or kitchens, or rather large as in concert halls and opera houses.

In our daily lives, we can well distinguish by listening whether or not we are in a room, and we can estimate the room volume and amount of absorption from the acoustic 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 the physical room acoustic effects at first from the viewpoint of 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 for describing room acoustic effects and for room acoustic modelling.

4.1 Cavities

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

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

$$\frac{\hat{p}}{p_0} = \frac{\hat{V}}{V_0} . \tag{4.1}$$
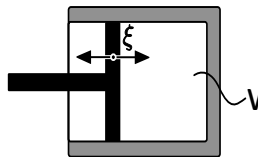
Fig. 4.1. Small cavity excited by piston
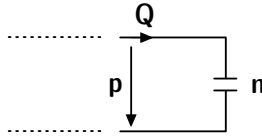
Fig. 4.2. Input impedance of a small cavity

For example, the volume variation can be created by a little piston of area S that vibrates with an amplitude ξ. Due to the 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 calculated easily.

$$p = \frac{p_0}{V_0} S\xi = \frac{\rho_0 c^2}{j\omega V_0} Sv = \frac{\rho_0 c^2}{j\omega V_0} Q .\tag{4.2}$$

The acoustic impedance of the cavity reacting on the piston is thus spring-such as (see also Eq. (3.13)).

$$\underline{Z} = \frac{1}{j\omega n}\tag{4.3}$$

with

$$n = \frac{V_0}{\rho_0 c^2 S}\tag{4.4}$$

According to this elementary relationship, the sound pressure signal created by a volume source can be modelled by a simple spring two-port.

4.2 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 equation (1.14). With a proper representation of boundary conditions concerning geometry and impedances, it can be solved, at least in principle. It must be decided in which domain the results shall be discussed. For determination of modes, it is appropriate to study the sound field in space and to discuss the spatial sound pressure distribution. By considering a harmonic volume source excitation $q = \hat{q}e^{j\omega t}$, the wave equation can be transformed into the Helmholtz equation (stationary case):

$$\Delta p + k^2 p = -jkZ_0 q .\tag{4.5}$$

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}:

$$f_{lmn} = \frac{c}{2\pi} k_{lmn} \, . \tag{4.6}$$

Each eigenfrequency can be addressed to a specific standing wave form a mode. This solution is easily found for mathematically elementary geometries such as rectangular boxes, cylinders or spheres. For example, for a rectangular box of dimensions L_x, L_y and L_z and $Z=\infty$ throughout the boundaries, we find a set of orthogonal mode functions,

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

and

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

with l, m and $n \in N_0$.

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

$$\frac{dN_f}{df} \approx 4\pi V \frac{f^2}{c^3} \, , \tag{4.9}$$

where V denotes the room volume.

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. 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 2000):

$$p(\vec{r}) = jkqZ_0 \sum_l \sum_m \sum_n \frac{\psi_{lmn}(\vec{r})\psi_{lmn}(\vec{r}_S)}{K_{lmn}\left(k^2 - k_{lmn}^2\right)}, \tag{4.10}$$

where $\vec{r}$ and $\vec{r}_S$ denote vectors to the receiver and the (point) source positions, respectively.
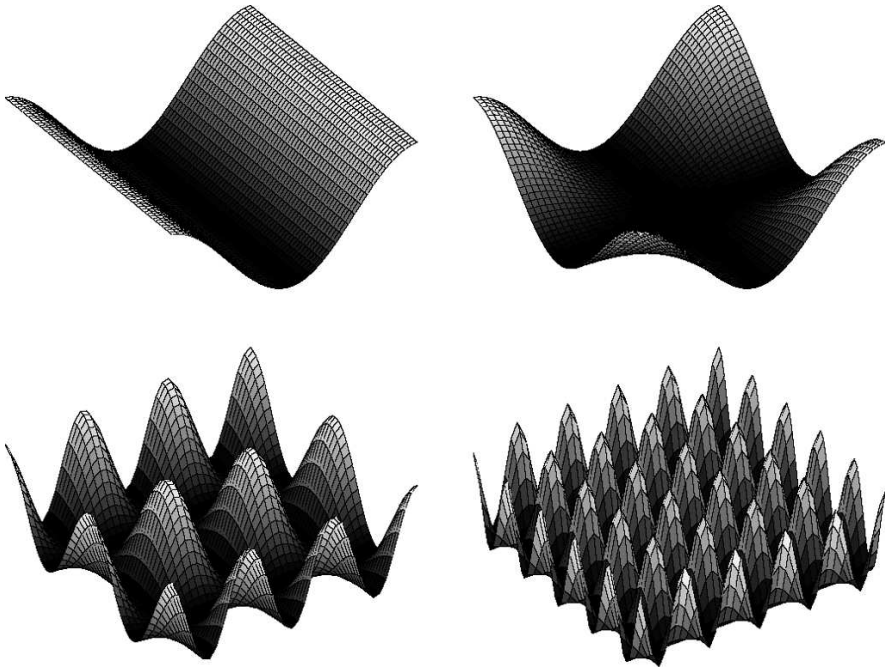
Fig. 4.3. Sound pressure distribution in 2-D modal fields: $(l,m) = (1,0)$, $(1,1)$, $(3,2)$ and $(5,5)$

4.2.1   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. 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 the solution without losses (see Sect. 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

$$\underline{k}_n = \frac{\omega_n}{c} + j\frac{\delta_n}{c} \,, \tag{4.11}$$
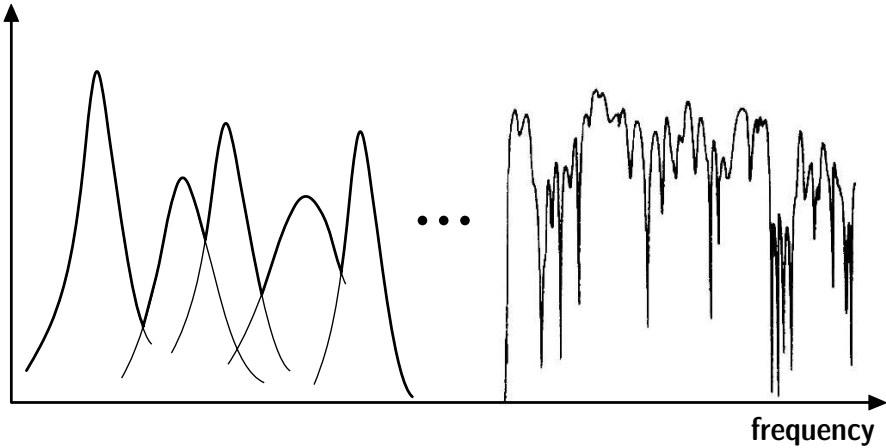
Fig. 4.4. Overlapping room modes at low and high frequencies

Eq. (4.10) changes into

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

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.

If, however, the resonances heavily overlap, i. e., when the frequency spacing is small compared with the mean half-width, a complicated super-position occurs. According to measurements by Kuttruff and Thiele (1954) and theoretical explanation by Schroeder (1954) and Schroeder and Kuttruff (1962) frequency curves in rooms are not useful 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. 4.4.1). The frequency range where this overlap is given, starts above

$$f_S \gg \sqrt{\frac{c^3}{4V\langle\delta_{lmn}\rangle}} \approx 1200\sqrt{\frac{T}{V}} , \tag{4.13}$$

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

Above the frequency limit,[8] 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 follow a Rayleigh distribution, for instance.

It is crucial to recognise 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 and, thus, 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 are crucial to describe the feedback condition.

Our listening experience in rooms is hardly related to comparing sound pressure at different locations anyway. In contrast, 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 geometric room acoustics.

4.3 Geometrical acoustics

At first, we introduce the concept of a "sound ray" or "sound particle." A sound ray is a thin line of sound energy transport. Accordingly, the energy transport of the sound wave can be addressed to a particle. 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. Definition of a ray

---

[8] which typically is very low (< 100).

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 to a relative phase shift, the superposition of the broadband sound pressures simplifies to energy addition,

$$\left|p\right|^2 \approx \left|\hat{p}_1{}^2\right| + \left|\hat{p}_2{}^2\right|. \tag{4.14}$$

The reason is found in the result of integration over frequencies (broadband signals), where the latter term vanishes:

$$\left|p\right|^2 = \left|\hat{p}_1 + \hat{p}_2\,\mathrm{e}^{jk\Delta x}\right|^2 = \left|\hat{p}_1{}^2\right| + \left|\hat{p}_2{}^2\right| + 2\left|\hat{p}_1\right\|\hat{p}_2\right|\cos\!\left(\omega\frac{\Delta x}{c}\right).$$
$$\approx \left|\hat{p}_1{}^2\right| + \left|\hat{p}_2{}^2\right| \tag{4.15}$$

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

$$\left|\bar{I}\right| = \left|\bar{I}_1\right| + \left|\bar{I}_2\right|. \tag{4.16}$$

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

$$w = w_1 + w_2, \tag{4.17}$$

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.

4.4 Statistical reverberation theory

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

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

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

The effective absorption coefficient, $\alpha_{diffuse}$, is then

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

with $\alpha(\vartheta)$ denoting the angular-dependent absorption coefficient. For locally reacting absorbers, $\underline{Z}(\vartheta) = const. = Z$. For real Z, Eq. (4.20) holds

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

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

$$\alpha_{diffuse} = 8\left[\frac{1-\sqrt{1-\alpha_0}}{1+\sqrt{1-\alpha_0}}\right]^2\left[\frac{2}{1-\sqrt{1-\alpha_0}} - \frac{1-\sqrt{1-\alpha_0}}{2} + 2\ln\left(\frac{1-\sqrt{1-\alpha_0}}{2}\right)\right] \cdot$$
(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 compared with the volume, the greater the energy exchange between the boundary and the exterior space. Thus,

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

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

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

the mean free path.

4.4.1   Reverberation

When sound power is fed continuously into a room and the source is suddenly switched off, the sound will not stop rapidly but decays. This observation is of importance for checking 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 derive the reverberation equation we consider 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 loss[9] of α, thus remaining with the energy $e_0(1-\alpha)$. After n reflections (after a time $t = n/\bar{n}$), the remaining energy of one ray is

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

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

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

According to the definition of the sound pressure level, the decay of sound energy density can be written in terms of a linear level decay.

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

The reverberation is traditionally described by a specific decay time, the time required for a level decrease of 60 dB. This time is denoted reverberation time, T:

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

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

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

---

[9] the average α for the diffuse field.

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

$$T = 0.16\frac{V}{S\alpha} = 0.16\frac{V}{A} \tag{4.29}$$

(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:

$$\bar{\alpha} = \frac{A}{S} = \frac{1}{S}\sum_{i=1}^{N}S_i\alpha_i; \quad S = \sum_{i=1}^{N}S_i; \quad A = \sum_{i=1}^{N}S_i\alpha_i. \tag{4.30}$$

The quantity A is denoted equivalent absorption area.

Now, we consider 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, air attenuation will also contribute to energy losses.

According to Eq. (3.33), the energy loss is given by e^{-mct}. The energy loss of sound rays, therefore, can be expanded to

$$e(t) = e_0 e^{[\bar{n}\ln(1-\alpha)-mc]t} \tag{4.31}$$
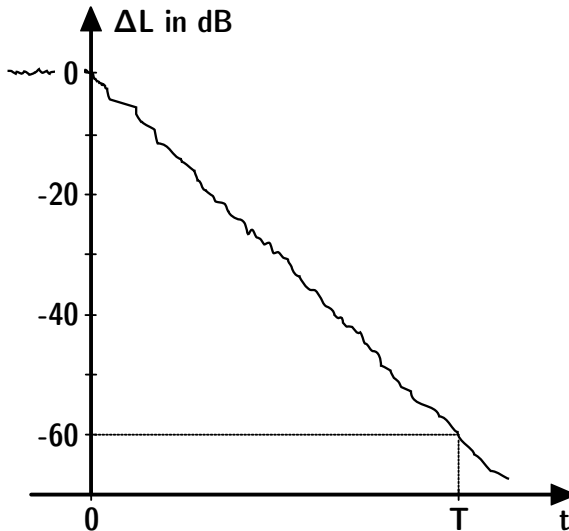
Fig. 4.6. Decay curve. Level drop $\Delta L = L_0 - L(t)$ after switch-off at $t = 0$

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

$$T = 0{,}16 \frac{V}{-S \ln(1-\alpha) + 4mV} \, , \tag{4.32}$$

$$T = 0{,}16 \frac{V}{A} \, , \tag{4.33}$$

with

$$A = \sum_{i=1}^{N} S_i \alpha_i + 4mV \, . \tag{4.34}$$

4.4.2 Steady-state energy density and level

Imagine that a sound source is radiating sound power P in a steady state. Sound rays are emitted that travel between the room boundaries and contribute to the total energy density. The first-order reflected sound and each later reflected component will completely add up – any time. The temporal series of reflections as described by Eq. (4.31) is radiated from the source with corresponding (reverse) delay: The later components at earlier times than the earlier components, just accounting for the travel time t' for each contribution, so that they coincide at 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 normalisation where $e_0 = P\mathrm{d}t$, it reads (in this case $P(t) = \mathrm{const}$)

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

The solution of this integral is

$$w_{\mathrm{diffuse}} = \frac{4P}{cA} e^{-A/S} \tag{4.36}$$

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. It can be interpreted either as

$$A_{\mathrm{Sab}} = \sum_{i=1}^{N} S_i \alpha_i + 4mV \tag{4.37}$$

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

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

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

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

The energy density accounts for the room reflections. Independently, we must consider the primary sound, called "direct sound," to obtain the total energy density in a steady state.

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

$$w_{\mathrm{direct}} = \frac{P}{4\pi c r^2} |\Gamma|^2_{\mathrm{max}=1} .$$ (4.40)
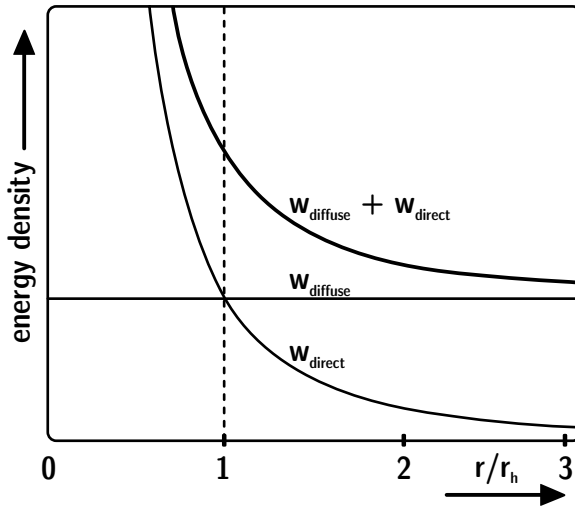
Fig. 4.7. Sound energy density in rooms with diffuse fields

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}:

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

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

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

Nondiffuse room sound fields

In nondiffuse room sound fields that are found in flat or long rooms or in coupled rooms, the sound pressure level can also be calculated from the sound power. The room surfaces, their absorption and scattering and, furthermore, the obstacle scattering of room fittings have a crucial influence on sound decay and level. Both quantities depend on the distance between source and receiver. It is important to know that in these rooms neither the decay curve is straight nor the room level is constant. More detailed information is given in (Kuttruff 2000).

The table below summarizes equations available for predicting 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 evaluating simulation models such as ray tracing (see Chap. 11). With these equations, we can also 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. It will be shown later that such an approach leads to artificial sound effects sufficient for sonification in computer games, for instance. The details of room acoustics as they relate to the specific physical phenomena and to aspects of perception, however, are not taken into account sufficiently, and better simulation models are required.
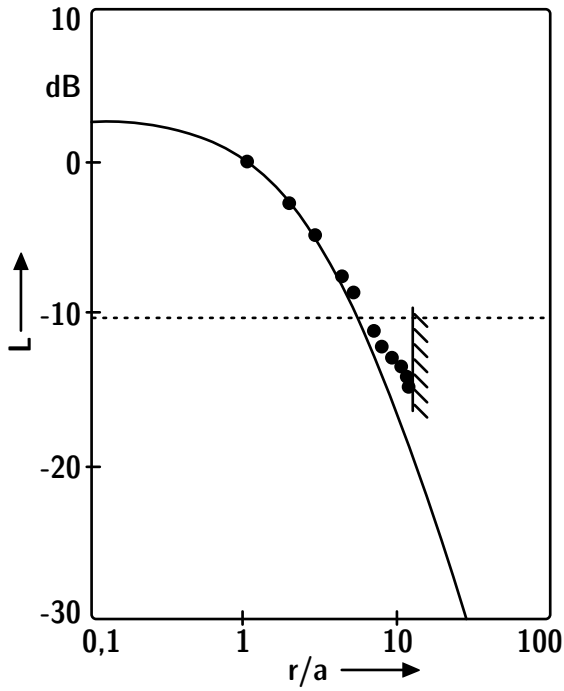
Fig. 4.8. Sound level in a flat room with height *a* and diffusely reflecting floor and ceiling (after (Kuttruff 1985; Vorländer 1988)). Analytic results in the solid curve compared with ray tracing results are shown. Level plot relative to direct sound in distance *a*/2. The side walls are located at *r*/*a* = 12.5

Table 4.1. Equations for calculating sound propagation curves in nondiffuse 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

$$w = \frac{P}{4\pi c} \sum_{k=1}^{\infty} \left(\frac{1/\rho_c + 1/\rho_f}{r_{2k-1}^2} + \frac{2}{r_{2k}^2} \right) (\rho_c \rho_f)^k$$

with $r_k = \sqrt{k^2 a^2 + r^2}$ and $r =$ source-receiver distance.
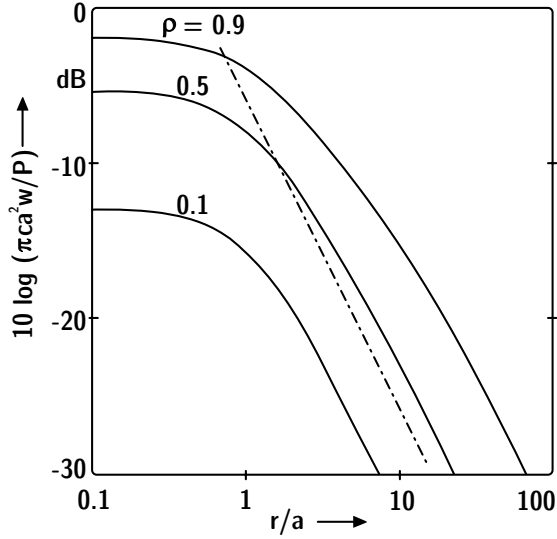
as above, $r \gg a$

$$w \approx \frac{P}{4\pi c r^2} \frac{\rho_c + \rho_f + 2\rho_c \rho_f}{1 - \rho_c \rho_f}$$

Table 4.1. (continued)

Room type	
Flat room of height a with scattering ceiling or floor, $r \gg a$	
Flat room of height a with scattering ceiling and floor, $r \gg a$	
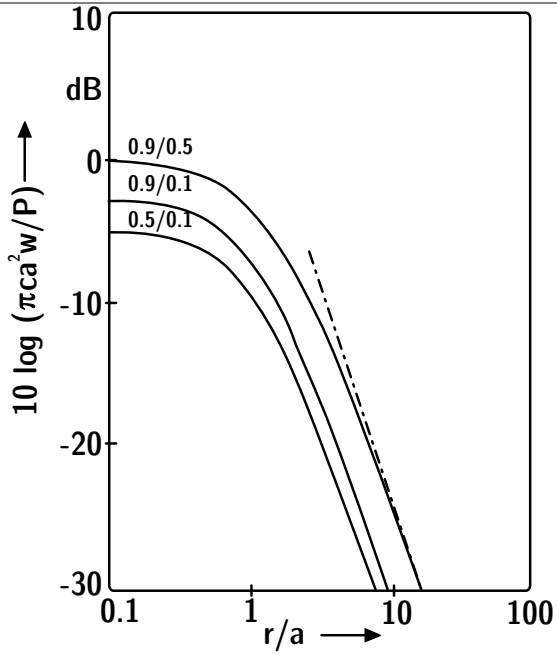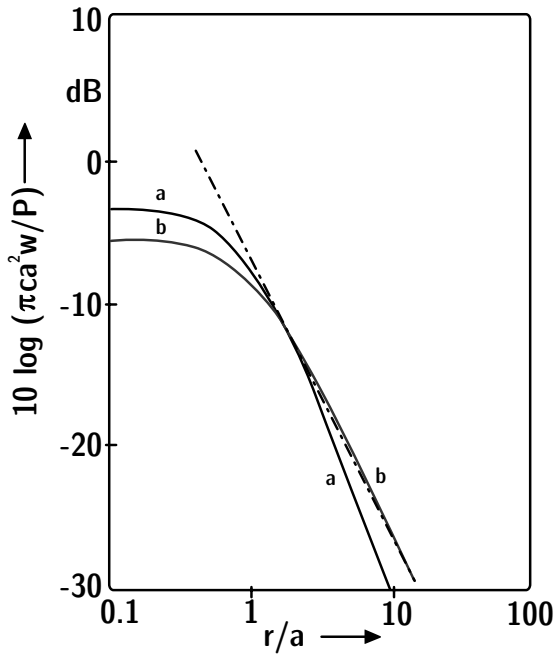
5   Structure-borne sound

The field of structure-borne sound is more complex than that of fluid acoustics. The reason is seen in the manifold possible degrees of freedom for force and torque and corresponding elastic reactions. In contrast to fluids, solids can react with elastic forces on 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.

Characterization of structure-borne sources is a very complex field on its own. The active processes of vibration generation must now be described mathematically. Additionally, 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. But in complex systems such as machines or buildings, this simple approach often is not sufficient since the definitions of force sources and impedances cannot be related to point contacts, but to multidimensional coupling with coherent or incoherent cross talk.

5.1   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 and Heckl 1973). The elastic constants for these strains are denoted σ and τ; the indices indicate 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, $\bar{s}$, and its components ξ, η, ζ; all are

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

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

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

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

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

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

$$c_{\mathrm{L}} = \sqrt{\frac{2\mu + \lambda}{\rho_0}}.$$
(5.4)

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

$$c_{\mathrm{T}} = \sqrt{\frac{\mu}{\rho_0}}.$$
(5.5)

Because the medium's reactive force and elasticity are less for shear forces than for compression forces, the speed of transversal waves is less than that of longitudinal waves. Figure 5.2 illustrates the direction of the displacement in these examples for longitudinal and transversal waves.

$\sigma_{31} + \dfrac{\partial \sigma_{31}}{\partial x_3} dx_3$

dx_1

dx_2

σ_{11}

f_3

σ_{21}

f_2

f_1

$\sigma_{21} + \dfrac{\partial \sigma_{21}}{\partial x_2} dx_2$

dx_3

$\sigma_{11} + \dfrac{\partial \sigma_{11}}{\partial x_1} dx_1$

σ_{31}

x_3

x_1

x_2
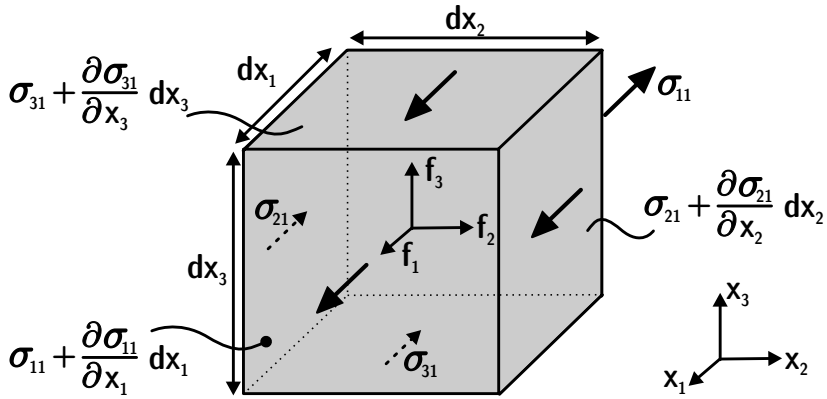
Fig. 5.1. Notation for coordinates in a solid medium
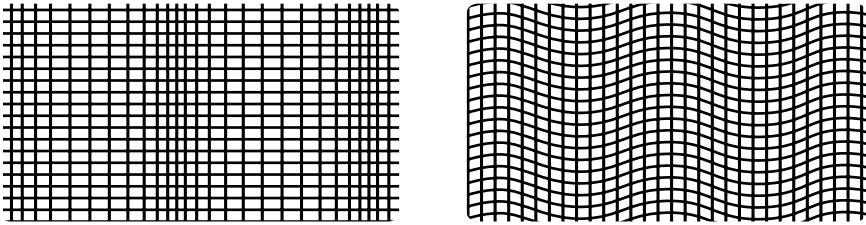
Fig. 5.2. Waves in solid media

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	6,420	3,040
Stainless steel	7.9	0.30	200	5,790	3,100
Copper	8.9	0.37	125	5,010	2,270
Brass	8.6	0.374	105	4,700	2,110
Lead	11.3	0.43	16	1,960	690
Glazing	3.9	0.224	54	3,980	2,380
Plexiglass	1.2	0.40	4	2,680	1,100

5.2   Waves on plates and their radiation

The wave forms in plates and beams are quite similar to those in un-bounded 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.

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.

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

$$\sin \vartheta = \frac{\lambda}{\lambda_{B}} = \frac{c}{c_{B}} \,.$$

(5.6)

This equation holds only 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

$$\omega = k_B^2 \sqrt[4]{\frac{B}{\rho}} \,,$$

(5.7)

where B denotes the bending stiffness of the plate,

$$B = \frac{h^2}{12} \cdot \frac{E}{1 - \gamma^2} \,,$$

(5.8)
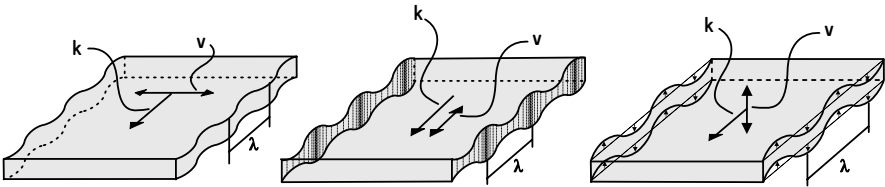
Fig. 5.3. Waves on plates
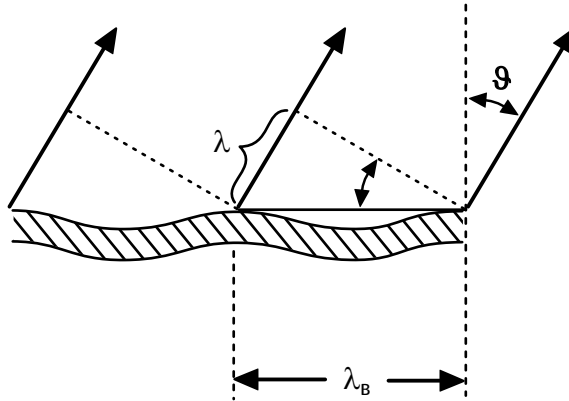
Fig. 5.4. Bending wave interaction with airborne sound

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

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

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

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

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

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

which also is a requirement for $\sin\vartheta \leq 1$ 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

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

where m'' denotes the mass per unit area in kg/m^2 and B' the bending stiffness in Nm normalized to the thickness h.
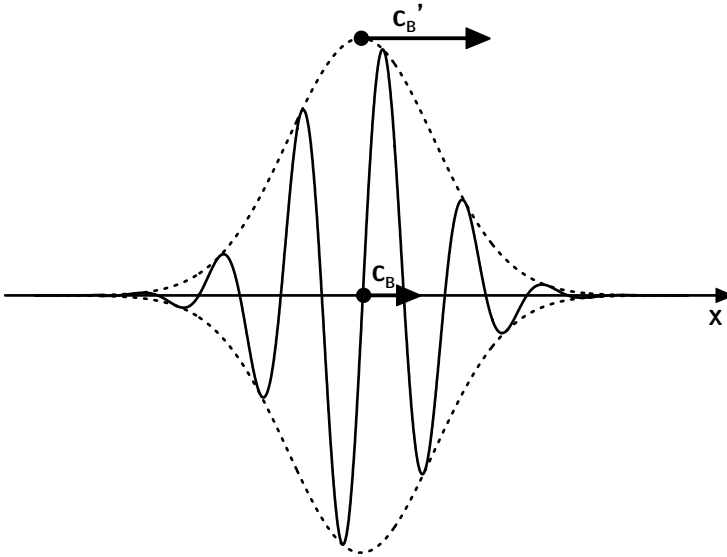
Fig. 5.5. Group velocity and phase velocity in bending waves, $c'_\mathrm{B} = 2\,c_\mathrm{B}$

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 identify the plate velocity component $j\omega\zeta$ as identical to the velocity of the airborne wave, $v\cos\vartheta$. This yields

$$\left|\bar{I}\right| = \frac{1}{2}\rho_0 c\left|v\right|^2 = 2\pi^2 \rho_0 c\left|\zeta\right|^2 \frac{f^3}{f - f_\mathrm{c}} . \tag{5.13}$$

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

An important quantity for describing the total plate as an airborne sound source is the radiation efficiency. It is defined as 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, $\widetilde{v}^2$,

$$\sigma = \frac{P}{\rho_0 c S \widetilde{v}^2} . \tag{5.14}$$

5.2.1   Finite-size plates

To study elastic waves on finite-size plates, we start with the differential equations and wave types for the unbounded case (Fig. 5.3). The boundary condition is then introduced for the plate support at its perimeter, such as "simply supported," "clamped" or "free" (Cremer and Heckl 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 location. For a simply supported plate, the velocity and the acceleration at the edges must vanish. The eigenfunctions accordingly are similar to those 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 a forced response, the plate follows the airborne excitation in wavelength, phase and angle. In a 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).

5.2.2   Internal losses and structural reverberation time

Damping of waves on plates can be caused by friction and viscoelastic 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 conditions or junctions, such as 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.

$$T_s = \frac{2,2}{f\eta} \tag{5.15}$$

where f denotes 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. For a four-sided plate,

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

5.3   Vibrational transmission over junctions

A crucial quantity, therefore, is the velocity amplitude of bending waves of plates interacting with air. In 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, $+$, the L-junction, $\llcorner$, and the T-junction, $\top$. 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 reflection and transmission. Energy will be transferred and exchanged according to the degrees of freedom of each displacement wave type. A pure bending wave travelling into a junction will result in a mixed bending/longitudinal wave in the reflection and in the transmitted domain, too.
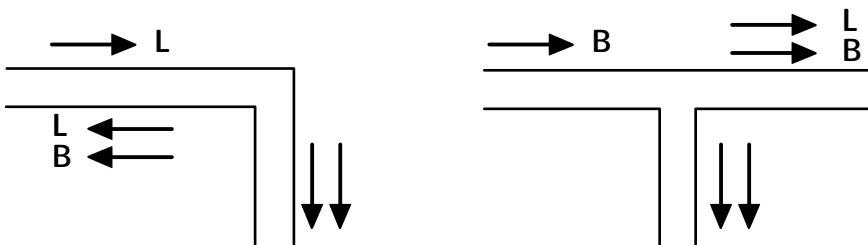
Fig. 5.6. Examples of wave type conversion of longitudinal (L) and bending (B) waves at junctions

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

$$K_{ij} = \frac{D_{v,ij} + D_{v,ji}}{2} + 10 \log \frac{l_{ij}}{\sqrt{a_i a_j}} \, . \tag{5.17}$$

$D_{v,ij}$ 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}. a_i and a_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 homogenous thin plate structures. In practice, however, multilayered 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 applied to describe the junction transmission loss of the specific case.

6 Psychoacoustics

A proper physical foundation of acoustics is essential for understanding the generation and transmission of sound. With respect to simulation and auralization, however, humans 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 will, thus, give significant insight into the way auralization must be organized and programmed, how exact it must be, which acoustic aspects are essential and which can be neglected in modelling.

The term psychoacoustics involves the description and modelling of human hearing. Psychoacoustic model functions try to extract the characteristic data related to specific hearing dimensions, such as loudness, sharpness, impulsiveness or spaciousness from physical data such as sound pressure time functions or spectra. These dimensions are straightforward physical descriptors, which as a set form the "character" of sound. Finally, involving also neural top-down processes, using our memory, what we have learned and considering what we expect, 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 judgment must be taken into account (Blauert 2005).

6.1 Anatomy of the peripheral hearing system

Knowledge of the peripheral hearing organ and corresponding signal processing leads 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 right ear – lead to special, so-called "binaural" and "interaural" cues such as source localization, precedence effect and spaciousness (see below).

Binaural hearing (Sect. 6.3) is a specific discipline in psychoacoustics. For a realistic or at least a plausible auralization result, it is essential. The consequences of binaural hearing are directional localization and distance perception. Furthermore, binaural hearing enables very effective noise
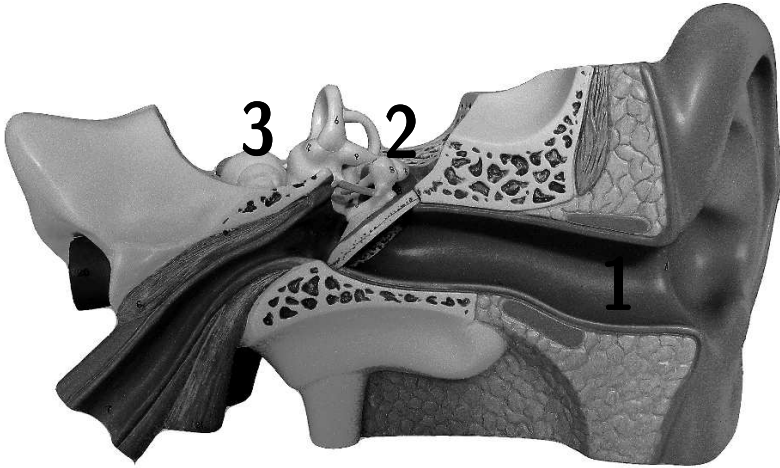
Fig. 6.1. Human hearing organ

suppression for understanding 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 field,[10] 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. The outer ear includes the pinna and the ear canal. The latter is a tube about 2.7 cm long and 6 to 8 mm in diameter. 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. Hearing sensitivity has its maximum in this frequency range. 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 into 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 as a force transducer lever, but as 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.

---

[10] with reference to the sound field with the head absent.

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 separated into two parts by several membranes. The basilar membrane plays a central role in the transduction of vibroacoustic stimuli into neural spike rate signals. It has a significant variation of width and stiffness along its length. The stiffness in the part near the oval window is greater than that toward 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 the 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 the location of its 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 higher than 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 dental drill).

6.2 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.

6.2.1 Loudness

A first-order approximation for many human sensory perceptions is that they are proportional to the relative change in measurable physical stimuli (law of Weber and Fechner). The loudness impression should then be represented by a logarithmic scale of sound pressure, thus, in the 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 a narrow band).

Equal loudness in wide frequency ranges is related to a frequency-dependent physical stimulus of the 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." F or a pure tone of 1 kHz, the decibel value is identical to the phon value.

A better estimation of the subjective impression of loudness is given by the scale in units of sone. One sone corresponds to the loudness of a 1 kHz pure tone at 40 dB sound pressure level. From 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
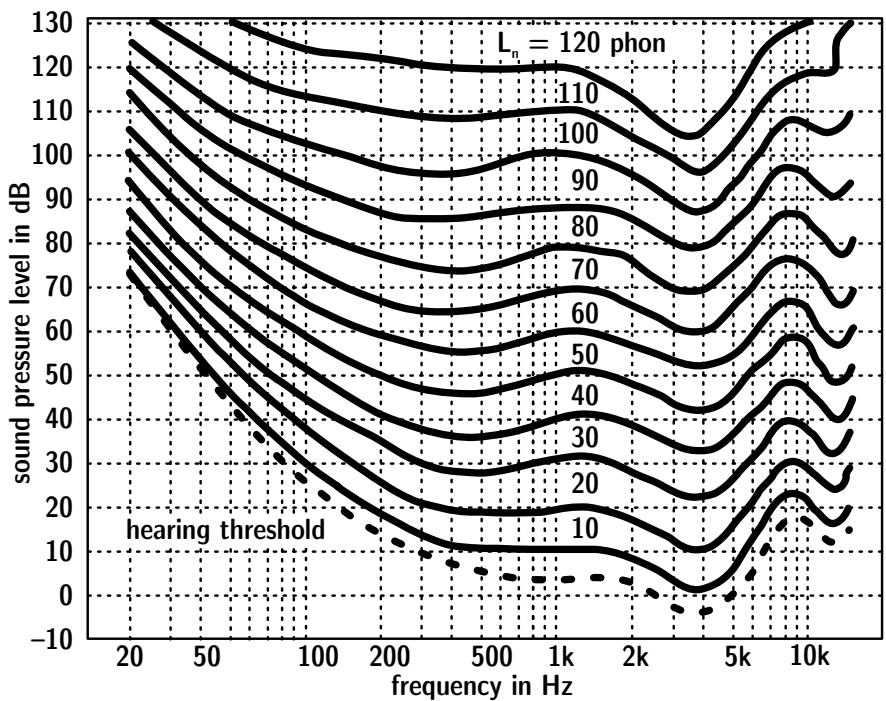
Fig. 6.2. Equal loudness level contours
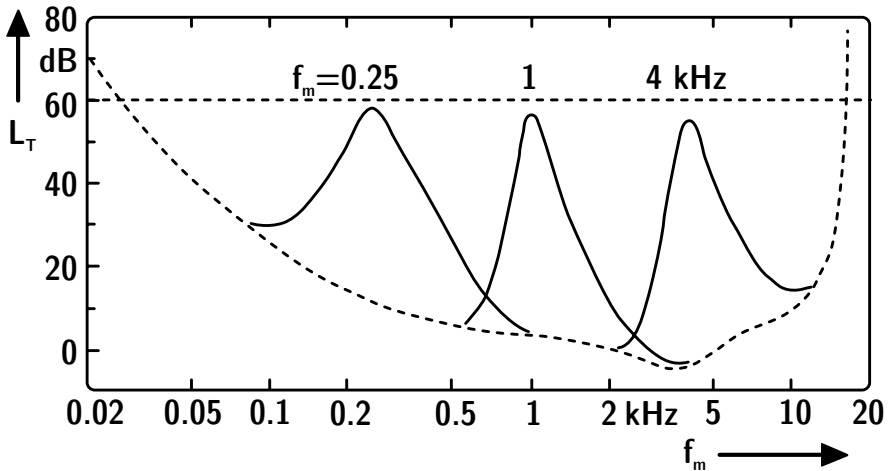
Fig. 6.3. Masking curves (shifted hearing threshold). A masker is a narrowband noise at midband frequencies of 250 Hz, 1 kHz and 4 kHz. A test tone is a pure tone with a 60-dB sound pressure level (after (Zwicker 1999))

so-called A-weighting and the level measured with an A-filter are expressed in dB(A). To express the ear's frequency sensitivity at higher levels, B, C and D filters were defined as well. 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 cannot model the subjective impressions sufficiently. Another consequence of nonlinear cochlea mechanics is so-called "masking."

Sound in one frequency band may mask sound in adjacent bands. This phenomenon is called "simultaneous masking." Determination of the A-weighted sound pressure is not sufficient to describe the perceived loudness. Better results are obtained by separating the sound spectrum into critical bands before calculating the specific loudness in each band and accounting for the masking curves between the bands (Zwicker 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 required to perceive a tone while another tone or masking sound is present. It is a threshold as well, but related to simultaneous masking sound (tone, noise, …).

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

6.2.2 Temporal masking

With rapidly changing transient sound such as a series of impulses, for instance, masking is present, too. After excitation of 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 "postmasking." It can easily be understood by this cochlear effect. Premasking also exists, which means that a previous impulsive sound may be masked by a later arriving loud sound. This effect, of course, contradicts cause and effect and can be explained only by a higher level neural process in sound evaluation in the brain. Obviously the priority of sound event processing is shifted toward a more "important" sound event, despite the ongoing processing of a previous, less intensive sound.

6.2.3 Time-varying loudness

The current standards and loudness models refer to stationary sounds. Fluctuating sound is 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. Models for future standards are being developed.[11]

6.2.4 Sharpness

The sensation of sharpness is a significant subjective impression and independent dimension of sound character. 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.

$$S = 0,11 \frac{\int\limits_{0}^{24\text{Bark}} N' g(z)\, z\, \mathrm{d}z}{\int\limits_{0}^{24\text{Bark}} N'\, \mathrm{d}z} \text{ acum} \tag{6.1}$$

---

[11] At the stage of publication of this book, the German standard DIN 45631 – Amendment 1 is in the voting process.

where S denotes the sharpness, N' the specific loudness on the Bark scale(z) and $g(z)$ is a weighting factor.

6.2.5 Fluctuation strength

Slow amplitude modulations cause perception of fluctuations. To be identified as a modulation of a clearly perceivable steady sound, a fluctuation should not be faster than 20 Hz. Modulations higher than 20 Hz will create the sensation of roughness (see below). At about 4 Hz, the ear has the 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% at 4 Hz. This corresponds to a fluctuation strength of 1 vacil, according to the following equation:

$$F = \frac{0{,}008 \int\limits_{0}^{24\text{Bark}} (\Delta L/\text{dB})/\text{Bark}\,\mathrm{d}z}{(f_{\text{mod}}/4\text{Hz}) + (4\text{Hz}/f_{\text{mod}})}\,\text{vacil}. \tag{6.2}$$

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

6.2.6 Roughness

Fast amplitude modulations cannot be perceived as fluctuations. Instead, their effect is best 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 clearly separated 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.

6.2.7 Tonality, pitch, pitch strength

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

Tonal components are recognized immediately. The human hearing system extracts tonal components from a complex mixed sound with great

efficiency. Also, when humans are listening to a complex sound, their attention is directed toward the tonal component. Tonality can be described by the prominence ratio or by spectral level difference between tonal and broadband components (DIN 45681). The tonal character determined from 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 that is extremely difficult to characterize 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, "strong." Other sounds such as narrow-band noise or high-pass noise limited to 1 kHz, have a certain tonal character, but not the same strength as a pure tone. It is confusing that these sounds mentioned are not periodic at all. It is therefore necessary to study 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 more 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 of the maximum spectral amplitude (pitch) and its variance (pitch strength). For details and current results, see (Fruhmann 2004a, 2004b).

6.3   Binaural hearing

Humans can localize sound due to binaural hearing. 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 thus is influenced by linear distortions. These distortions are dependent on direction. In the case of spherical waves, distortions are dependent on the distance, too. At the two ears, the sound signals arrive with differences in time and amplitude. Sound from a source located at the side of the head travels a longer time to the contralateral ear[12] and suffers frequency-dependent damping due to diffraction and absorption. Both effects are noticeable as differences between the ear signals, as interaural time differences (ITD) and interaural level differences (ILD).

---

[12] The ear in the shadow zone of the head.

In the median plane (the vertical plane between the ears along the fontanelle), ILD and ITD are small. Therefore the precision of localisation is less than with clear interaural cues. Nevertheless, humans can distinguish between frontal, up, or back direction, due to 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.

6.3.1 Head-related transfer functions

Monaural and binaural cues are introduced into the eardrum sound pressure by diffraction of sound incident on the head and torso. Usually a plane wave is considered as reference. The amount of diffraction is described by the head-related transfer function (HRTF), the details of which were discovered in pioneering work by (Shaw 1982). It is defined by the sound pressure measured at the eardrum or at the ear canal entrance divided by the sound pressure measured with a microphone at the centre of the head but with the head absent. Accordingly, HRTF is dependent on the direction of sound incidence.
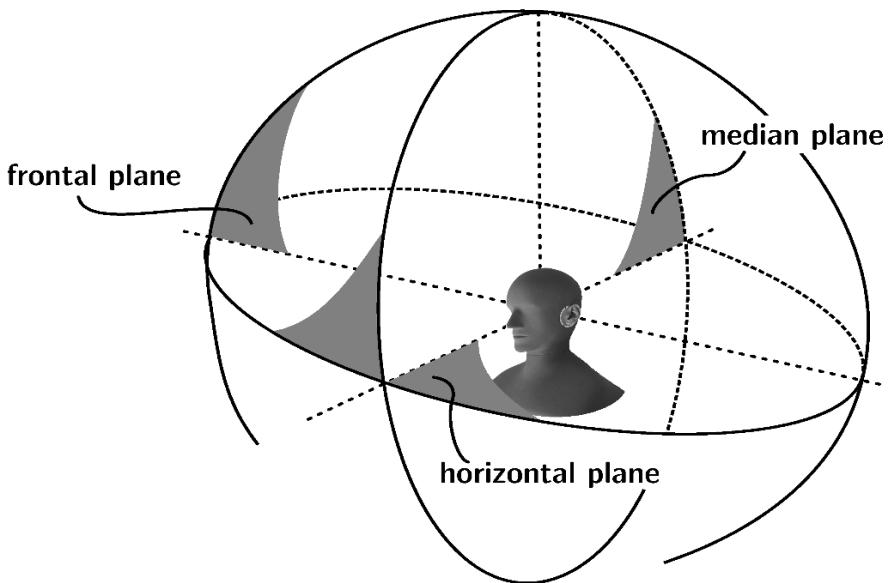
Fig. 6.4. Head-related coordinate system

The principal components of the HRTF are in a frequency range higher than 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). Note that the HRTF is dependent on individual anatomic dimensions.

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

In Fig. 6.5, some typical HRTFs are illustrated. They were measured with four test subjects, frontal incidence. In Fig. 6.6, an example is shown of an 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 the right (grey) ear. Furthermore it can be seen in the impulse responses 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).
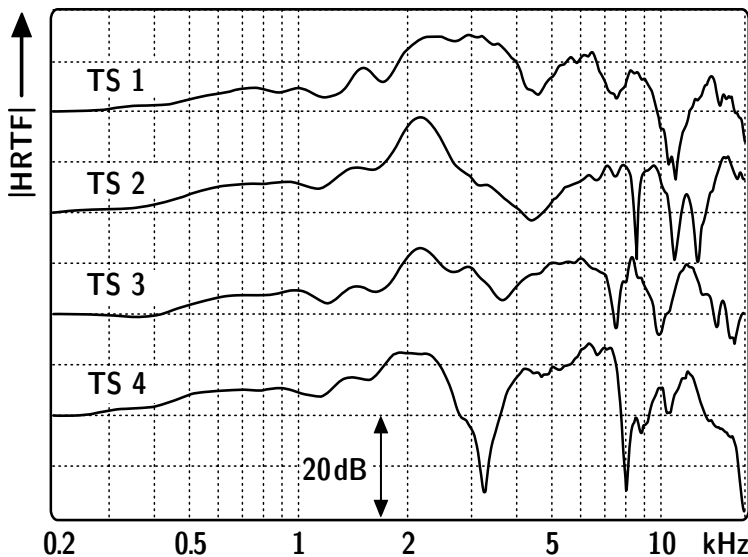
Fig. 6.5. Head-related transfer functions of four human test subjects, frontal incidence
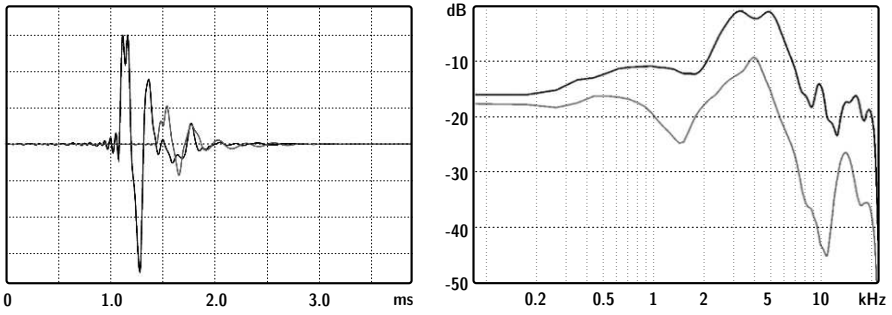
Fig. 6.6. Head-related transfer function (right) for sound incidence from the left and corresponding head-related impulse response HRIR (left)

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° in the frontal direction ($\varphi = 0°$), 10° on the right/left axis ($\varphi = 90°$ or 270°) and 5° in the back direction ($\varphi = 180°$). In the median plane at $\theta = 90°$, the jnd is about 20°.
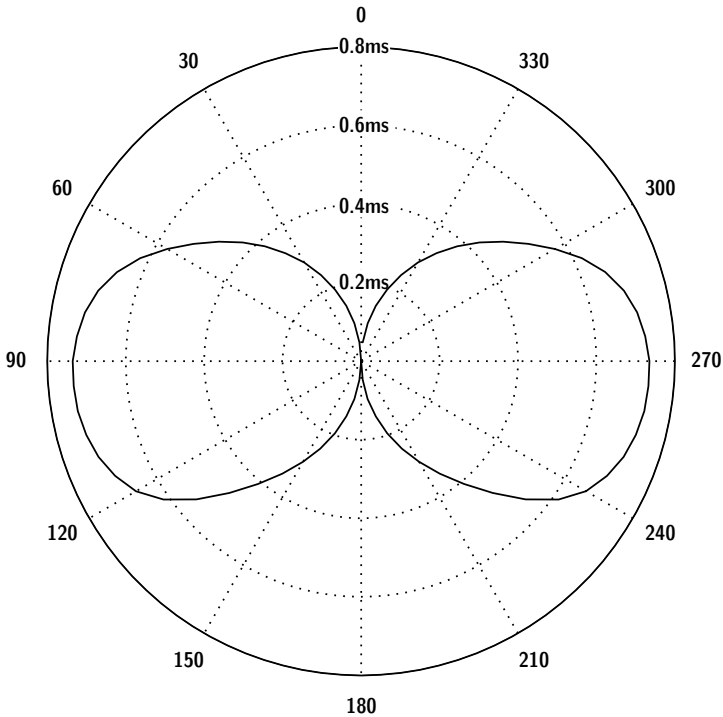
Fig. 6.7. An example of ITD for sound incidence in the horizontal plane

The reason for a more accurate localization in the horizontal plane is the availability of interaural time differences, whereas in the median plane, just monaural (spectral) cues are fed into the binaural hearing system. The gain in performance by spatial hearing due to interaural differences offers great advantages for localizing sources, for hearing in noisy environments in general and for noise suppression in particular (precedence effect (Blauert 1996) and "cocktail party effect"). The precedence effect (law of the first wave front) can be explained by temporal effects of neural processing with a preference for 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 the speech intelligibility of our specific communication partner is enhanced, just as the speech level was increased (binaural intelligibility level difference).[13]

6.3.2   Artificial heads

The first relevant studies aiming at a standard dummy head for measurement and simulation were presented in the 1960s by the pioneering work of Burkhard and Sachs (1975) and of Shaw (1982). In 1972, KEMAR was introduced. Its first main application was measurement of hearing aids under in situ conditions. KEMAR, however, was used as reference dummy head in a wider sense. Today, it still serves as one of the reference databases of the HRTF.[14]

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

---

[13] The specific communication partner need not be in the front direction.
[14] HRTF data can be found in the Internet, for example:
KEMAR http://sound.media.mit.edu/KEMAR.html
CIPIC   http://interface.cipic.ucdavis.edu
IRCAM http://www.ircam.fr/equipes/salles/listen/download.html
AUDIS http://www.eaa-fenestra.org/Products/Documenta/Publications/09-de2

Further research aimed at a technical description of head and torso geometry, however, with the same acoustic behaviour as natural replicas or 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 for correct localisation cues.

When using the standard dummy heads for recordings and for research in perception of direction and distance, a rather large portion of the test subject population reports disturbances of the listening experience. The test results typically suffer from front–back confusion and large uncertainties in localisation in the median plane. The reason can be found by multidimensional analysis (modulus, phase, frequency, ILD, ITD, …) of the HRTF, comparing individuals with dummy heads. There might be some individuals who match the dummy head performance quite well, while others' ears are "far away" from the dummy head's ears.

Approaches for developing best matched 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 success 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. Note that this approach is not based on a geometric or structural average but on a selected HRTF including significant HRTF details such as peaks and notches.

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 test subject population represented by any dummy head is a crucial process of selection. 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 has changed during the last three decades (typically the standard group of normal hearing subjects, adults of age 18–25 are taller than they were years ago). And it might be asked whether one standard dummy head might be sufficient to cover all humans.

Today, dummy heads and binaural technology are of high quality. International standards for telecommunication technology and for measuring hearing aids are available. They represent the state of the art published in

recent decades and the applications of dummy heads in research and development. Besides dummy head recordings and reproduction for home entertainment, which were not too successful commercially, several fields such as 3-D audio, virtual environments, headphone development and sound quality could be created and extended with high relevance in applied acoustics and audio engineering. However, some questions still remain, related to the applicability of average head dimensions for both sexes, humans in all parts of the world, and particularly for children.[15]

Head movements

Head movements are essential if front–back confusion is affecting the precision in localizing the source (Börger et al. 1977; Mackensen 2003). The possibility of moving and rotating the head slightly to enhance the localization precision must therefore be provided in virtual sound environments; see Sect. 15.1.1. Current research is thus focused on the dynamic features in the HRTF by considering relative head and torso movement, such as in normal behaviour, to enhance sound localization accuracy. The degrees of freedom are then extended toward a larger number of spherical coordinate systems related to head orientation in horizontal and vertical axes and the torso orientation (Moldrzyk et al. 2004).

6.4   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 multidimensional and influenced by many factors, some of which are non-acoustical. Psychological room acoustics can be understood as a bridge 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 important for describing the overall subjective effect in rooms. Unlike reverberation, they depend on the specific listener's 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

---

[15]   which is relevant for fitting hearing aids, for example.
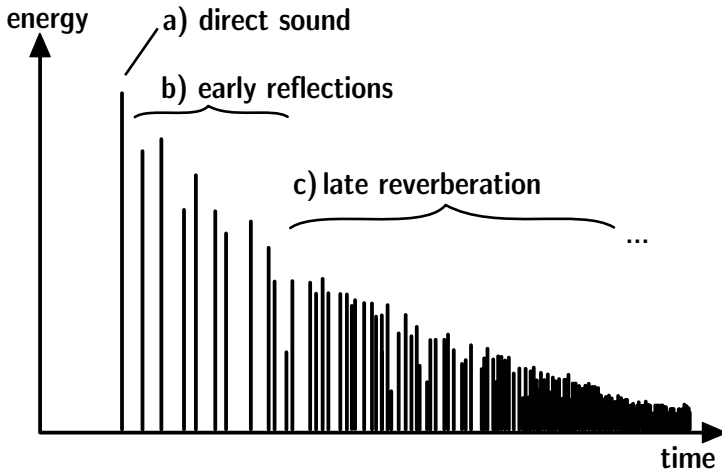
Fig. 6.8. Energy vs. time curve, Energetic room impulse response

delayed reflections, too. The response of the room to excitation with an ideal impulse, therefore, serves well as a basis for interpretation of room acoustics. A typical example of a room impulse response ("energy time curve") is illustrated in Fig. 6.8.

The components in the energy time curve are arranged in a characteristic series of pulses. The pulse density grows 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 cancel 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 a higher level than the direct sound (up to 10 dB), which might happen with secondary loudspeaker sound or when the direct sound is blocked by a barrier, the localisation 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, 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.

6.4.1 Reverberance

The sensation of reverberance must be seen in relation to the signal presented. Music and speech do have transient components. However, the mean syllable duration and also the mean tempo of music does not permit following 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. For this reason, several reverberation times were introduced which evaluate early parts of the energy impulse response[16] curve (Fig. 6.8).

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 in relation to 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 a steady-state signal can be understood as a permanent excitation with densely repeated pulses. Thus, the response to steady-state excitation is the sum of all components of the impulse decay. In the stationary case, there is direct sound 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 right to arrive in time with the direct sound). The same holds for all reflections. Retardation and coincidence[17] 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. But then, the energy from the temporal series of reflections will vanish one by one. This process is modelled by the so-called "integrated impulse response" (Schroeder 1965). It consists of two steps (see Fig. 6.9):

---

[16] The term "impulse response" will be introduced in Sect. 7.2.2 by using signal theory.

[17] delayed components vs. coincident components.

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

$$C = h^2(t < 0) = N_0 \int_0^\infty p^2(\tau)\mathrm{d}\tau . \tag{6.3}$$

b) Switch-off: Subtraction

$$h^2(t) = C - N_0 \int_0^t p^2(\tau)\mathrm{d}\tau = N_0 \left[\int_0^\infty p^2(\tau)\mathrm{d}\tau - \int_0^t p^2(\tau)\mathrm{d}\tau \right]$$
$$= N_0 \int_t^\infty p^2(\tau)\mathrm{d}\tau \tag{6.4}$$

The reverberation time[18] can now be evaluated from the decay curve, h^2. And the perceived reverberance, too, correlated with the early decay time (EDT), can now be evaluated. It is defined as six times the time interval corresponding to the average decay between 0 dB and -10 dB (Atal et al. 1965).
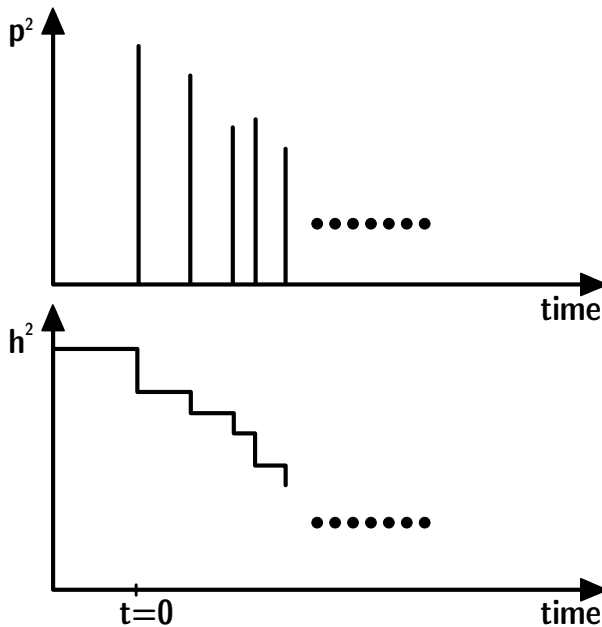
Fig. 6.9. Top: Energetic impulse response. Bottom: Integrated impulse response. Note the temporal coincidence of impulses and subtraction steps

---

[18] In measurements, the standard reverberation time is evaluated from a logarithmic slope regression between -5 dB and -35 dB.

6.4.2 Strength

As explained in the previous section, the total energy for the steady state 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 independently of the sound power of the source. By choosing a reference distance in a 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.

$$G = 10 \log \frac{\int\limits_{t_0}^{\infty} p^2(t)\mathrm{d}t}{\int\limits_{0}^{\infty} p_{10}^2(t)\mathrm{d}t} . \tag{6.5}$$

$p(t)$ denotes the sound pressure of the room impulse response and $p_{10}(t)$ the reference sound pressure at 10 m distance with the same source in a free field. The integration limit, t_0, is defined as the arrival time of the direct sound (including the direct sound).

6.4.3 Speech intelligibility and transparence

Because the early reflections support speech intelligibility, their energy sum can be used to characterize the syllable intelligibility. For simplicity, 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:

$$D = \frac{\int\limits_{0}^{50\,\mathrm{ms}} p^2(t)\mathrm{d}t}{\int\limits_{0}^{\infty} p^2(t)\mathrm{d}t} . \tag{6.6}$$

A similar expression is intended with the introduction of "clarity," 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 transparence in fast pieces of

music, the ability to recognise musical details behind the "curtain" of re-
verberation.

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

$$C_{50} = 10 \log \frac{\displaystyle\int_0^{50\text{ms}} p^2(t)\,dt}{\displaystyle\int_{50\text{ms}}^{\infty} p^2(t)\,dt}. \tag{6.8}$$

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

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

The balance between the early and late parts of the impulse response can
also be expressed by using the first moment of the impulse response. This
way we define the "centre time," T_S:

$$T_S = \frac{\displaystyle\int_0^{\infty} t\,p^2(t)\,dt}{\displaystyle\int_0^{\infty} p^2(t)\,dt}. \tag{6.10}$$

Definition, clarity and centre time are interrelated and correspond to
clarity of speech and music. Which parameter is best cannot be decided.
Extensive listening tests and questionnaires used in the laboratory and in
music halls have shown that all quantities are useful to describe this spe-
cific auditory impression of transparence related to the linguistically or
musically basic elements such as 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, independently. The so-called
"speech transmission index" (STI) expresses the degree of changes of
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 the high frequencies.

6.4.4 Spatial impression

The subjective impression of spaciousness was identified as 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. The early part, up to 80 ms, contributes to the auditory source width (ASW). As described above, source localization in complex sound fields is evaluated by the human hearing system such 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):

$$
\mathrm{LF} = \frac{\displaystyle\int_{5\mathrm{ms}}^{80\mathrm{ms}} p_\infty^2(t)\,\mathrm{d}t}{\displaystyle\int_{0}^{80\mathrm{ms}} p^2(t)\,\mathrm{d}t} = \frac{\displaystyle\int_{5\mathrm{ms}}^{80\mathrm{ms}} \left(p\cos\vartheta\right)^2(t)\,\mathrm{d}t}{\displaystyle\int_{0}^{80\mathrm{ms}} p^2(t)\,\mathrm{d}t}\;,
$$

where p_∞ denotes the sound pressure impulse response obtained using a gradient (figure-of-eight) microphone oriented to the horizontal axis through the ears of the listener.
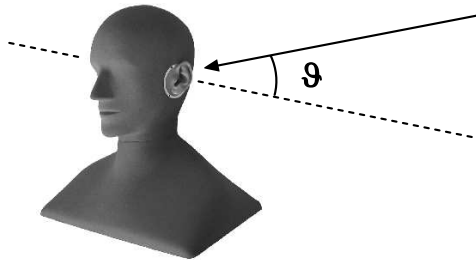
Fig. 6.10. Sound incidence from lateral reflections

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:

$$LG = 10 \log \frac{\int\limits_{80\text{ms}}^{\infty} p_\infty^2(t)\mathrm{d}t}{\int\limits_0^{\infty} p_{10}^2(t)\mathrm{d}t} = \frac{\int\limits_{80\text{ms}}^{\infty} (p\cos\vartheta)^2(t)\mathrm{d}t}{\int\limits_0^{\infty} p_{10}^2(t)\mathrm{d}t} . \tag{6.11}$$

When sound arrives at the listener's ears from lateral directions, it will lead to a loss in correlation in the binaural pattern. Particular 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 ears of a listener or a dummy head, p_r and p_l, respectively, is defined as follows:

$$IACF_{t_1,t_2}(\tau) = \frac{\int\limits_{t_1}^{t_2} p_l(t) \cdot p_r(t+\tau)\mathrm{d}t}{\sqrt{\int\limits_{t_1}^{t_2} p_l^2(t)\mathrm{d}t \int\limits_{t_1}^{t_2} p_r^2(t)\mathrm{d}t}} . \tag{6.12}$$

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

$$IACC_{t_1,t_2} = \max\left[IACF_{t_1,t_2}(\tau)\right]. \tag{6.13}$$

With the choice of the integration interval in Eq. (6.12), early and late lateral reflections can be treated separately, thus, leading 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 distinguish different situations. As in other psychoacoustic areas, the just noticeable differences (jnd) give important information about the precision necessary 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. Just noticeable differences of subjective room acoustical impressions (after (Vorländer 1995; ISO 3382-1)).

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

6.4.5   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 diffuse sound field in a room, this coherence between the two sound pressure signals, p_1 and p_2, at a distance d is (Kuttruff 2000)

$$\Psi_{p1p2}(x) = \frac{\sin(kd)}{kd} \, . \tag{6.14}$$

If we assume that our sound signal has a frequency content dominant at low to midfrequencies 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. On the 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 are, at least typically, of the order of magnitude of 1 m.

6.4.6   Estimation of the monaural subjective parameters

Before starting to discuss a detailed sound field simulation in Chap. 11, we estimate the order of magnitude of the subjective parameters in a room of volume, V, surface, S, and reverberation time, T. This kind of estimation gives no exact result but only the order of magnitude of room acoustic parameters. We might use this information as expectation values and as a reference for checking the plausibility of simulation results. Also, 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.

For an approximate approach, assuming a diffuse sound field and the prerequisite exponential decay function (Sect. 4.4.1), the strength and other parameters in a room can be estimated roughly. The lower integration limit of the series of reflections is set as an average to $t_0 = 1/\overline{n}$. 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 agreement with Eq. (4.39), we find for the strength

$$G = 10\log\left(\frac{1}{r^2} + 310\frac{T}{V}\right) + 4{,}34A/S + 20 \, ,$$

$$G = 10\log\left(\frac{1}{r^2} + \frac{50}{A}\right) + 4{,}34A/S + 20 \, ,$$

(6.15)

or as an approximation for large distances r,

$$G = 37 - 10\log A + 4{,}34A/S, \quad r \gg \sqrt{A}/7 \, .$$

(6.16)

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

$$w_{\text{late}} = \frac{1}{V}\int_{0.05 + 1/\overline{n}}^{\infty} e^{-13.8t/T}\mathrm{d}t = \frac{T}{13.8V}e^{-\left(\frac{276}{T} + A/S\right)} \, ,$$

(6.17)

$$w_{\text{early}} = \frac{1}{V} \int_{1/\bar{n}}^{1/\bar{n}+0.05} e^{-13.8t/T}\, \mathrm{d}t + \frac{1}{4\pi c r^2}$$

$$= \frac{T}{13.8V}\left(e^{-A/S} - e^{-\left(\frac{276}{T}+A/S\right)} \right) + \frac{1}{4\pi c r^2} \tag{6.18}$$

Accordingly, definition, clarity and centre time result in

$$D = \frac{\dfrac{T}{13.8V}\left(e^{-A/S} - e^{-\left(\frac{276}{T}+A/S\right)} \right) + \dfrac{1}{4\pi c r^2}}{\dfrac{T}{13.8V}e^{-A/S} + \dfrac{1}{4\pi c r^2}},$$

$$\approx \frac{\dfrac{T}{13.8V}\left(1 - e^{-\frac{276}{T}} \right) + \dfrac{1}{4\pi c r^2}}{\dfrac{T}{13.8V} + \dfrac{1}{4\pi c r^2}}, \tag{6.19}$$

$$C_{80} = 10\log \frac{\dfrac{T}{13.8V}\left(e^{-A/S} - e^{-\left(\frac{13.8}{T}0.08+A/S\right)} \right) + \dfrac{1}{4\pi c r^2}}{\dfrac{T}{13.8V}e^{-\left(\frac{13.8}{T}0.08+A/S\right)}},$$

$$\approx 10\log\left(e^{\frac{13.8T}{V}}\left(1 + \frac{13.8V}{4\pi c r^2 T} \right) - 1 \right) \text{ for } A \ll S \text{ or } \bar{\alpha} \ll 1, \tag{6.20}$$

$$t_s = \frac{T}{13.8}\left(\frac{A}{S} + 1 \right)$$

$$\approx \frac{T}{13.8} \text{ for } A \ll S \text{ or } \bar{\alpha} \ll 1 \tag{6.21}$$

the latter with the influence of the direct field neglected.

7 Signal processing for auralization

The fundamentals and techniques of theoretical and engineering acoustics as introduced in the previous chapters, with all available analytic calculation models, allow predicting the generation and radiation of sound. Psychoacoustics offers performance models of the human auditory system and focused evaluation of the technical character of sound. Accordingly, numerous methods for analysis of acoustic signals are available. As described in the previous chapter, these can be based on signal theory, wave field physics or psychoacoustics. Sound can give information about 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 number extracted from acoustic signals can represent an average impression, at least approximately. But the subjective sound event as such is only covered by a full auditory experience. The perception, the impression, the interpretation and the meaning of sound is not covered by this technical approach. The full characterization and interpretation of sound, in the end, can be achieved only when hearing and other senses are involved directly. Therefore the technique of auralization offers an important extension 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.

7.1 The concept of auralization

> Auralization is the technique for creating 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
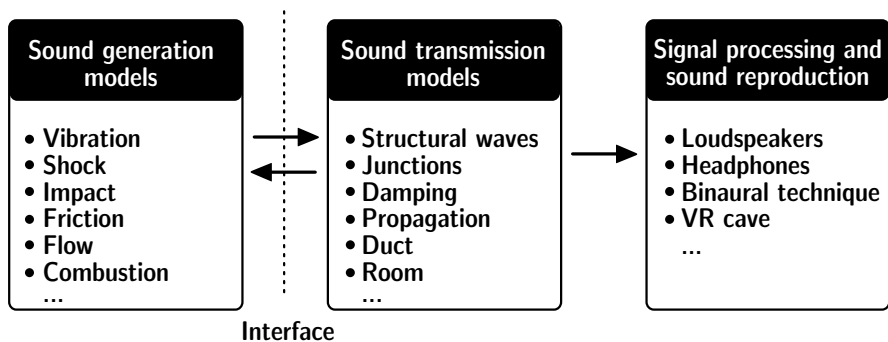
Fig. 7.1. Principle of auralization

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, however, is not a problem of physical feedback, but of psychological response. In a purely physical sense, the signal flow can be modelled only in the forward direction. In contrast, in problems of structure-borne sound, the situation changes completely. The vibrational velocity and displacement in beams and plates depends 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 can be represented typically by a two-port model, whose components can be determined by simulation or measurement. If the transfer functions of the elements are known by calculation or measurements, then the signal transmitted in the structure, duct, room or in a free field can be processed by convolution.

This looks simple at first glance, but the task of generating an appropriate filter becomes more difficult when more details are considered. 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 shall be high: 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, movement of the source or the receiver, just to list a few keywords.

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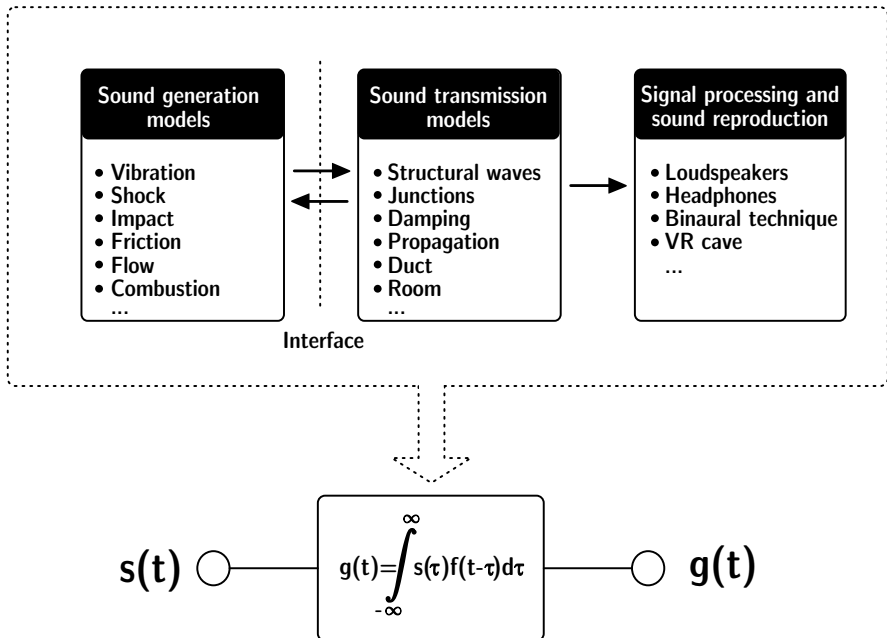
Fig. 7.2. Convolution of source signal *s*(*t*) with a filter impulse response *f*(*t*) to obtain a receiver signal *g*(*t*)

of freedom of vibration in structural paths, and the identification of interfaces between sound and vibration.

A historic example was mentioned in the preface. In the year 1929 in Munich, Spandöck and colleagues tried to model a room for speech and music performance. 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.

Today, with powerful computers available, the components of the auralization are typically 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 simulation 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 vibrational 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 perceivable and ready for sound reproduction (most simply over headphones). The steps necessary for proper auralization are performed by using tools of the field of 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.

7.2   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.

7.2.1   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, vibrational velocity or a similar signal.[19] We denote this function by $s(t)$ 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. A linear system affects signals in a linear way, which means that signal superposition can be treated as linear combination. Amplification just results in an amplitude change. For any transmission[20] (transformation, Tr) of a signal fed into a system, the following holds

$$Tr\left\{\sum_i (a_i \cdot s_i(t))\right\} = \sum_i \left(a_i \cdot Tr\{s_i(t)\}\right) = \sum_i (a_i \cdot g_i(t)), \qquad (7.1)$$

[19] output from any kind of sensor.

[20] A "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.

where s_i denotes the input signals and g_i the output signals, $i = 1,2,3\ldots$ 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,

$$Tr(s(t\text{-}t_0)) = g(t\text{-}t_0) \ . \tag{7.2}$$

By far, most systems in acoustics 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 (also here 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 reactions to signals in the time and frequency domains. This reaction is uniquely represented by the impulse response (in the time domain) or the stationary transfer function (in the frequency domain).

7.2.2   Impulse response and transfer function

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

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

$h(t)$ 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.
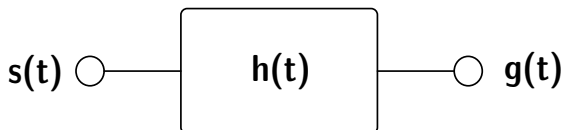
Fig. 7.3. Processing of source signal $s(t)$ with a filter impulse response $h(t)$ to obtain a receiver signal $g(t)$
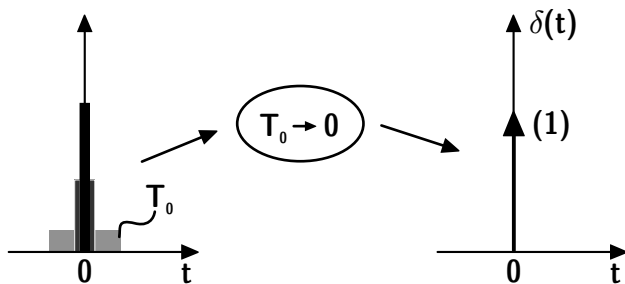
Fig. 7.4. Dirac pulse

It allows in particular the construction of filters. In some examples, it is a direct measure of 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, whose width tends to zero and height to infinity:

$$\lim_{T_0 \to 0} \frac{1}{T_0} \mathrm{rect}(\frac{t}{T_0}) . \tag{7.4}$$

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

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

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 by a factor (amplification):

$$a\delta(t) * s(t) = as(t) . \tag{7.6}$$

Time shift (propagation path, delay line):

$$\delta(t - t_0) * s(t) = s(t - t_0) . \tag{7.7}$$

Integration (step function $\varepsilon(t)$):

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

Excitation of a system with a Dirac pulse:

$$h(t) * \delta(t) = \int_{-\infty}^{\infty} h(\tau)\delta(t-\tau)\mathrm{d}\tau = h(t) \,. \tag{7.9}$$

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 ($\mathrm{Re}\{\underline{S}(f)\}$ and $\mathrm{Im}\{\underline{S}(f)\}$) or in an equivalent form as modulus and phase($|\underline{S}(f)|$ and $\varphi(f)$):

$$\underline{S}(f) = \mathrm{Re}\{\underline{S}(f)\} + \mathrm{Im}\{\underline{S}(f)\} = |\underline{S}(f)| \cdot \mathrm{e}^{\mathrm{j}\varphi(f)} \,. \tag{7.10}$$

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, $\underline{S}(f)$. Generally complex, the steady-state transfer function (related to harmonic signals) $\underline{H}(f)$ is defined as

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

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
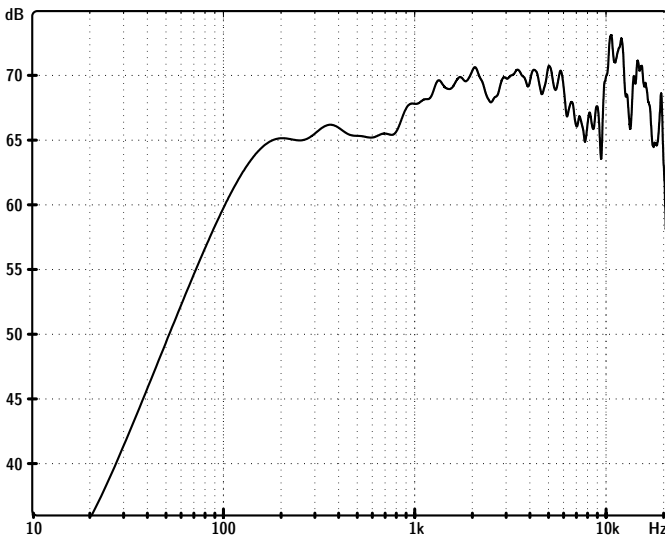
Fig. 7.5. Example of a loudspeaker sensitivity function
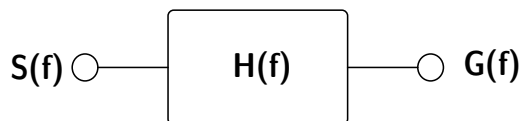
Fig. 7.6. Processing of source signal $\underline{S}(f)$ with a stationary transfer function $\underline{H}(f)$ to obtain a receiver signal $G(f)$

phase and calculation according to Eq. (7.11) yields the transfer function at this frequency. Repetition of this procedure in certain frequency steps gives a sample of the transfer function.

Accordingly the signal flow expressed in frequency domain reads

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

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

7.3   Fourier transformation

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

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

Thus, LTI system can be described in time or frequency domain uniquely. Signal flow through LTI systems, therefore, can 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
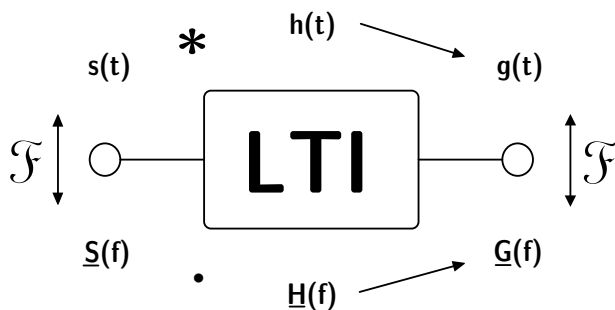
Fig. 7.7. Input and output signals of LTI systems

stage of signal transmission. Even in the temporal calculation process, the convolution integral, can be "transformed" into the frequency domain, thus giving a multiplication. This is not a surprise since the Fourier transformation is known in mathematics as the key to solving integrals of the convolution type.

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

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

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

In transforming signals, it reads

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

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

The Dirac pulse must be mentioned again. We completely understand its function as the identity function of convolution, since its spectrum is 1, the neutral element of multiplication. The latter equation can also be interpreted as a definition of the Dirac pulse.

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

$$\delta(t) = \int_{-\infty}^{\infty} e^{j2\pi f t} df . \qquad (7.19)$$

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, therefore, is consideration of discrete signal processing and the proper representation of continuous functions by sampling.

7.4   Analogue-to-digital conversion

To feed signals into a computer memory and to process them, 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 certain steps and sampled in time, thus yielding a discrete series of scaled binary data.

The precision of quantization depends on the amplitude resolution chosen. The range of numbers used is normalized and transformed to 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 ± full scale (assuming AC signals with an average close to zero). With a resolution of 16 bits, this is related to 65536 integers between –32768 and +32767, 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

$$N_{quant} = -20\log 2^n \approx -6n \,. \tag{7.20}$$

Sampling rates of 40–50 kHz are typical for sound in the hearing range are and quantization of 16 bits, in measurement or sound recording hardware, also up to 24 bits. Dynamic ranges caused by hardware limitations are thus available with same range as for 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 implemented as mathematical operations.

According to the theory of linear time invariant systems, the sampling process can be described as follows. An analogue signal[21] $s(t)$ is sampled at times nT with $n = 0,1,2,\ldots$ and $T = 1/f_{Sample}$ and instantaneous voltage is measured at each sample. This process corresponds to a multiplication of the analogue signal by a series of Dirac pulses,

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

---

[21] To be typically considered preconditioned in volt units at the input of the A/D device (sound card).
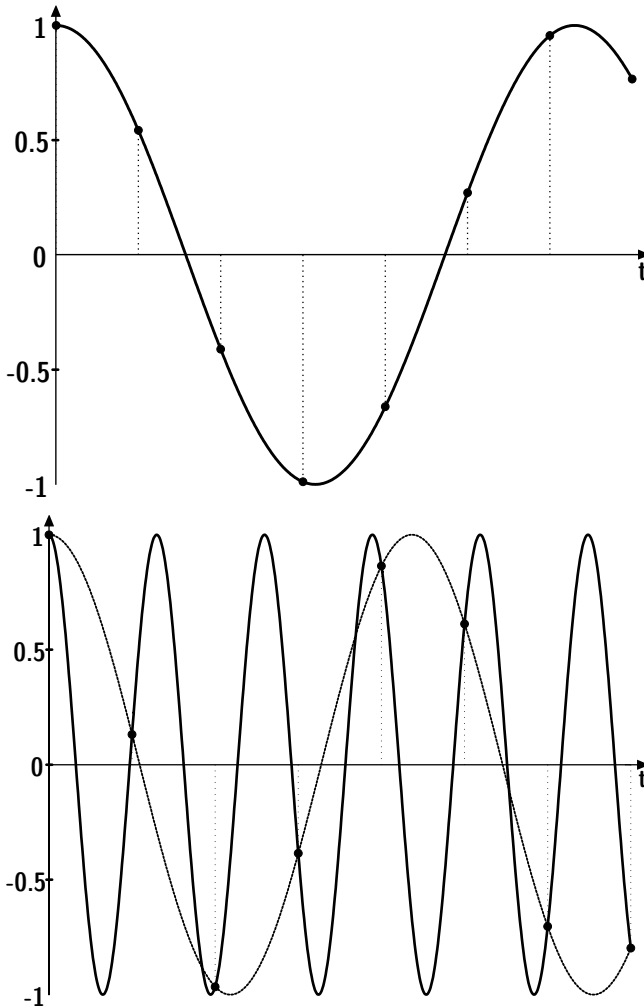
Fig. 7.8. Top: Sampling of a signal. Bottom: Ambiguity of the discrete samples matching to sinusoidals

and the sampled signal reads

$$s(n) \equiv s(t) \cdot \text{III}(t)\Big|_{t=nT} = \sum_{n=-\infty}^{\infty} s(nT)\delta(t-nT)\Bigg|_{t=nT} , \tag{7.22}$$

see Fig. 7.8.

At least two samples must cover one period of the harmonic signal, as illustrated in Fig. 7.8, to exclude any ambiguity, so-called "aliasing." Otherwise, harmonic signals with integral 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.

   More generally, we can 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

$$s(t/T) \cdot \mathrm{III}(t/T) \quad \circ\!\!-\!\!\bullet \quad \underline{S}(Tf) * \mathrm{III}(Tf) \tag{7.23}$$

with

$$\mathrm{III}(Tf) = \int_{-\infty}^{\infty} \mathrm{III}(t/T) \cdot \mathrm{e}^{-\mathrm{j}2\pi ft}\, \mathrm{d}t \tag{7.24}$$

denoting the Fourier transform of the Dirac series. The time and frequency axes are normalized to the sampling rate $1/T$. Note the inverse relationship between $\mathrm{III}(t/T)$ and $\mathrm{III}(Tf)$. A narrow Dirac sequence in the time domain corresponds to a wide series of spectral lines in the frequency domain.
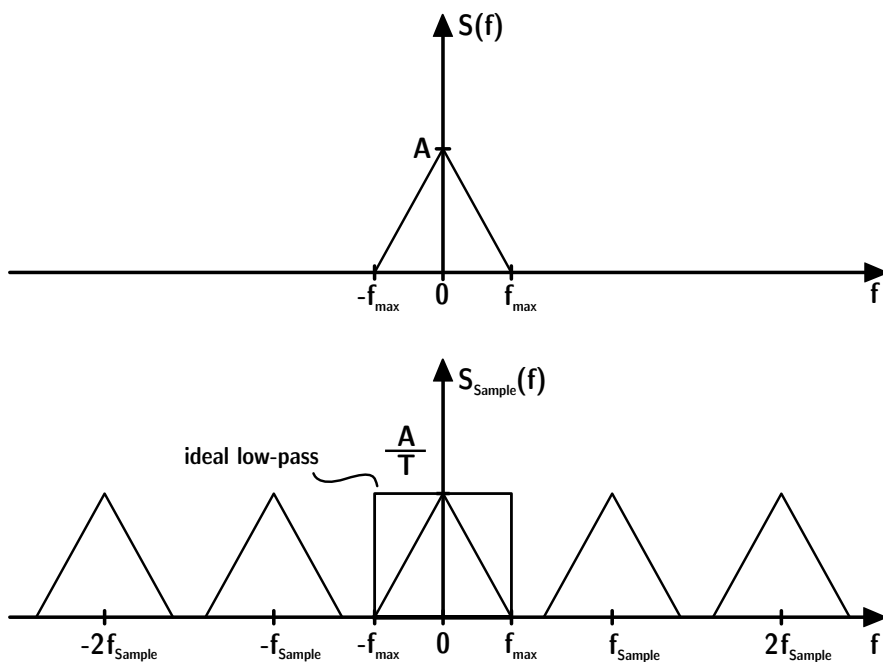
Fig. 7.9. Reconstruction of the analogue signal

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

$$f_{Sample} \geq 2 f_{max} . \tag{7.25}$$

7.5 Discrete Fourier transformation

For sampled signals, there remains the question regarding an efficient Fourier transformation. The calculation algorithm for the discrete Fourier transformation (DFT) is (compare Eq. (7.16))

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

The variable n represents the time domain, k the frequency domain. For solution of this sum, N^2 (complex) multiplications are required.
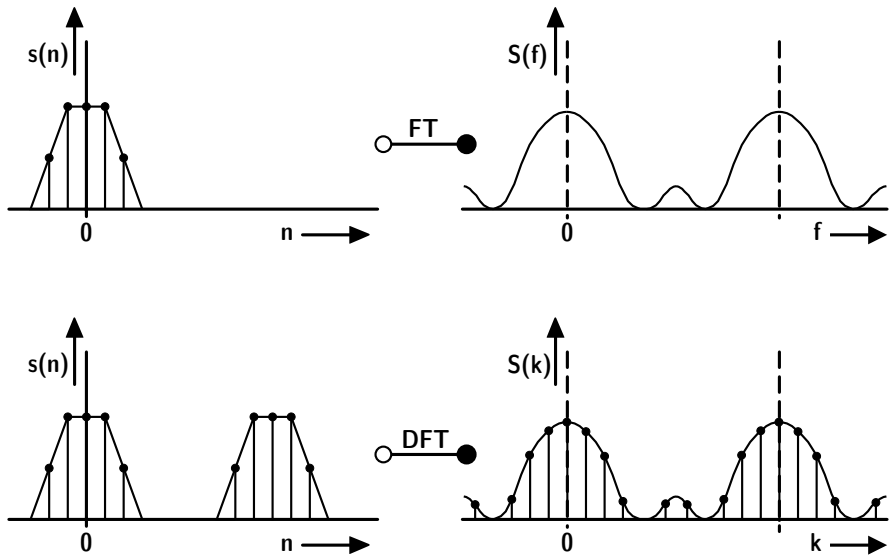
Fig. 7.10. Sampling and processing of a signal, $s(n)$ in top left. Top right: Corresponding theoretically continuous spectrum, $S(f)$. Bottom right: Numerical (discrete) spectrum $S(k)$. Bottom left: The periodic signal corresponding to the discrete spectrum $s(n)$ (after (Lüke 1999))

As a consequence of sampling, the spectrum of a sampled signal will be periodic (Eq. (7.23)) and continuous. But in digital representation, the spectrum can be stored only in digital form at certain frequency lines.[22] 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 having a sufficient distance between the temporal periods (see Fig. 7.9).

7.6 Fast Fourier transformation

The so-called fast Fourier transformation, (FFT) is a special version of the DFT. It is one of the key algorithms in virtual acoustics, in acoustic measurements, in speech and image processing and 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

$$N = 2^m \ (4, 8, 16, 32, 64, \ldots). \tag{7.27}$$

The reason for the accelerated calculation is preprocessing with the result of presorting symmetric terms and reducing 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 $N = 4$:

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

with

$$W = e^{-j2\pi/N} . \tag{7.29}$$

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

---

[22] We cannot store continuous data in the computer memory.

a so-called "bit reversal," a specific interchange of columns and rows, so that quadratic blocks of zeros (2×2, 4×4, 8×8, …) are created. Of course, all multiplication terms involving zeros can be omitted. In our example, the transformed matrix is

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

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.

Worth mentioning is 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:
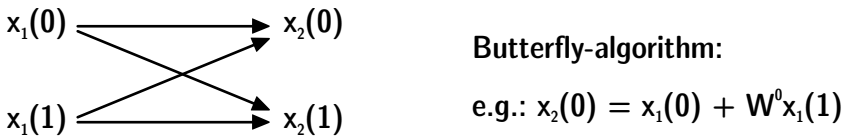
Fig. 7.11. FFT butterfly

The solution of an $m\times m$ matrix can finally be found by a cascade of m butterflies, which reduces the necessary number of multiplications from N^2 to $N\log_2(N/2)$, for example, for $N=4096$ by a factor of 372 from 16777216 down to 45056.

7.6.1   Sources of errors, leakage and time windows

At the given boundary conditions, several sources of errors are possible. At first, it must not be forgotten that FFT as a special form of DFT is related to periodic signals. If the signal to be transformed is 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.
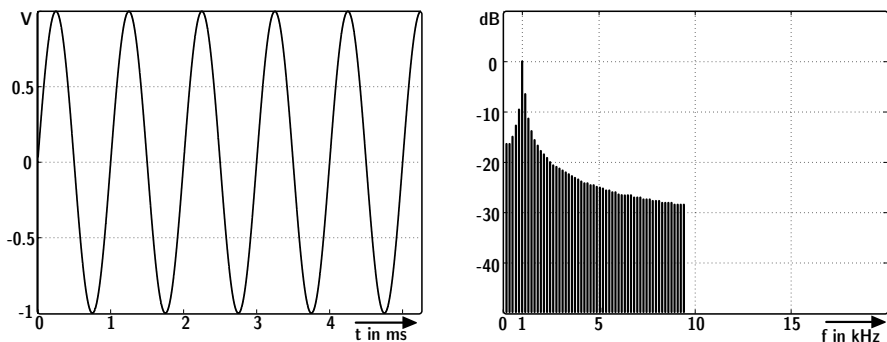
Fig. 7.12. A 1 kHz pure-tone signal and spectrum by using DFT (6.25 periods)

Another approximate method is the 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 the time domain. A symmetric window reduces early and late components in the signal and lets the midtime part pass unchanged. 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. Typical window functions (Examples, $n = 0,1,2,\ldots, N\!-\!1$)

Window	Function
Rectangular	$w(n) = 1$
Triangular	$w(n) = \begin{cases} \dfrac{n}{N/2} & ; \quad n = 0,1,\ldots,\dfrac{N}{2} \\ \dfrac{N-n}{N/2} & ; \quad n = \dfrac{N}{2},\ldots,N-1 \end{cases}$
Hanning	$w(n) = \sin^2\left(\dfrac{n}{N}\pi\right)$
Hamming	$w(n) = 0{,}54 - 0{,}46\cos\left(\dfrac{2\pi}{N}n\right)$
Blackman-Harris	$w(n) = a_0 - a_1\cos\left(\dfrac{2\pi}{N}n\right) + a_2\cos\left(\dfrac{2\pi}{N}2n\right) - a_3\cos\left(\dfrac{2\pi}{N}3n\right)$

7.7 Digital filters

Digital filters are used for pre- and postprocessing of signals. In measurements, they serve as high-pass, low-pass or band-pass filters. In auralization and sound reproduction they serve as a basis for filtering, convolution and for final adjustment of audio effects, including special cues such as 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 of 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, also for treatment of harmonic functions. By introducing the Laplace variable,

$$z = e^{j\omega}, \tag{7.31}$$

the frequency response is mapped to a complex two-dimensional function in the complex plane. The following rules can be applied to designing 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 by $e^{-j\omega t}$, thus affecting the phase without changing amplitudes. A pole (or zero) on the unit circle corresponding to a filter response $\underline{H}(j\omega)$ becomes infinite (or zero) at a certain frequency. A pole outside the unit circle creates an instability with increasing filter impulse response
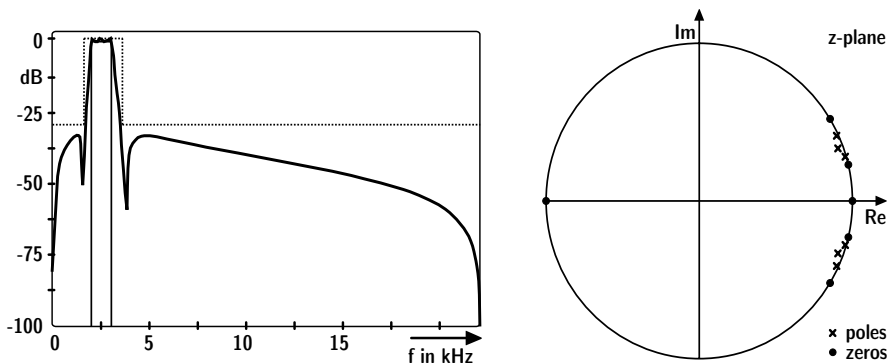
Fig. 7.13. Digital band-pass filter of the sixth order. Top: Frequency response. Bottom: Pole-zero diagram

$h(t)$. Poles outside the real axis generally correspond to oscillations of the filter impulse response.

Now, the frequency response can be constructed easily from the pole-zero plot. 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 to groups: IIR and FIR filters.

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

$$H(z) = \frac{\displaystyle\sum_{n=0}^{N} b(n)z^{-n}}{\displaystyle\sum_{n=0}^{N} a(n)z^{-n}}, \qquad (7.32)$$

which should approximate the desired response with the least possible order N. This can also be illustrated using a block diagram with forward and feedback lines. z^{-1} means a shift by one sample, and the triangles mean multiplication (amplification) by factors a or b.
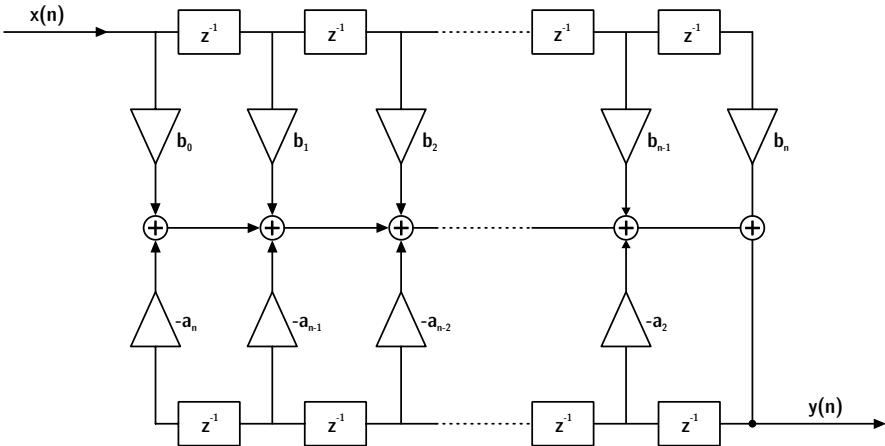
Fig. 7.14. Block diagram of an IIR filter. $x(n)$ and $y(n)$ are input and output signals, respectively

The output signal, $y(n)$, is created by 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. IIR filters also require less effort and complexity than FIR filters and usually have a lower order.

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

$$\sum_{n=0}^{N} a(n)z^{-n} = 1. \tag{7.33}$$

Thus the transfer function of the filter reads

$$H(z) = \sum_{n=0}^{N} b(n)z^{-n} , \tag{7.34}$$

with the following block diagram:
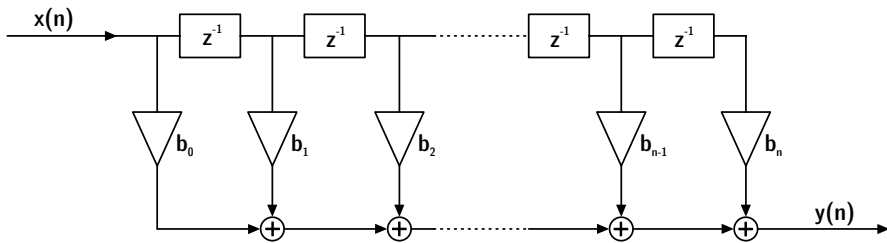
Fig. 7.15. Block diagram of an FIR filter

FIR filters are stable in each case. The output depends only on input data and not on feedback. The impulse response is identical to 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.

Filter concepts are useful and applicable for auralization of 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; Morjopoulos 1994; Kirkeby and Nelson 1999).

8 Characterization of sources

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.

Auralization

Sound generation models	Sound transmission models	Signal processing and sound reproduction
• **Vibration** • **Shock** • **Impact** • **Friction** • **Flow** • **Combustion** ...	• Structural waves • Junctions • Damping • Propagation • Duct • Room ...	• Loudspeakers • Headphones • Binaural technique • VR cave ...

Interface
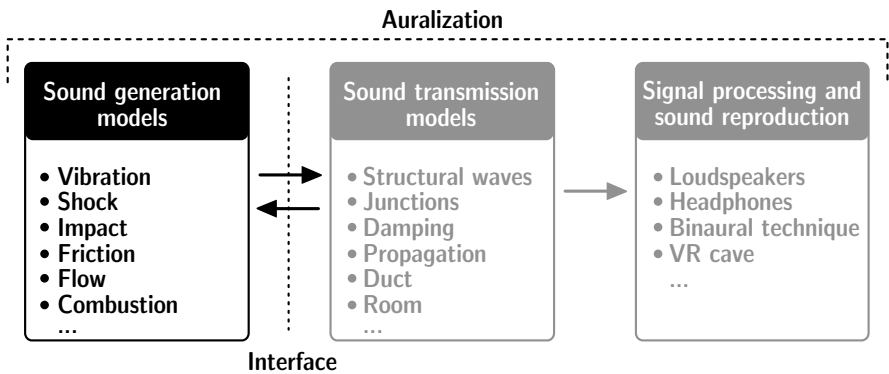
Fig. 8.1. Sound source characterization

In this section we consider how sound generation and sound sources can be recorded and processed.

8.1 Airborne sound sources

Input signals, $s(t)$, are related to specific sources. Typical examples are musical instruments, the human voice and a noise-generating machine. At this point, the question of feedback will be shortly discussed or rather the neglect 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, feedback might occur as a severe problem. The reason is the relation between the inner impedance of the source

and the radiation impedance (Sect. 2.3). As long as the radiation imped-
ance is small compared with the inner impedance of the source mecha-
nism, the radiated volume velocity is invariant to changes in the environ-
ment (room, duct, etc.).

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 data-
base. For musical instruments, the human voice or loudspeakers such data
can be found. For noise sources, however, such generally applicable direc-
tivity data are typically not available. The directivity must be modelled by
assuming specific elementary radiation conditions, or it must be measured
in an anechoic situation or in the near field. Microphone arrays on a meas-
urement surface such as arrays used in sound power measurements, are an
option for collecting data for source directivity encoding. In this respect, it
is also of interest to reconstruct, at least in approximation, the surface vi-
bration 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). Here
we focus on some specific examples. As an example of a detailed mathe-
matical model, multipole synthesis is introduced. Another option explained
in this book already is the functional basis of spherical harmonics.

8.1.1   Multipole synthesis

The prerequisite of multipole synthesis is that sound fields can be devel-
oped into a series of spherical harmonics (Sect. 2.5) or of multipoles
(Sect. 2.4). The aim of this procedure is coding of measured or calculated
directivity patterns. In simulations. the sets of multipole coefficients are
used to reconstruct the original source directivity. This is achieved by lin-
ear superposition.

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 har-
monics, the principle will yield identical results for sources of spherical
geometry. The decomposition of the sound radiation problem into spheri-
cal 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 these.

For radial symmetry, the spherical Hankel function of zero order is equivalent to the concept of a monopole. Now, higher orders of m include higher orders of sources, for instance, monopoles and dipoles in the three Cartesian orientations, x, y, z, are given by

$$
\underline{p}_{\text{o}}(k,\vec{r}) = \frac{\rho_0 c k^2 \hat{Q}}{4\pi} \psi_{00}^1(k,\vec{r}) \qquad \underline{p}_{8,\text{x}}(k,\vec{r}) = \frac{\rho_0 c k^3 d\hat{Q}}{4\pi} \psi_{11}^1(k,\vec{r})
$$
$$
\underline{p}_{8,\text{y}}(k,\vec{r}) = \frac{\rho_0 c k^3 d\hat{Q}}{4\pi} \psi_{01}^1(k,\vec{r}) \qquad \underline{p}_{8,\text{z}}(k,\vec{r}) = \frac{\rho_0 c k^3 d\hat{Q}}{4\pi} \psi_{10}^1(k,\vec{r})
\qquad (8.1)
$$

in direction $\vec{r} = (r,\vartheta,\varphi)^T$. Quadrupoles exist as lines or square arrays. The general expansion into spherical harmonics is

$$
\underline{p}_{\text{xy}}(k,\vec{r}) = \frac{\rho_0 c k^4 d^2 \hat{Q}}{24\pi} \psi_{22}^{-1}(k,\vec{r})
$$
$$
\underline{p}_{\text{xz}}(k,\vec{r}) = \frac{\rho_0 c k^4 d^2 \hat{Q}}{24\pi} \psi_{21}^{1}(k,\vec{r})
$$
$$
\underline{p}_{\text{yz}}(k,\vec{r}) = \frac{\rho_0 c k^4 d^2 \hat{Q}}{24\pi} \psi_{21}^{-1}(k,\vec{r})
$$
$$
\underline{p}_{\text{xx}}(k,\vec{r}) = \frac{\rho_0 c k^4 d^2 \hat{Q}}{24\pi} \left(\psi_{22}^{1}(k,\vec{r}) - 2\psi_{20}^{1}(k,\vec{r}) - 2\psi_{00}^{1}(k,\vec{r}) \right)
$$
$$
\underline{p}_{\text{yy}}(k,\vec{r}) = \frac{\rho_0 c k^4 d^2 \hat{Q}}{24\pi} \left(-\psi_{22}^{1}(k,\vec{r}) - 2\psi_{20}^{1}(k,\vec{r}) - 2\psi_{00}^{1}(k,\vec{r}) \right)
$$
$$
\underline{p}_{\text{zz}}(k,\vec{r}) = \frac{\rho_0 c k^4 d^2 \hat{Q}}{24\pi} \left(4\psi_{20}^{1}(k,\vec{r}) - 2\psi_{00}^{1}(k,\vec{r}) \right)
\qquad (8.2)
$$

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 for approximating and interpolating arbitrary source data. We obtain the velocity addressed to spherical wave functions by using the pressure gradient

$$
\vec{v} = \frac{j}{\rho_0 \omega} \nabla p ,
\qquad (8.3)
$$

which yields

$$
\nabla\psi_{mn}^{1} = \begin{pmatrix} k(2m+1)^{-1}\big[mh_{m-1}^{2}(kr)-(m+1)h_{m+1}^{2}(kr)\big]P_{m}^{n}(\cos\vartheta)\cos(n\varphi) \\ r^{-1}h_{m}^{2}(kr)\big[(m-n+1)P_{m+1}^{n}(\cos\vartheta)-(m+1)\cos\vartheta\,P_{m}^{n}(\cos\vartheta)\big]\cos(n\varphi) \\ -n(r\sin\vartheta)^{-1}h_{m}^{2}(kr)P_{m}^{n}(\cos\vartheta)\sin(n\varphi) \end{pmatrix} \quad 8.4)
$$

and

$$
\nabla\psi_{mn}^{-1} = \begin{pmatrix} k(2m+1)^{-1}\big[mh_{m-1}^{2}(kr)-(m+1)h_{m+1}^{2}(kr)\big]P_{m}^{n}(\cos\vartheta)\sin(n\varphi) \\ r^{-1}h_{m}^{2}(kr)\big[(m-n+1)P_{m+1}^{n}(\cos\vartheta)-(m+1)\cos\vartheta\,P_{m}^{n}(\cos\vartheta)\big]\sin(n\varphi) \\ -n(r\sin\vartheta)^{-1}h_{m}^{2}(kr)P_{m}^{n}(\cos\vartheta)\cos(n\varphi) \end{pmatrix} \quad (8.5)
$$

As mentioned above, straightforward expansion of the sound field in spherical wave functions can produce severe problems of poor convergence, if the source is nonspherical symmetry. Multipoles can be placed 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).
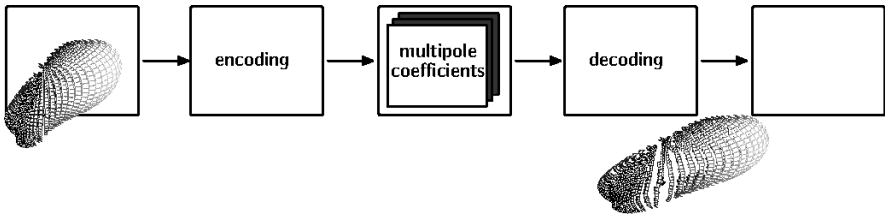
Fig. 8.2. Source directivity encoding

8.1.2 Musical instruments

Signals from musical instruments for auralization can be recorded in anechoic environment (anechoic chamber or, at least a highly absorbing room such as a recording studio) (Giron 1996). Recording must be done 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 for all signal components. This fact 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 instruments have a fluctuating radiation pattern
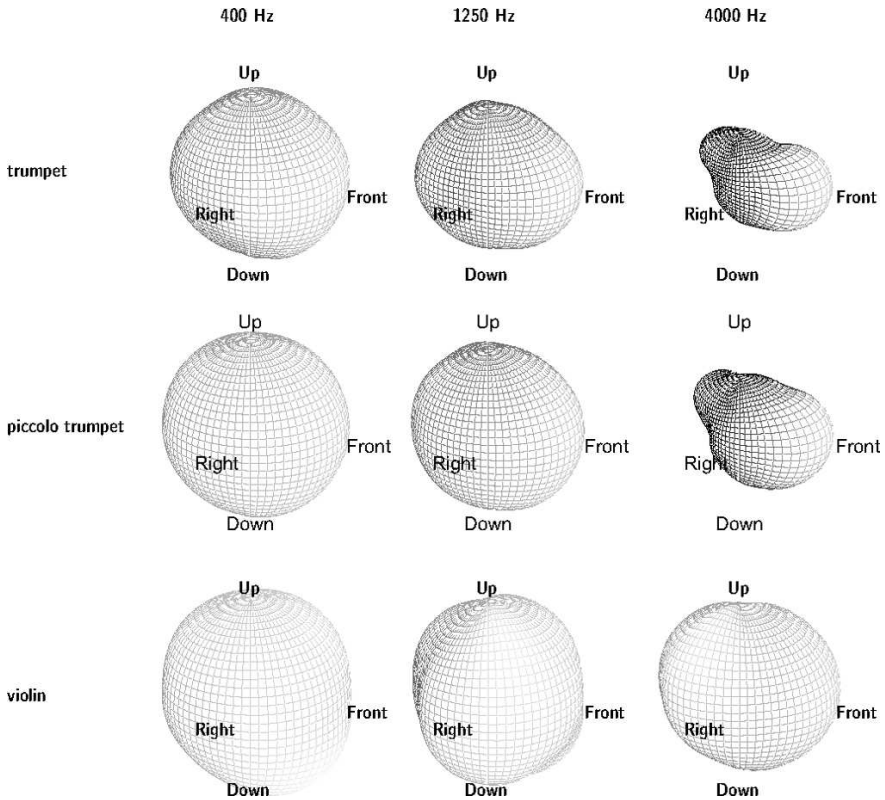
Fig. 8.3. Examples of directional characteristics ("balloons") of musical instruments

because the valves are opened and closed while playing the instrument.[23] 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 is one way to overcome this problem (Otondo and Rindel 2005).

Another aspect is the floor reflection. In hemianechoic rooms, the floor reflection will be included in the recording, such as in the actual performance on a stage. In this case, the floor reflection must not be included in the simulation as well.

The most comprehensive collection of directional characteristics of musical instruments was published by (Meyer 1995).

---

[23] Note that the valves form a line array of volume sources.

8.1.3 Singing voice

The radiation pattern of the human voice in talking or singing 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 an anechoic environment with a two-channel technique. One channel was recording the signal sung (glissando over at least one octave) in the far field at 2 m distance while the other channel was serving as reference near the mouth. With proper equalization and normalization to the frontal incidence the directional pattern is

45° Azimuth, 30° Elevation
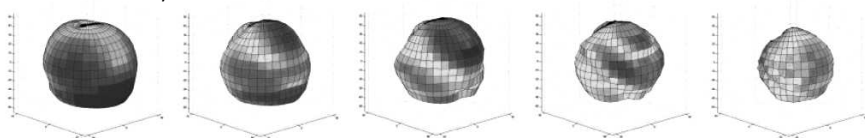
135° Azimuth, 30° Elevation
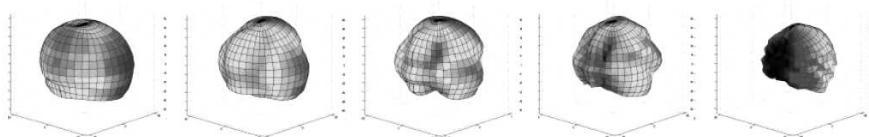
| 800Hz | 1250Hz | 1600Hz | 2500Hz | 8000Hz |

Fig. 8.4. Directivity of the singing voice at two view angles (after (Kob 2002))

Fig. 8.5. Artificial head for the singing voice (after (Behler and Müller 2000; Kob 2002))

obtained after placing the singer on a large turntable and repeating the procedure in angular steps.

8.1.4   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 an artificial head can be used. These data are found in telecommunication standards (ITU p. 58), and artificial "talking" heads are available.

8.1.5   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.

For some instruments, however, such as woodwind and string instruments, the directivity changes dynamically. This 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. In the reproduction situation, the channels must be treated separately and independently for the
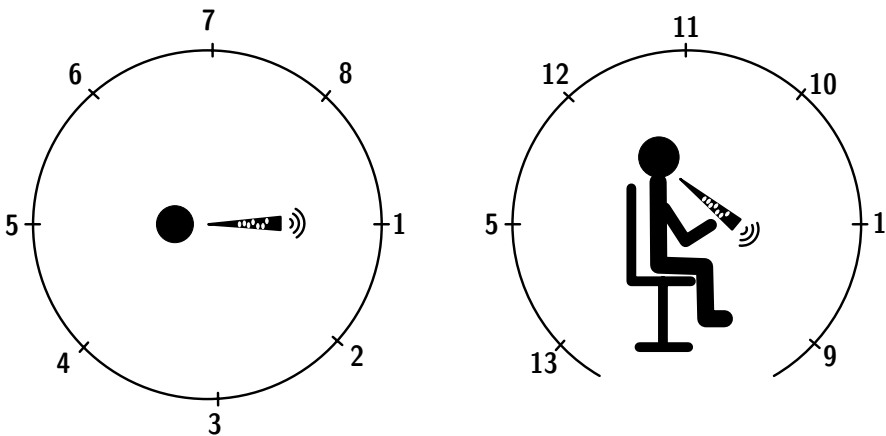
Fig. 8.6. Multichannel source recording (after (Otondo and Rindel 2005))

specific direction. This is easy for room simulation (Sect. 11.1) 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 to loudspeaker reproduction, the situation is more difficult. The channels now must be mapped to a kind of omnidirectional loudspeaker with an adjustable directional characteristic. If the characteristic is not constant but depends on the music played, such as 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.

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 recordings were made with orchestras and choirs in anechoic chambers or in near field conditions with the least possible cross talk. Other signals must be created in each case specifically. The "cleanest" 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).[24]

Later, about the mid-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.

---

[24] For 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 8.1. 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'Arlèsienne
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

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. 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

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. 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

Fig. 8.7. Multichannel source recording session (after (Vigeant et al. 2007), courtesy of J.H. Rindel, DTU Lyngby)

Finally, a study by (Vigeant et al. 2007) examined which recording technique is superior in realism and source width. Multichannel auralizations compared to single channel auralizations for both solo instruments and a full orchestra were compared in listering tests. The recordings made were based on the multichannel technique proposed by (Otondo 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 to channel separation between instruments and directions or radiation.

Table 8.4. Five-channel anechoic recordings of an orchestra of solo instruments using the technique developed in the DoReMi project (Otondo 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

8.2   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 represents the ideal case with contact conditions:

– point force,
– high contact impedance of the transmitting system.

8.2.1   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 a 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 its electrical equivalent. The force source, thus, is given by an ideal source with blocked force F_0, and a serial impedance, Z_i or mobility Y_i. Real force sources can be characterized in a first approximation with this simple model. Any source can be described by an open-circuit (blocked) 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 easily by using different (known) load impedances, Z_{a1} and Z_{a2}, to determine force and velocity.

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. 8.2.2.
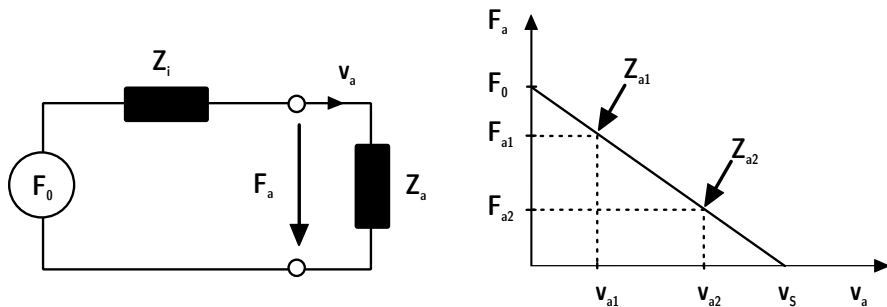
Fig. 8.8. Characterization of real force sources

$$Z_i = \frac{F_{a2} - F_{a1}}{v_{a2} - v_{a1}},$$ (8.6)

$$F_0 = F_{a1} - v_{a1} Z_i \text{ or } F_0 = F_{a2} - v_{a2} Z_i.$$ (8.7)

To discuss the interface problem further, we choose the model of power flow. The power of a source injected into the structure is

$$P = \frac{1}{2} |v_s|^2 \frac{\text{Re}\{Y\}}{|Y_i + Y|^2}.$$ (8.8)

Two extreme cases are easily defined. As mentioned above and valid for the tapping machine used on heavy construction (force source)

$$P \approx \frac{1}{2} |F_0|^2 \text{Re}\{Y\},$$ (8.9)

and for heavy sources working on lightweight construction,

$$P \approx \frac{1}{2} |v_s|^2 \text{Re}\left\{\frac{1}{Y^*}\right\}.$$ (8.10)

If the source is acting on a plate, its mobility is of interest. Plate mobilities depend strongly on the modal distribution. In a very large area, the point mobility of a plate with bending stiffness B and mass surface m'' can be approximated by

$$Y_{\text{plate}} = \frac{1}{8\sqrt{B' m''}}.$$ (8.11)

For finite-size plates, the modal pattern must be calculated first (Cremer and Heckl 1973) to obtain the point mobility (see also Sect. 5.2.1). One important aspect is that the modal pattern is not independent of the point of contact.

8.2.2 3-D 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 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. (8.10), but the force/mobility coupling is expressed in a matrix. Petersson and Plunt (1982) and Gibbs et al. (2007) extend this approach toward practical cases in building structures. The main extension is that a multiple contact is considered (see also (Petersson and Gibbs 2000)):

$$P = \frac{1}{2}\left((v_s)^T \left[(Y_i)^T \right]^{-1} (v_s^*) \right). \tag{8.12}$$

For multiple point contact, the relative force amplitude, F_j/F_i, and the coupling transfer mobility, Y_{ij}, between the contacts i and j yields the effective point mobility of the ith contact point (Gibbs et al. 2007):

$$Y_i^\Sigma = Y_i + \sum_j \frac{F_j}{F_i} Y_{ij}. \tag{8.13}$$

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,

$$\left| Y_i^\Sigma \right|^2 = \left| Y_i \right|^2 + \sum \left| Y_{ij} \right|^2. \tag{8.14}$$

9 Convolution and sound synthesis

If 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 the 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 for framewise convolution in slowly time-variant systems. Also, the technique of convolution or "filtering" (IIR, FIR) is valid for LTI systems exclusively. For 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.

sound source   audio DSP system   audio output
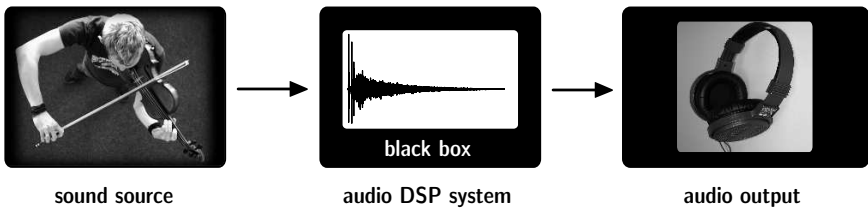
Fig. 9.1. Block signal processing for auralization

9.1 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 are involved (see also Chap. 15).

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 to avoid data full scale overflow. Although being faster in convolution, they might show disadvantages due to a 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:

$$s(n) * h(n) = \sum_{k=0}^{N-1} s(k)h(n-k) = g(n) \, . \tag{9.1}$$

MATLAB® code for discrete convolution

Explicit code:

```
function g = convolution(s,h)

%Initialize signal vectors
s = s(:); %input signal
h = h(:); %impulse response

%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).

9.2 FFT convolution

Here, we use the convolution integral in the frequency domain. The input data are stored in a vector as temporal sequence and the impulse response and the output signal are considered as 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);
```

9.2.1 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 data, we could not wait for the complete input signal being passed and stored in the output vector.[25] Furthermore, with adaptive filters, the latency would become unacceptably large. Therefore, we have to cut the signal into temporal segments, process them frame by frame, and transfer the results to the output unit sequentially. From frame to frame,

---

[25] This is obviously not possible in real-time applications, where the output signal must be replayed instantaneously.
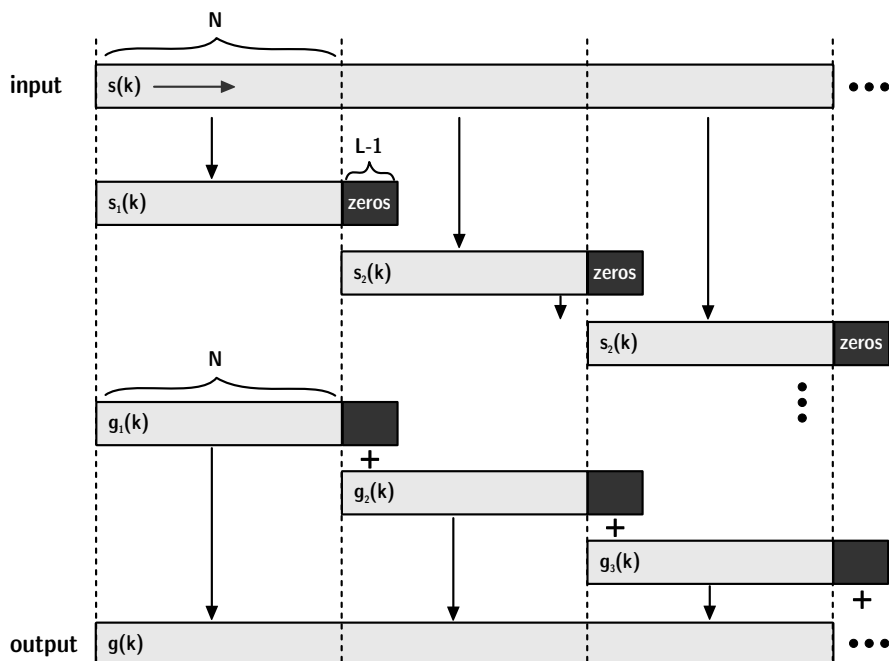
Fig. 9.2. Overlap-add convolution

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.

A segment of the input signal is now discussed using one segment between t_1 and $t_2 + t_{\text{fade}} = t_1 + \Delta t + t_{\text{fade}}$. It 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.
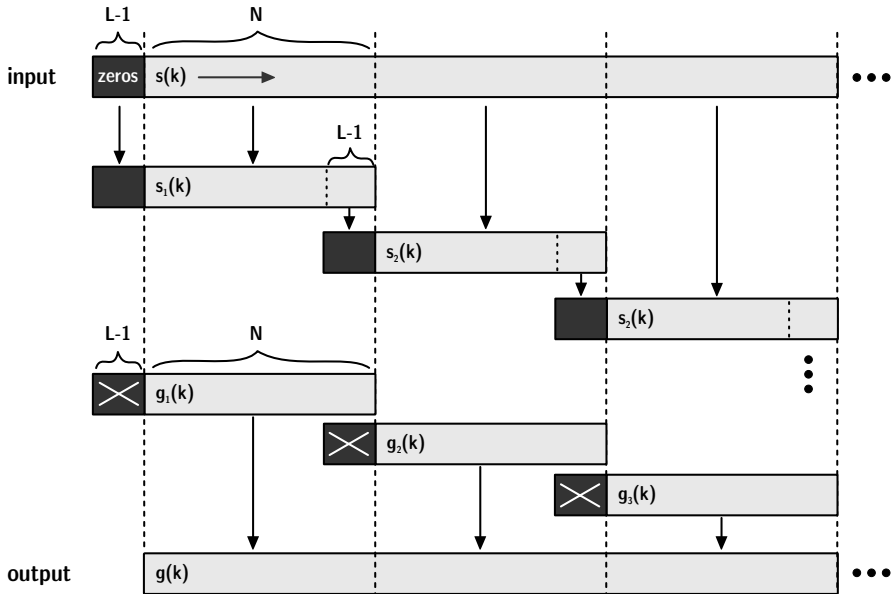
Fig. 9.3. Overlap-save convolution

9.3   Binaural synthesis

With the tools described in the previous chapter, it is a very small step toward the most important component of virtual reality systems: Filtering is used to connect a sound signal to a spatial cue. To achieve user immersion in VR systems, it is indispensable to create spatial sounds that 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 Sect. 8.1, can be processed so 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 Sect. 16.2.1) which can produce an appropriate spatial effect. A binaural mixing console can be used for processing headphone signals by using HRTF.
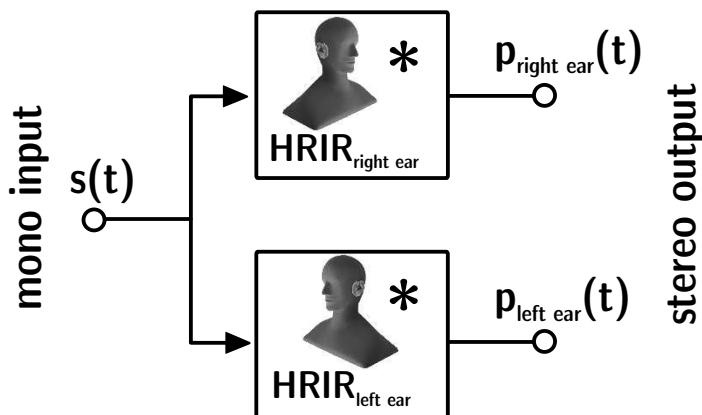
Fig. 9.4. Binaural synthesis

As mentioned in Sect. 6.3, 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 the HRTF,[26] any direction of sound incidence can be simulated, when a mono source $s(t)$ is convolved with a pair of head-related impulse responses:

$$
\begin{aligned}
p_{\text{right ear}}(t) &= s(t) * HRTF_{\text{right ear}} \\
p_{\text{left ear}}(t) &= s(t) * HRTF_{\text{left ear}}
\end{aligned}
\tag{9.2}
$$

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. IIR filters, for instance, were also created to simulate the poles and nodes of the 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 on 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 in Chaps. 14 and 15.

---

[26] Databases of the HRTF are available on the Internet, see Footnote 14 on page 90.

9.4   Binaural mixing console

Binaural mixing consoles are devices for creating multichannel binaural synthesis. This tool is ideal for auralization of free field environments where a small number of sources or reflections are to be modelled.[27] 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 suited. In a most elementary case, this is solved by sound recording at a certain distance and correcting for the propagation law (spherical wave, for instance) to account for the distance of the receiver. Binaural synthesis will take care of the direction of sound incidence. Of course, multiple sources can be simulated and auralized at the same time.

For spherical waves, the free field spectrum corresponding to one source reads:

$$\underline{H}\big|_{\text{left,right}} = \frac{\mathrm{e}^{-j\omega t}}{ct} \cdot \underline{H}_{\text{source}}(\theta, \phi) \cdot \underline{H}_{\text{air}} \cdot HRTF(\vartheta, \varphi)\big|_{\text{left,right}}, \tag{9.3}$$

where $\underline{p}$ denotes the sound pressure 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. 2.4) in source coordinates (θ, ϕ), $\underline{H}_{\text{air}}$ the low pass of air attenuation (Sect. 3.6) and HRTF the head-related transfer function (Sect. 6.3.1) of the sound incidence in listener coordinates at a specified orientation (ϑ, φ).

The total free-field spectrum is used as a convolution filter for processing source signals $s(t)$. In case of N sources,[28] the resulting signals are superposed. In the time domain, this procedure leads to

$$p(t)\big|_{\text{left,right}} = \sum_{i=1}^{N} s_i(t) * IFT(\underline{H}_i\big|_{\text{left,right}}), \tag{9.4}$$

where $\underline{H}_i$ denotes the spectra of the propagation function of source i.

Sound transmission or diffraction is modelled by adding attenuation filters (see Sects. 3.4 and 12.2) into Eq. (9.3), whose parameters 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.

---

[27] This will be no problem up to 32 or 64 channels, but it is not applicable for simulating binaural reflections in room acoustics.

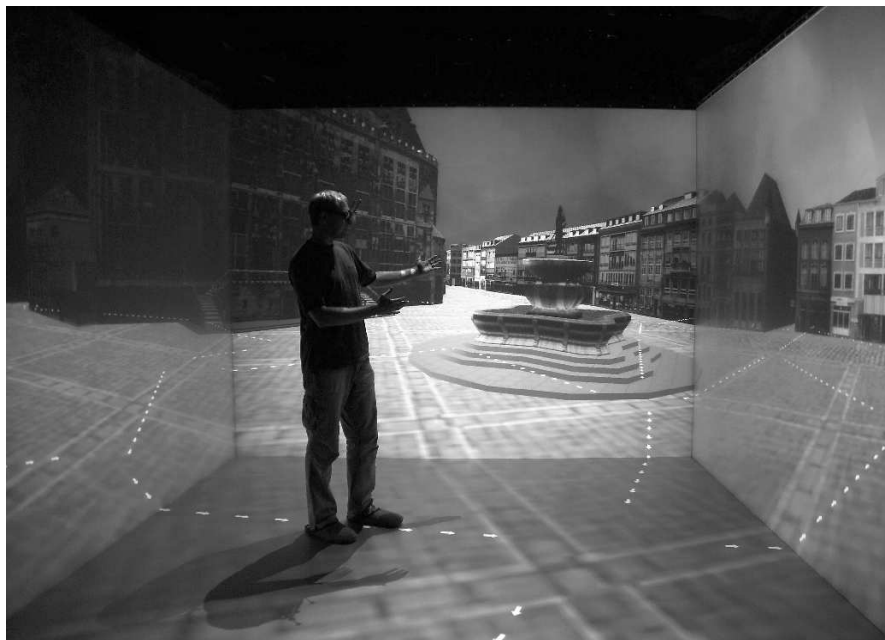[28] which may also include "image sources"; see Sect. 11.3.

Fig. 9.5. Free-field scene in a simulation of the market place of the City of Aachen (courtesy of the Virtual Reality group of RWTH Aachen University)
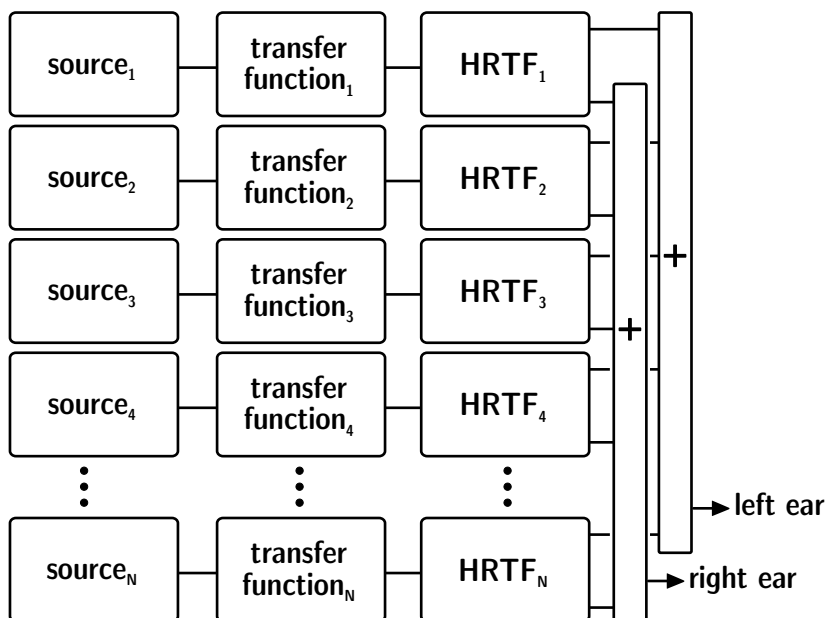

Fig. 9.6. Block diagram of free-field auralization

With the possibility of placing and moving 3-D 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 samples.[29] FIR or IIR filters (see Sect. 7.7) can be used, either on DSP platforms or by using PC environments with standardized audio interfaces. Investigations regarding the 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 also be performed in the frequency domain by using overlap-add or overlap-save FFT and multiplication by 64 complex data.

9.5 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 can discriminate between differences of 1° in azimuth in the frontal direction in the horizontal plane (see Sect. 6.3). In other directions, the angle intervals can be chosen larger. By using interpolation between the HRTF pairs available or by encoding the HRTF in a suitable functional basis (Torres et al. 2004), the size of the database can be reduced further (Minaar et al. 2005).

In applications of auralization with small distances between source and head, standard HRTF is not sufficient (Brungart et al. 1996). Standard HRTF is defined for plane wave incident 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 is the relative level evaluated. Spectral changes and, thus, changes of ITD and ILD are also caused by sources approaching the head at constant angles, but from distances closer than 1 m. This kind of data set must be determined with special care and unambiguous definition. A reasonable definition for a reference sound field is the spherical wave. Accordingly, the near-field HRTF is defined in the classical way, but it corresponds to a point source at a near distance.[30]

---

[29] The sampling rate for audio processing is usually 44.1 kHz.

[30] Proper point sources must be constructed for this kind of measurement. Alternatively, near-field HRTF can be calculated by using BEM.
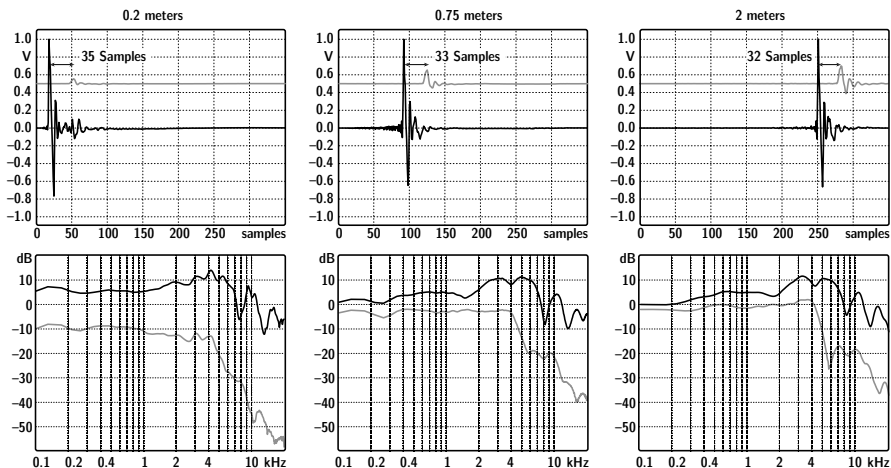
Fig. 9.7. Right to left: From standard to near-field HRTF (after (Lentz 2007))

One can easily calculate the memory space required for such a high spatial resolution of HRTF. Several approaches for data compression are possible. 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 where sound is used as carrier of information come into play. One tool in this respect is the so-called "auditory display."

10 Simulation models

After getting familiar with the principle of auralization, i. e., the separation of the acoustic problem in a model of signal transmission and binaural synthesis, and after having the sources characterized, 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. 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 future more and more simulation methods, however, will be applicable. The focus, therefore, is set to the physical background of the simulation methods and not on computational constraints. In Chaps. 11–15, we will discuss in detail up-to-date simulations methods applicable to virtual reality systems.
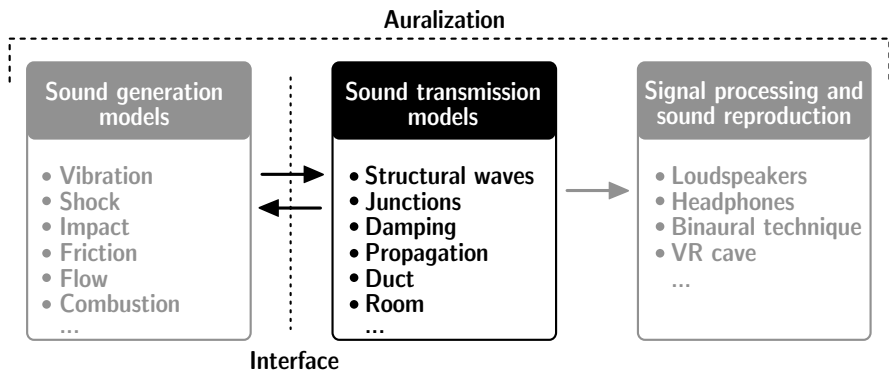
Fig. 10.1. Simulation models in acoustics

10.1 Simulation methods for sound and vibrational fields

Modelling of sound and vibrational propagation is one of the main problems in theoretical and numerical acoustics. All basic features of sound
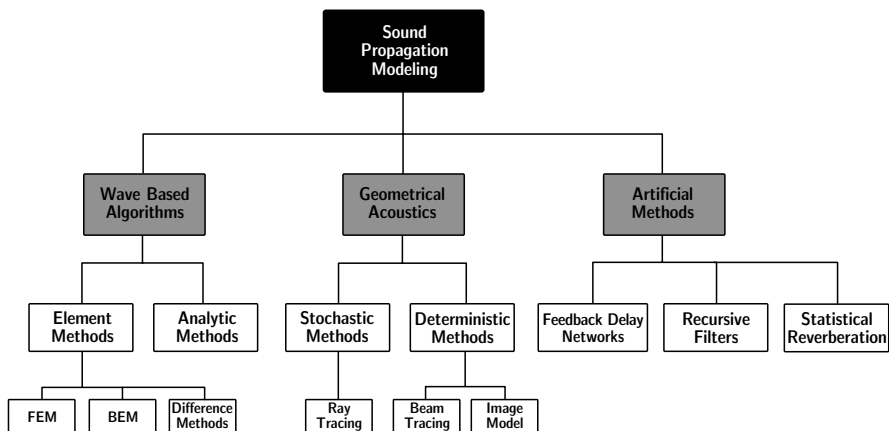
Fig. 10.2. Classification of models for simulating sound propagation

radiation (Chap. 2) and of sound fields (Chap. 3) come into play now. Nevertheless, specific methods particularly for nonanalytic approaches must be discussed in this chapter. Boundary conditions and field geometries mostly do not match the elementary conditions of standard coordinate systems such as Cartesian, spherical or cylindrical geometry. The basic solutions we found in Chaps. 2 and 3 are still interesting because they show the basic features of sound sources and propagation. The details and the fine structure in the results, however, can be obtained only 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 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 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 the density of samples in the time and frequency domains.

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 derivations in space and time (Skudrzyk 1971, Mechel et al. 2002):

$$p = \rho_0 \dot{g} \, ,$$
$$v = -\nabla g \, . \tag{10.1}$$

The benefit of Green's formulation is easily understood when it is applied to sources and spatial propagation paths from a point r_0 to a point r. For example, for a point source with volume flow Q (see Sect. 2.1),

$$Q = \lim_{a \to 0}\left(-4\pi a^2 \frac{\partial g}{\partial r} \right). \tag{10.2}$$

Its Green's function in free space is

$$g(r|r_0) = \frac{e^{-jk|\vec{r} - \vec{r}_0|}}{4\pi|\vec{r} - \vec{r}_0|}. \tag{10.3}$$

Coming back to the acoustic radiation problem and to the Helmholtz equation with source term ($f = jkZ_0q$; see Eq. (4.5)),

$$\Delta p(r) + k^2 p(r) = -f(r_0), \tag{10.4}$$

the Green's formulation leads to the Helmholtz–Huygens integral:[31]

$$p(r) = \iiint f(r_0)g(r|r_0)\mathrm{d}V_0$$
$$+ \iint \left(g(r|r_0)\frac{\partial p(r_0)}{\partial n} - p(r_0)\frac{\partial g(r|r_0)}{\partial n} \right)\mathrm{d}S_0 \tag{10.5}$$

With this integral, the resulting sound pressure of various kinds of source distributions in a volume and 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 and dipoles.

The discussion of Green's functions was, so far, related to harmonic signals. If we expand the radiation in the time domain by assuming an impulse excitation,[32] the tremendous importance of Green's functions becomes clear. They are *filters* transporting signals to the receiving point, expressed in the convolution of source functions with Green's functions (note the temporal relationship between source and receiver point):

$$p(r,t) = \int_{-\infty}^{t} f(r_0,t_0)g(r|r_0,t - t_0)\mathrm{d}t_0. \tag{10.6}$$

---

[31] The derivations in the surface integral defined in the direction normal to the surface elements (Mechel et al. 2002).

[32] With impulse excitation, we consider a constant harmonic spectrum with zero phase, see Sect. 7.2.2.

By Fourier transformation, we obtain adequate solutions in the frequency domain. The kernel of the convolution can then be expressed in terms of source volume velocity, Q, which leads to the formulation of a transfer impedance,

$$Z = \frac{p(\omega, r)}{Q(\omega, r_0)}.$$
(10.7)

Also, the concept of a transfer function between the sound pressure at one point to the sound pressure of another point can be chosen, if the source characterization is based on a near-field pressure signal.

$$H = \frac{p(\omega, r)}{p(\omega, r_0)}.$$
(10.8)

Similarly, transfer functions and Green's functions for structural acoustics can be defined.

Which approach, Green's function, transfer impedance or transfer function, is preferable depends on the kind of source. Force or pressure sources are more straightforward to be coupled to Green's functions, velocity sources to transfer functions. Most easy to remember is the fact that multiplication of the source signal by the transmission function should yield a sound pressure signal in the end.

10.1.1 Reciprocity

Green's functions are reciprocal (Lyamshev 1959). Reciprocity is one of the most powerful tools in determining acoustic transfer functions. The problem of sound and vibrational transmission in a passive linear time-invariant system can be solved in both ways. 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 vibroacoustic transmission from a force source exciting a structure that radiates to a receiver point (Fahy 1995).

Airborne sound reciprocity

The direct formulation consists of a real volume source emitting sound that is received in the field space. The ratio between the sound pressure at the receiver to the volume flow of the source[33] is $p_{receiver}/Q_{source}$. In the reciprocal

---

[33] = acoustic transfer impedance.
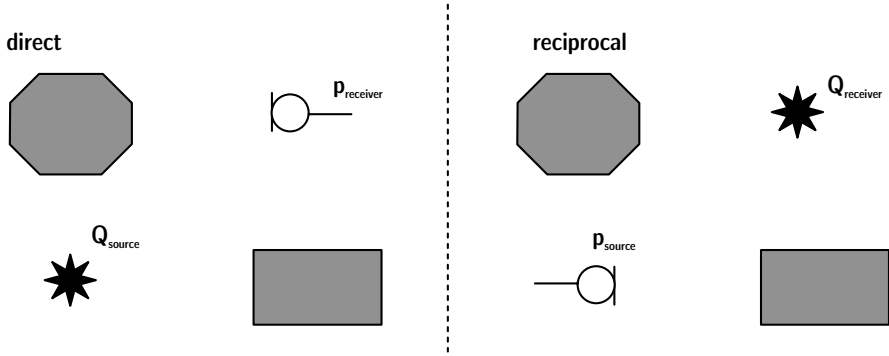
Fig. 10.3. Reciprocity of airborne sound propagation from a monopole source (after (Fahy 1995))

arrangement, this ratio is identical to the ratio of the pressure at the source point to the volume flow at the receiver, $p_{\text{source}}/Q_{\text{receiver}}$. However, specific reference conditions must be defined. The reference conditions concern the determination of the sound pressure which must be obtained in a mechanically blocked state ($Q=0$).[34] Thus,

$$\left.\frac{p_{\text{receiver}}}{Q_{\text{source}}}\right|_{Q_{\text{receiver}}=0} = \left.\frac{p_{\text{source}}}{Q_{\text{receiver}}}\right|_{Q_{\text{source}}=0}. \tag{10.9}$$

Vibroacoustic transfer function reciprocity

In a problem of a force source exciting a structure that radiates sound, the reciprocity relationship is given by

$$\left.\frac{p_{\text{receiver}}}{F_{\text{source}}}\right|_{Q_{\text{receiver}}=0} = \left.\frac{v_{\text{source}}}{Q_{\text{receiver}}}\right|_{F_{\text{source}}=0}. \tag{10.10}$$

We consider the sound pressure received from the direct formulation depending on the force injected on the sound-radiating structure. This ratio is identical to the ratio in the reciprocal approach between the volume flow at the receiver point in the fluid to the free ($F=0$) normal velocity on the structure.

---

[34] equivalent to an open-circuit situation for obtaining the voltage in terminated electric circuits.
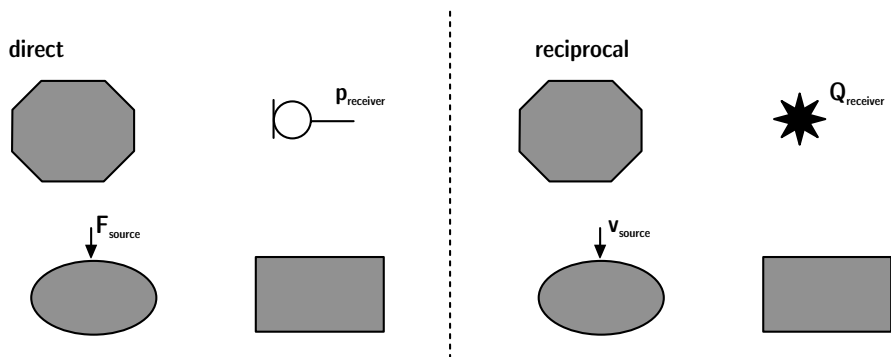
Fig. 10.4. Vibroacoustic transfer function reciprocity (after (Fahy 1995))

Vibroacoustic Green's function reciprocity

When the velocity of a vibrating surface element not the force is the input at the source side, the reciprocity relationship is given by

$$\left.\frac{p_{\text{receiver}}}{v_{\text{source}}\text{d}S}\right|_{Q_{\text{receiver}}=0} = \left.\frac{p_{\text{source}}}{Q_{\text{receiver}}}\right|_{v_{\text{source}}=0}. \tag{10.11}$$

Here, we consider the sound pressure received from the direct formulation depending on the local normal velocity of a surface element $\text{d}S$ of the sound-radiating structure. This ratio is identical to the ratio in the reciprocal approach between the sound pressure at the receiver point in the fluid to the blocked ($v = 0$) sound pressure on the structure. This relationship is valid for all surface elements of the structure.
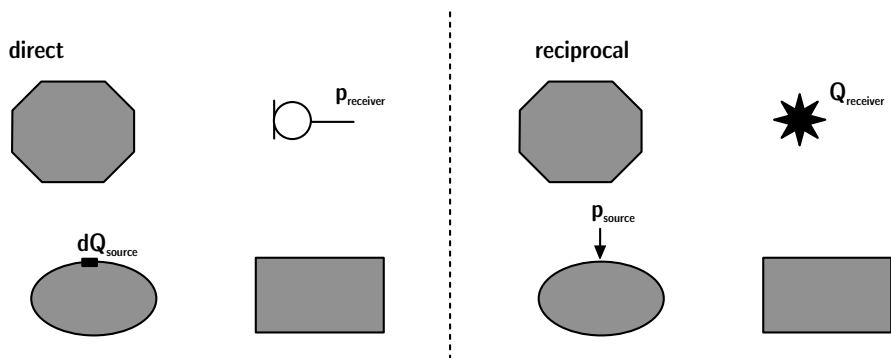
Fig. 10.5. Vibroacoustic Green's function reciprocity (after (Fahy 1995) with $\text{d}Q_{\text{source}} = v_{\text{source}}\text{d}S$)

After having discussed the principles of sound propagation between points in a fluid and between structures and fluid in analytic examples, we will focus in the next section on numerical methods in the frequency and time domains. These methods are applicable in a more general sense to any kind of geometric conditions of source, receiver and environment.

10.1.2 Frequency domain models

In frequency domain calculations, a constant frequency is considered. Hence, the discussion of transfer functions is based on harmonic signals.

$$s(t) = \underline{A} \cdot e^{j\omega t} , \tag{10.12}$$

$s(t)$ denoting any complex-amplitude harmonic signal of vibration, source volume velocity or sound pressure in the field point. In this case the wave equation reduces to the homogeneous Helmholtz equation (compare Eq. (4.5)), for instance, for the sound pressure in the field volume, $x \in V$.

$$\Delta p(x) + k^2 p = 0 . \tag{10.13}$$

To obtain the free (modal) response, we consider a source-free field domain, where the right-hand side of the equation is zero. Instead, this equation enables us to calculate the sound pressure relationships between field points, and this is the perfect approach for calculating transfer functions (Eq. (10.8)).

Depending on the geometry, as mentioned above, generally we cannot solve the Helmholtz equation straightforwardly, 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 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) use of the Helmholtz equation in an 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, the latter in the formulation of the finite element method.

Meshing

Numerical models require spatial discretization by introducing a "mesh." 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,
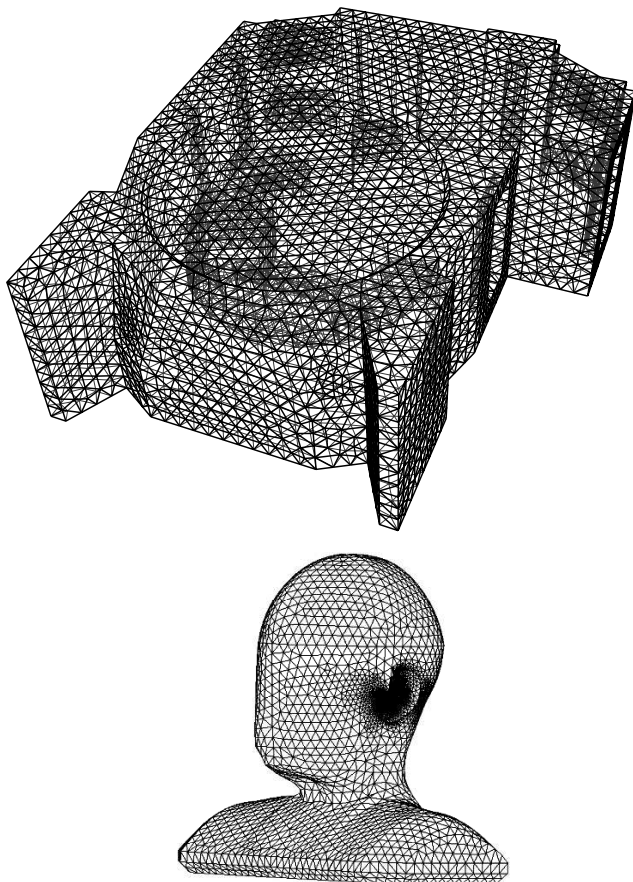
Fig. 10.6. Examples of meshes for numerical wave propagation analysis. A recording studio (top) and a customized dummy head (bottom)

triangular, 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 vibrational 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 are 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.

Boundary Element Method

The boundary element method (BEM) is explicitly related to 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 rearranged into the Helmholtz–Kirchhoff integral equation and discretized. The Helmholtz–Kirchhoff integral is the source-free Helmholtz–Huygens integral (Eq. (10.5)):

$$p(r) = \iint \left(g(r|r_0) \frac{\partial p(r_0)}{\partial n} - p(r_0) \frac{\partial g(r|r_0)}{\partial n} \right) \mathrm{d}S_0 , \qquad (10.14)$$

with Green's functions in 3-D free space:

$$g(r|r_0) = \frac{e^{-jk|\vec{r}-\vec{r}_0|}}{4\pi|\vec{r}-\vec{r}_0|} , \qquad (10.15)$$

which fulfil the far-field radiation (Sommerfeld) condition of vanishing sound pressure for $r \to \infty$.

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 vibrational velocity as a driving source. This integral is formulated in discretized form on a surface mesh and solved numerically in matrix algebra. The crucial point of the BEM formulation is the numerical nonuniqueness. It is worth mentioning 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 detail. 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 are even in free software.[35]

The complexity of BEM can be roughly summarized as follows: A simulation which must be calculated up to a frequency of f required a mesh element size of at most $c/6f$. The resulting model size of a surface S is

$$N = \frac{36Sf^2}{c^2} . \qquad (10.16)$$

---

[35] The reference software for acoustic BEM is "Sysnoise"™
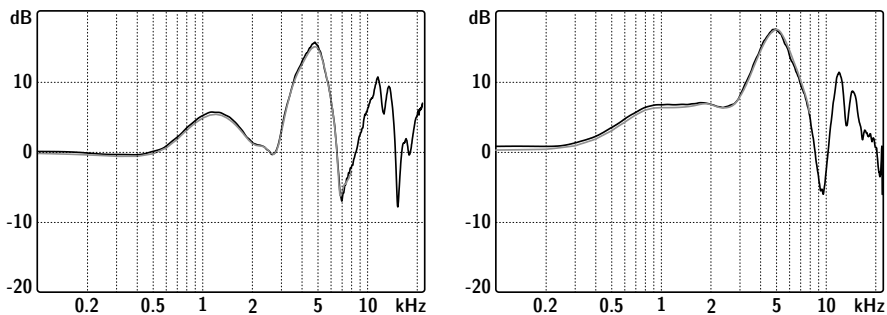http://www.lmsintl.com/SYSNOISE (renamed "Virtual.Lab Acoustics")

Fig. 10.7. BEM-calculated HRTF of a customized dummy head, left: Frontal inci-dence; right: 45° in the horizontal plane for the ipsilateral ear. Head model corre-sponding to Fig. 10.6 (bottom). Example follow the method described in (Fels et al. 2004), see also Sect. 6.3.1

The BEM matrix then contains N^2 entries. Solvers of PC software[36] to-day are capable of inverting a matrix of 8000 nodes in 60 seconds. This re-sult holds for one frequency. In a problem of required frequency resolution of 1 Hz, the calculation time multiplies by f, which yields some $f/60$ hours for numerically generating a complete set of transfer functions for all field points.

Advanced techniques such as fast multipole BEM (Sakuma and Yasuda 2002; Yasuda and Sakuma 2003; Marburg and Schneider 2003) are devel-opments that allow separating meshes into regions of high discretization and others with the effect of transfer propagation. The complex linking be-tween mesh elements is thus rearranged in a hierarchical way.

Finite Element Method

Finite elements are created by 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 equilibrium[37] leads back to a stable and minimum energy state. Due to its general energetic formu-lation, this principle is used for mechanical problems of static load and de-formation (also for crash test simulation), for fluid dynamics, heat conduc-tion, electromagnetic or acoustic field problems, and it is also the basis for the finite element method (FEM) (Zienkiewicz 1977).

---

[36] on a machine with a 2 GHz dual-core processor and sufficiently large RAM to keep the matrix problem in-core.

[37] by applying virtual displacement, for example.

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.

All elements' entries are combined into a so-called "stiffness" matrix, S,[38] 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.
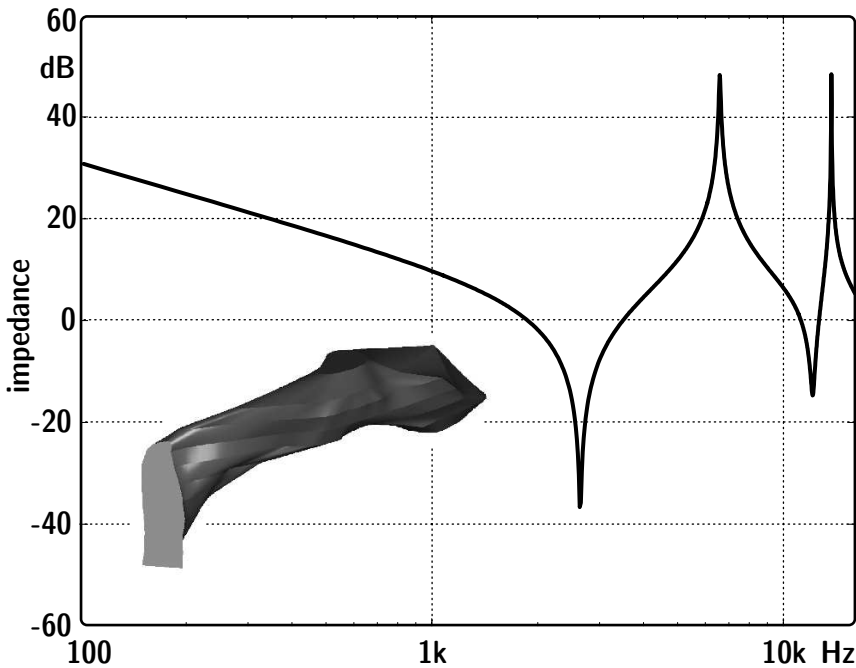
Fig. 10.8. Example of an ear canal impedance $20\log(|\underline{Z}|/Z_0)$ (for the CAD model shown, ear canal entrance and reference plane on the left) calculated using FEM

---

[38] Although other forces may be used here, the historic name is related to problems of static deformation by Hooke's forces.
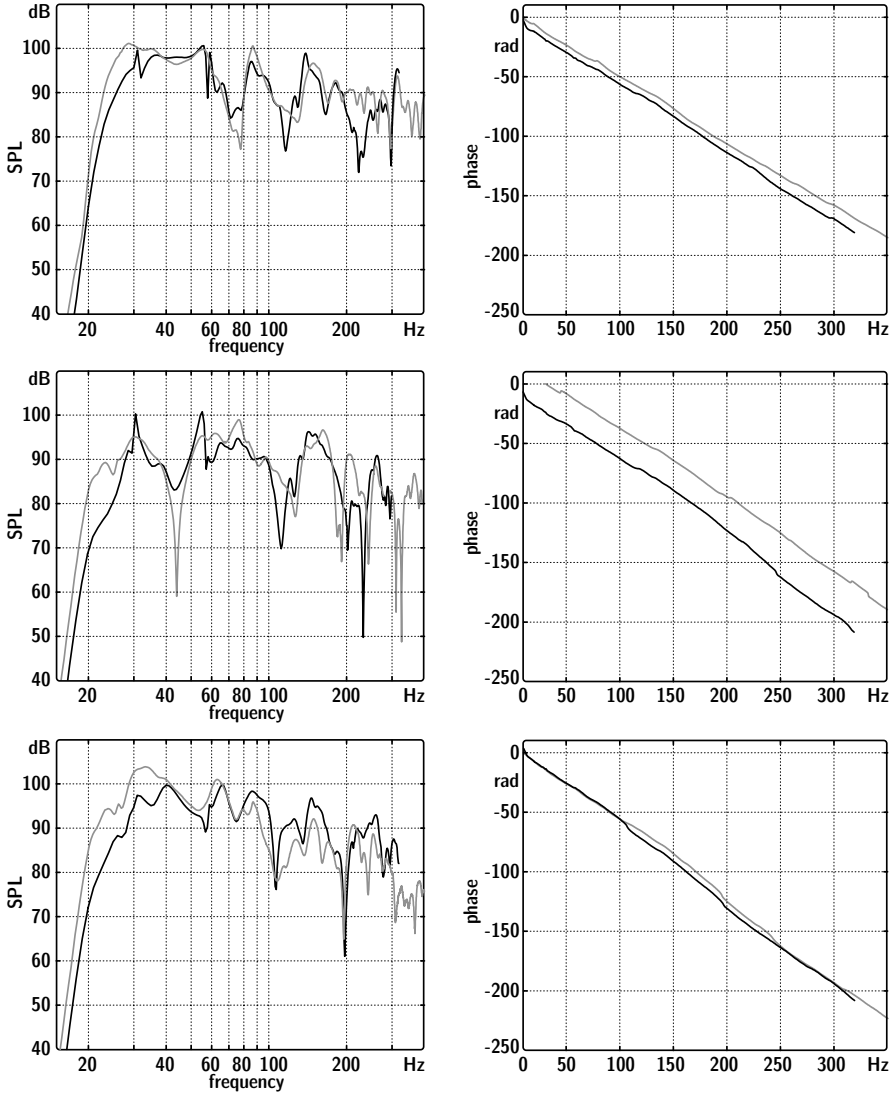
Fig. 10.9. Modulus and phase of results from FEM calculations and measurements in a recording studio (corresponding to the mesh shown in Fig. 10.6 (top), example after (Aretz 2007))

In FEM solvers, the direct solution to determine the eigenvalues of the matrix can be used, just using the matrix equation without further subspace conditions. In the indirect method, the problem is projected on a modal basis into 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.

State of the art is FE used for sound pressure calculation in small or midsize rooms for frequencies up to some kHz. Using PC software, typical mesh sizes are in a range of 100,000 nodes, and typical calculation times are of the order of magnitude of 5 minutes per frequency.

Modal approach and modal superposition

As in the finite element formulation, modes may serve as set of orthogonal function 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. 4.1) with midband frequency, half-width or quality. This way of modal analysis is well known as powerful tool in measurement of complex systems. Also here, 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 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. 4.2.1, but it is generally interesting for all problems of acoustics and vibration.

Frequency spacing

The resolution of numerically determined spectra should 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

$$\Delta_{f,mode} = \frac{\delta}{\pi} \qquad (10.17)$$

shall be covered by at least two frequency lines. $\Delta_{f,mode}$ denotes the half-width of a typical resonance. This kind of line spectrum can approximate well a second-order band-pass function with a damping of δ. The decay time of this mode is approximately

$$T = \frac{6.9}{\delta} . \qquad (10.18)$$

The required frequency spacing to model the system is thus

$$\Delta f = \frac{2\delta}{\pi} = \frac{4.4}{T} ,$$ (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 for understanding the basic principles. On this basis, any complex system of coupled resonances, in problems of airborne sound or structural vibration, or both coupled can be described by using energy balance and energy flow. At this stage of abstraction, SEA offers a powerful technique for calculating sound pressure levels and so 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 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.

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 for each medium, material or 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.
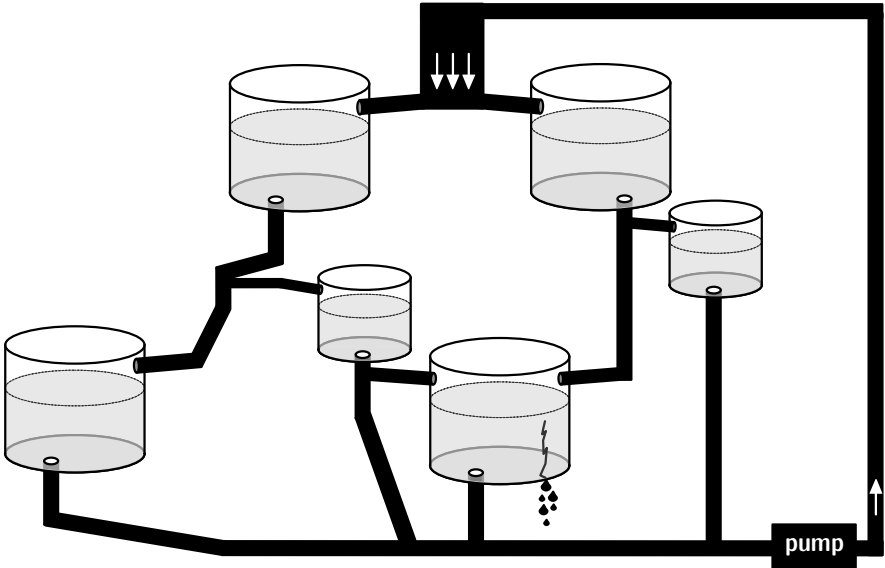
Fig. 10.10. Water flow in a system of basins and tubes

The power extracted from a subsystem i is given by

$$\Pi_i = \omega\eta_i E_i,$$
(10.20)

where η_i is the damping loss factor and E_i the steady-state energy stored in modes of spectral density v_i. When two subsystems are coupled, conservation of energy and balance leads to equilibrium when

$$\Pi_{ik} = \omega v_i \eta_{ik}\left(\frac{E_i}{v_i} - \frac{E_k}{v_k}\right).$$
(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 statistical sense since the probability of energy exchange in modal coupling depends on several factors including interaction of normal and tangential modes, fluid-structure coupling and geometric factors. It is essential that subsystems are independent, weakly coupled and contain statistically many independent modes.

Furthermore, reciprocity can be found in SEA, and this is a powerful tool to describe coupling loss factors.

$$v_i \eta_{ik} = v_k \eta_{ki}$$
(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.

$$\Pi_{i,\text{input}} = \omega\eta_i E_i + \sum_k \omega v_k \eta_{ik}\left(\frac{E_i}{v_i} - \frac{E_k}{v_k}\right) \qquad (10.23)$$

The equation of modal energies of all subsystems, coupling and power flow can thus be expressed in matrix form, such as

$$\mathbf{\Pi} = \mathbf{CE}, \qquad (10.24)$$

which for three subsystems is expanded in the form

$$\mathbf{C} = \begin{pmatrix} \omega v_1(\eta_1 + \sum_{k\neq 1}\eta_{1k}) & -\omega v_1\eta_{12} & -\omega v_1\eta_{13} \\ -\omega v_2\eta_{21} & \omega v_2(\eta_2 + \sum_{k\neq 2}\eta_{2k}) & -\omega v_2\eta_{23} \\ -\omega v_3\eta_{31} & -\omega v_3\eta_{32} & \omega v_3(\eta_3 + \sum_{k\neq 3}\eta_{3k}) \end{pmatrix} \qquad (10.25)$$

Typical applications in acoustic engineering involve hundreds or thousands of subsystems. Free software is available on the Internet.[39]

When the energy density or the sound intensity is known, the sound pressure can be estimated on the basis of a specific sound field type. In diffuse field condition in a room, for example, the squared sound pressure is related to the energy density by using Eq. (1.31).

10.1.3 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 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 wave fronts propagating in time. Reflections, diffraction and other propagation effects can be calculated. At chosen field points, impulse responses can be obtained.

---

[39] The reference (but not free) software is "AutoSEA2"[TM]:
http://www.esi-group.com/SimulationSoftware/Vibro_acoustics.
For free software, see, for example: http://opensea.mub.tu-harburg.de/

Waveguides

The imagination of wave propagation in one dimension is perfectly represented by waveguides. They introduce propagation delay and attenuation due to divergence and damping. A spherical wave (omnidirectional) also can be described by waveguides. Points of reflections due to interfaces between different impedances are also easy to implement by connecting waveguides (delay lines) with transfer functions of reflection and transmission factors. At this point, it is obvious that waveguides contain forth- and backtravelling waves and thus are bidirectional.

In 2-D or 3-D wave propagation, the waveguide model is mapped to a corresponding CAD model, such as a room. The delay lines then are geometrically fixed at some point or patches on the walls. Otherwise, the combinations of geometric paths would increase exponentially. With geometric concentration of the waveguide nodes, a finite number of node connections results (Krämer 1994).

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 section. The famous "Kelly–Lochbaum" model of the vocal tract explains the formation of formants in vowels.

Frequency-dependent losses are included adding digital low-pass filters into the delay loop. FIR and IIR filter networks are used, partly involving sophisticated phase models and subsample shifts for better adjusting the actual geometric relations between nodes (Karjalainen 2005). The geometric conditions and corresponding wave divergence must be specifically included, except for 1-D application such as for wind instruments and vocal tracts.
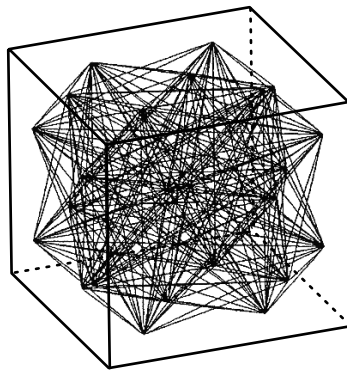
Fig. 10.11. Mapping room geometry on a set of coupled delay lines (after (Krämer 1994))
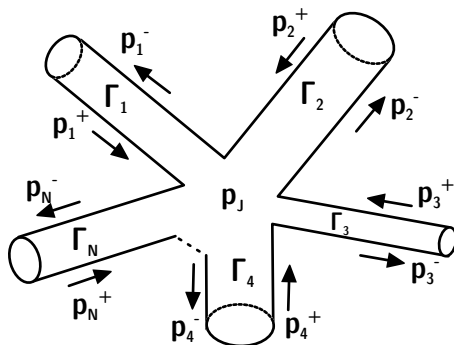
Fig. 10.12. Nodes coupling delay lines (after (Karjalainen 2005))
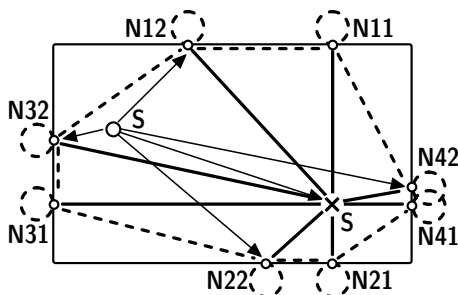
Fig. 10.13. Room model with waveguide network (after (Karjalainen 2005))

Geometrical acoustics

This model, already introduced for calculating basic features of room acoustics in Sect. 4.3, can be applied to 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 sound energy received. Because 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 immission prognosis are the most interesting fields. Accordingly, geometrical acoustics is well developed in these areas.

Ray construction is the key to geometrical acoustics. In elementary definition, rays will travel along the path with shortest travel time (Fermat's principle). Depending on the medium, this might lead to straight lines and specular reflections. In layered media such as in the atmosphere with height-dependent temperature or a wind profile, the rays are bent by refraction. But as in geometrical optics and discussion of lenses, refraction can well be included in the model. The strategy to construct rays is twofold:

– forward geometric construction from the source to the receiver
– reconstruction from the receiver to the source.

Basically these two approaches are equivalent, which is inherently given in the law of reciprocity (source and receiver may be interchanged). But the approaches also show partly 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 weighting are constructed. Impulse responses also may 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 irradiation and reradiation of energy from surface elements.[40] 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 formulated. This model is also known in illumination simulation on diffusely reflecting surfaces.

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

$$B(r,t) = \frac{1}{\pi} \iint_S (1 - \alpha(r'))B(r',t - \frac{R}{c}) \frac{\cos\vartheta\cos\vartheta'}{R^2} dS' + B_d(r,t). \qquad (10.26)$$

B_d is the direct sound intensity. Lambert's law of diffuse scattering is included in this integral equation. Note that exact timing is also included due to retardation by $t - R/c$.

---

[40] often called "patches."
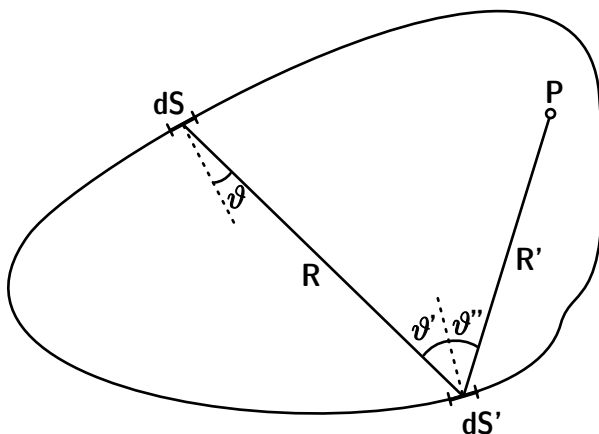
Fig. 10.14. Radiosity based on Kuttruff's integral equation (after (Kuttruff 2000))

The pure calculation of energy irradiation on the walls is interesting, but the energy at field point r in the interior field space is more relevant. This is achieved by

$$w(r,t) = \frac{1}{\pi c}\iint_S (1-\alpha(r'))B(r',t-\frac{R'}{c})\frac{\cos\vartheta''}{R'^2}\mathrm{d}S' + w_\mathrm{d}(r,t) \,. \qquad (10.27)$$

w_d denotes the direct sound (see Eq. (4.40)). In discretized form, the Integral (10.27) is exactly the basis for acoustic radiosity. It is used for calculating the total energy density at 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).

10.2 Two-port models

Many acoustical/mechanical and electrical equations are formally similar, for example,

$$U = L\frac{\mathrm{d}I}{\mathrm{d}t} \qquad F = m\frac{\mathrm{d}v}{\mathrm{d}t} \qquad I = C\frac{\mathrm{d}U}{\mathrm{d}t} \,. \qquad (10.28)$$

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 methods in analyzing electrical circuits must be seen in coupling current and voltage in complex circuits and in identifying

major and minor current flow paths. The model, therefore, is excellently qualified to analyze force and velocity ratios over in 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. Electromechanical analogies

	Analogy I		Analogy II
Voltage U	⟷ Force F	⟷	Current I
Current I	⟷ Velocity v	⟷	Voltage U
Electr. impedance	⟷ Mech. impedance	⟷	Electr. conductivity
Z_{el}	Z_m		Y_{el}
Resistance R	⟷ Friction losses w	⟷	Conductivity $1/R$
Inductivity L	⟷ Mass m	⟷	Capacity C
Capacity C	⟷ Spring n	⟷	Inductivity L

Analogy I is impedance conserving and analogy II is conserving the circuit plan. This means that an equivalent mechanical circuit developed from an electrical circuit will have the same general structure. Some examples are shown in Fig. 10.15. It is worth mentioning that electroacoustic transducers and, thus, coupled electrical-mechanical devices can also be modelled.
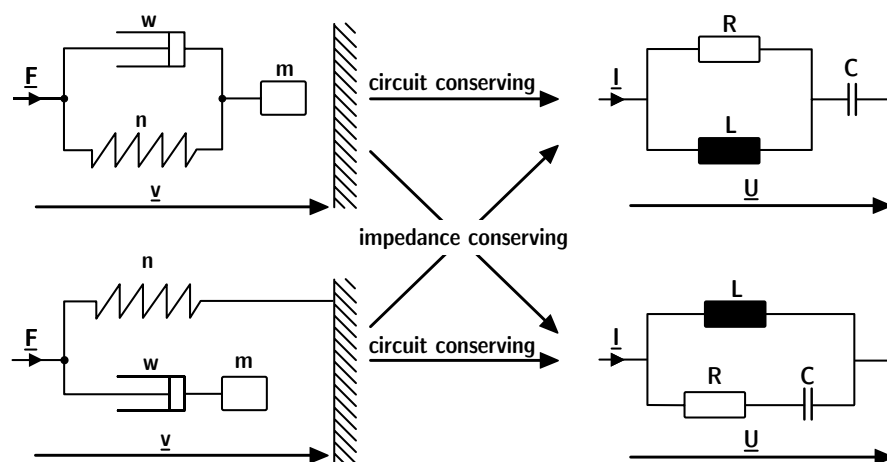

Fig. 10.15. Equivalent electromechanical circuits

Concentrated elements are valid in one-dimensional or multi-dimensional orthogonal signal transmission. Particularly 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 toward 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 a difference in the potential electric field. In the language of mechanics and acoustics, this difference is to be interpreted as force or 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, too, help in connecting two-ports in complex networks. Transfer functions, transfer impedances, matrix elements, etc., are spectral data that can be used in coupling two-ports and in simulating networks. If the network represents a system acting between a source and a receiver, it can be used perfectly for calculating auralization filters.

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) to the velocities on the opposite side,

$$\frac{F_{2o}}{v_1} = \frac{F_{1o}}{v_2} \, , \tag{10.29}$$

remains invariant. Infinite mechanical loads are achieved by blocking the motion.

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.

Fig. 10.16. Mechanical two-port, analogy I (force/voltage)
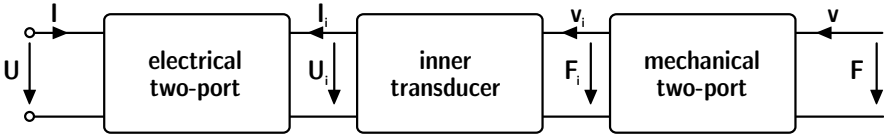
Fig. 10.17. Network representing an electromechanical transducer (analogy I)

Reciprocity also holds here, for instance, in the form of

$$\frac{F_{v=0}}{I} = \pm \frac{U_{I=0}}{v}$$

$$\frac{v_{F=0}}{I} = \mp \frac{U_{I=0}}{F}$$

(10.30)

with the upper sign denoting the rule for electric field transducer (force/voltage transduction), and the lower for magnetic field transducers with (force/current transduction). The left side represents an actuator (or loudspeaker), the right side a sensor (microphone).

Electroacoustic transducers such as loudspeakers and microphones are modelled as mechanical transducers with division of the force by their membrane area, S. The force, thus, becomes a pressure and the velocity a volume flow. This is represented exactly 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.

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, even distributed fields can be used as input and output data, if the fields are clearly defined in their modal contribution factors.
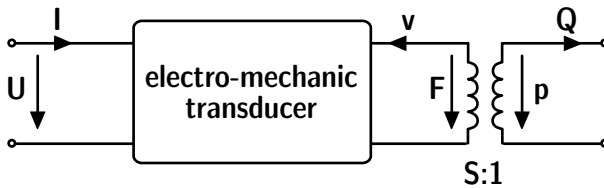
Fig. 10.18. Electroacoustic two-port (analogy I)

10.2.1 Transfer path models

Complex networks can be established on the basis of general two-port the-
ory. They can be used for airborne sound propagation, vibration propaga-
tion, structure-borne sound radiation and auralization. The energy trans-
mitted in the network is separated into paths, as illustrated in Fig. 10.19.
Ideal or real force or velocity sources are connected directly to the two-
port network. The difference between ideal and real sources is the imped-
ance coupling between the source and the transmission system.

For 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 should be modelled with feedback,
and this requires network analysis in a first step and signal flow calculation
in a second step.
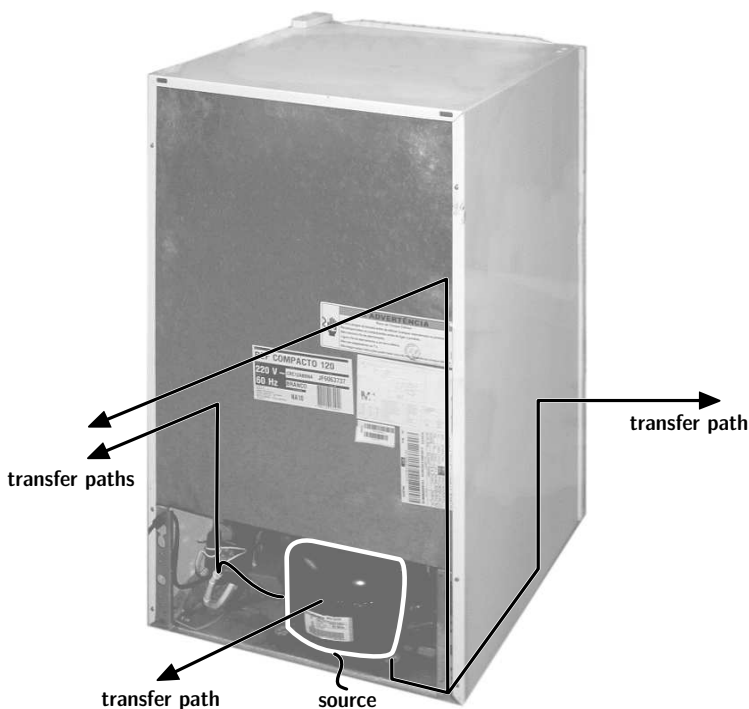
Fig. 10.19. Transfer path separation. Illustrated with the example of a refrigerator
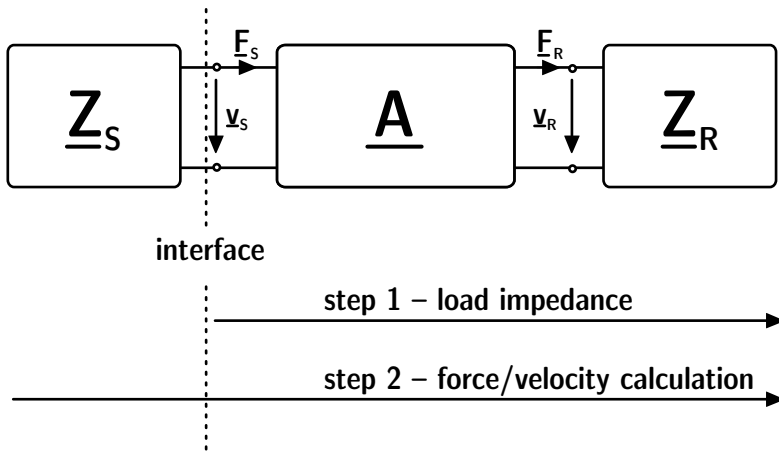
Fig. 10.20. Two-port network of source impedance $\underline{Z}_S$, transfer matrix $\underline{\mathbf{A}}$ and receiver impedance, $\underline{Z}_R$ in analogy II (force/current)
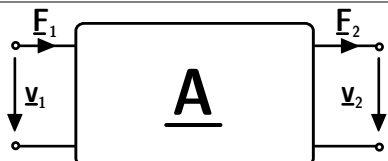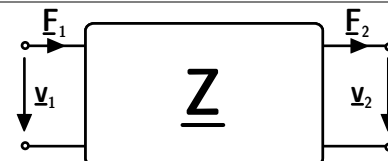
As illustrated in Fig. 10.20, in the first step, the two-port network must be analyzed for the total impedance at the interface to the source. Only when this load impedance is known, can the power output of the source be calculated, in dependence on the inner impedance of the source. This general principle must be applied to simple one-dimensional networks and to complex matrix formulations.

The matrix $\underline{\mathbf{A}}$ (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. 10.2). The matrix can also be rearranged into a pure impedance matrix, $\underline{\mathbf{Z}}$:

$$\begin{pmatrix} \underline{F}_1 \\ \underline{F}_2 \end{pmatrix} = \begin{pmatrix} \underline{Z}_{11} & \underline{Z}_{12} \\ \underline{Z}_{21} & \underline{Z}_{22} \end{pmatrix} \cdot \begin{pmatrix} \underline{v}_1 \\ \underline{v}_2 \end{pmatrix} = \underline{\mathbf{Z}} \cdot \begin{pmatrix} \underline{v}_1 \\ \underline{v}_2 \end{pmatrix}. \tag{10.31}$$

The two-port parameters $\underline{Z}_{11}$ to $\underline{Z}_{22}$ are called input and output impedances, respectively, while $\underline{Z}_{12}$ and $\underline{Z}_{21}$ are called transfer impedances. $\underline{\mathbf{A}}$ can be easily transformed into $\underline{\mathbf{Z}}$. For reciprocal networks, det $\underline{\mathbf{A}} = 1$ holds. The task in transfer path analysis is determining the matrix elements; see also Chap. 14.

Table 10.2. Two-port matrix conversion, after (Dohm 2004)

Chain form	Impedance form

$$\begin{pmatrix} \underline{F}_1 \\ \underline{v}_1 \end{pmatrix} = \begin{pmatrix} \underline{A}_{11} & \underline{A}_{12} \\ \underline{A}_{21} & \underline{A}_{22} \end{pmatrix} \cdot \begin{pmatrix} \underline{F}_2 \\ \underline{v}_2 \end{pmatrix} = \underline{\mathbf{A}} \cdot \begin{pmatrix} \underline{F}_2 \\ \underline{v}_2 \end{pmatrix} \qquad \begin{pmatrix} \underline{F}_1 \\ \underline{F}_2 \end{pmatrix} = \begin{pmatrix} \underline{Z}_{11} & \underline{Z}_{12} \\ \underline{Z}_{21} & \underline{Z}_{22} \end{pmatrix} \cdot \begin{pmatrix} \underline{v}_1 \\ \underline{v}_2 \end{pmatrix} = \underline{\mathbf{Z}} \cdot \begin{pmatrix} \underline{v}_1 \\ \underline{v}_2 \end{pmatrix}$$

$$\begin{pmatrix} \underline{A}_{11} & \underline{A}_{12} \\ \underline{A}_{21} & \underline{A}_{22} \end{pmatrix} \qquad\qquad \underline{\mathbf{A}} = \begin{pmatrix} \dfrac{\underline{Z}_{11}}{\underline{Z}_{21}} & \dfrac{\det \underline{\mathbf{Z}}}{\underline{Z}_{21}} \\[2mm] \dfrac{1}{\underline{Z}_{21}} & \dfrac{\underline{Z}_{22}}{\underline{Z}_{21}} \end{pmatrix}$$

$$\underline{\mathbf{Z}} = \begin{pmatrix} \dfrac{\underline{A}_{11}}{\underline{A}_{21}} & \dfrac{\det \underline{\mathbf{A}}}{\underline{A}_{21}} \\[2mm] \dfrac{1}{\underline{A}_{21}} & \dfrac{\underline{A}_{22}}{\underline{A}_{21}} \end{pmatrix} \qquad\qquad \begin{pmatrix} \underline{Z}_{11} & \underline{Z}_{12} \\ \underline{Z}_{21} & \underline{Z}_{22} \end{pmatrix}$$

The prerequisite of concentrated elements should be discussed in more detail. The elastic and viscous effects of large construction elements such as dampers or springs, cannot always be separated into concentrated elements. At small wavelengths, too, they act as waveguides of propagating or standing waves. As in modelling electromagnetic waveguides, they can be modelled as quasi-continuous networks of parallel circuits, each representing a small interval Δx which is appropriate for one-dimensional wave propagation; see also Sect. 10.1.3, Waveguides.

10.3 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 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 for creating mixtures of tones and noise 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 modelling musical sounds.

For studies of the propagation paths without the need for predicting propagation spectra exactly, wave-front synthesis by using finite time differences offer an insight into the field of travelling waves. In mesh-based time domain models, there is no a priori assumption of wave types. The waves are developing 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 discretized acoustic field equations (Eqs. (1.11) and (1.12)). They yield (here, in one-dimensional form)

$$-\frac{\Delta p}{\Delta x} = \rho_0 \frac{\Delta v}{\Delta t}, \tag{10.32}$$

$$-\rho_0 \frac{\Delta v}{\Delta x} = \frac{1}{c^2} \frac{\Delta p}{\Delta t} + \rho_0 q. \tag{10.33}$$

The problem, however, is that the wave propagation, its speed and its interaction with boundaries can be solved exactly only for 1-D cases. Examples of 2-D problems were presented as approximations, but FTD in 3-D domains suffers from severe artefacts such as dispersion, unless corrections are introduced (Savioja and Välimäki 2000). The reason 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. The finite time difference method, however, is useful for illustration and animation of wavefront propagation.

11 Simulation of sound in rooms

11.1 General

Computer modelling of room acoustics was 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 still a powerful tool 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, 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 programs are based on geometrical acoustics (Sect. 10.1.3). In geometrical acoustics, the description of the sound field is reduced to energy, transition time and the direction of rays. This approach is correct as long as the dimensions of the room are large compared to wavelengths and if broadband signals are considered. These approximations are valid with sufficient accuracy in large rooms above the cutoff frequency f_c; see Eq. (4.13).

In intercomparison tests (so-called "round-robin tests") (Vorländer 1995; Bork 2000, 2005a, 2005b), the efficiency and the limits of room acoustics computer simulation (see Sect. 11.5) were evaluated. Finally, auralization has been included in room acoustics simulation since the beginning of the 1990s (see Sect. 11.6).

In this chapter, we will discuss the fundamental algorithms, the variations of their implementation in software and their efficiency. It will be shown that pure specular models (image models) are not capable of simulating room sound fields accurately enough, but combinations of models will allow simulation with acceptable plausibility. Hybrid models that can handle specular and diffuse reflections for estimating the late reverberation spectrum are the solution to obtaining impulse responses very near measurement results.

The methods are, at first, used to calculate the room acoustic criteria (T, EDT, D, C, T_S, LF, IACC …); see Sect. 6.4 to get a more specific result than given by the estimate (Sect. 6.4.6).

Two techniques of geometrical acoustics have to be distinguished: "ray tracing" and "image sources." Independent of their software implementation, they represent different physical approaches. To achieve a clear understanding, this point must be stressed, since ray tracing algorithms can well be used to calculate and handle image sources. The main difference between ray tracing (in the classical definition) and image sources is the way energy detection and the internal nature of physical energy propagation are implemented.

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

11.1.1 CAD room model

To implement 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 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 such as 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 polygonal complexity, it would also be wrong. As will be seen 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 Sect. 3.3). Objects or surface corrugations that are not large compared with wavelengths, have to be taken out the CAD model and replaced by flat surfaces with adequate acoustic properties. This holds 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.

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 and
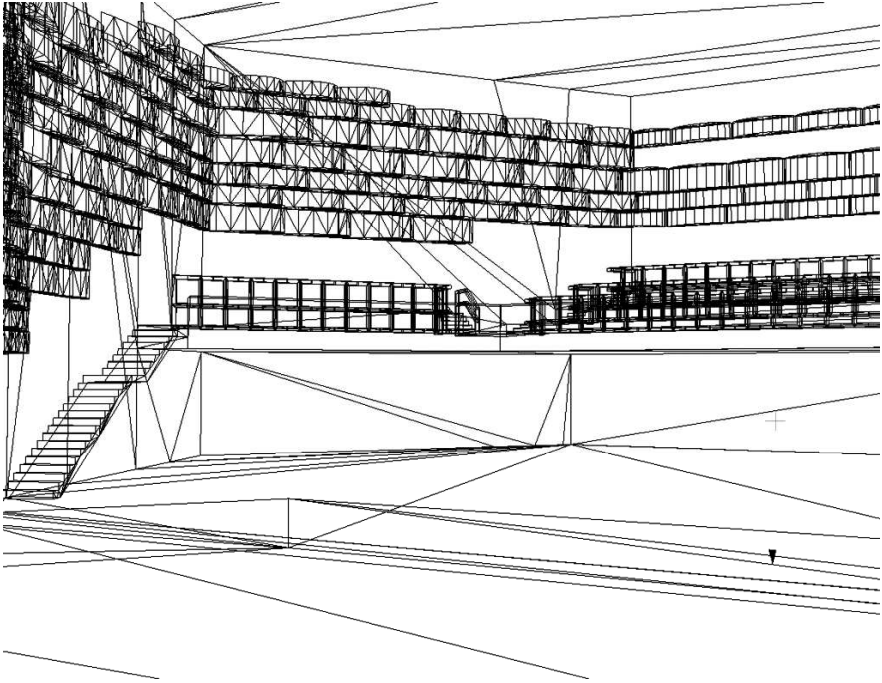
Fig. 11.1. Visual wire-frame model of a multipurpose hall

above that frequency, as partly scattering surfaces or objects. 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 wavelengths small compared with 0.5 m, say, $\lambda = 5$ cm. Then, frequencies around 7 kHz and above are discussed. For several reasons, this would be of interest only in very special cases. The main arguments for choosing a range of lower frequencies are as follows:

1. The spectral content excited by natural sources such as the voice of 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 details toward a specific acoustically relevant resolution. Polygonal smoothing by using spatial low-pass filters with a cutoff wave number of $125\,\text{m}^{-1}$ (equivalent to a minimum length resolution of 50 cm) could be an interesting option for future automated CAD user interfaces.
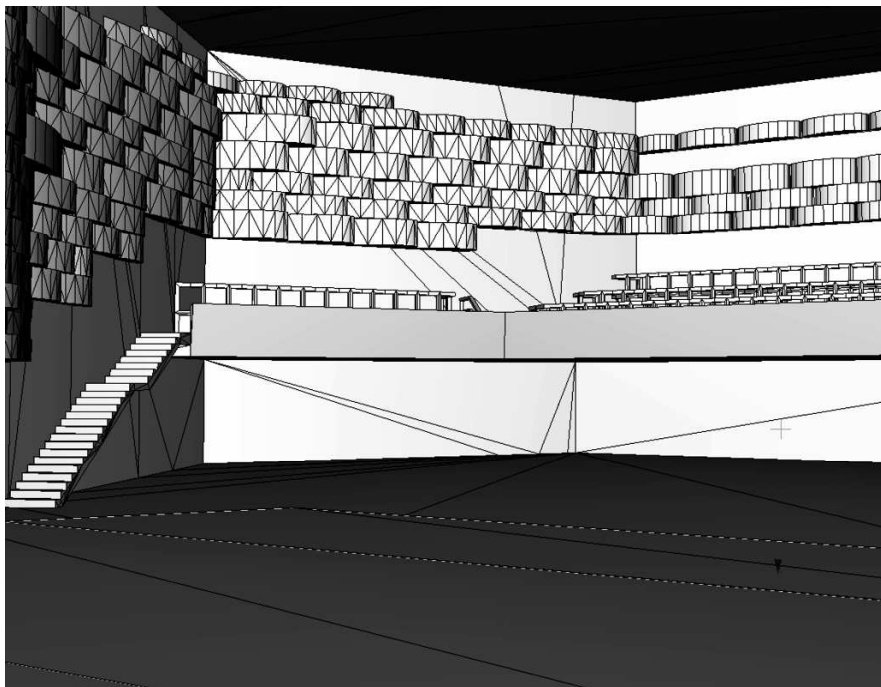
Fig. 11.2. Visual surface model of a multipurpose hall

In the end, the room must be approximated by analytic surfaces, typically by planes. Surfaces of higher order (cylindrical, spherical, parabolic, …) can be used in principle (see below). All surfaces lying in a common plane form a "wall." A wall, thus, need not 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 at all in modelling a room shape.

All plane surfaces are defined by three points in a previously fixed co-ordinate system. Preferably points (called "vertices"[41]) 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 polgon's plane can be allowed, if the new corners are automatically adjusted to match the plane (at least numerically) exactly.

---

[41] from "vertex"
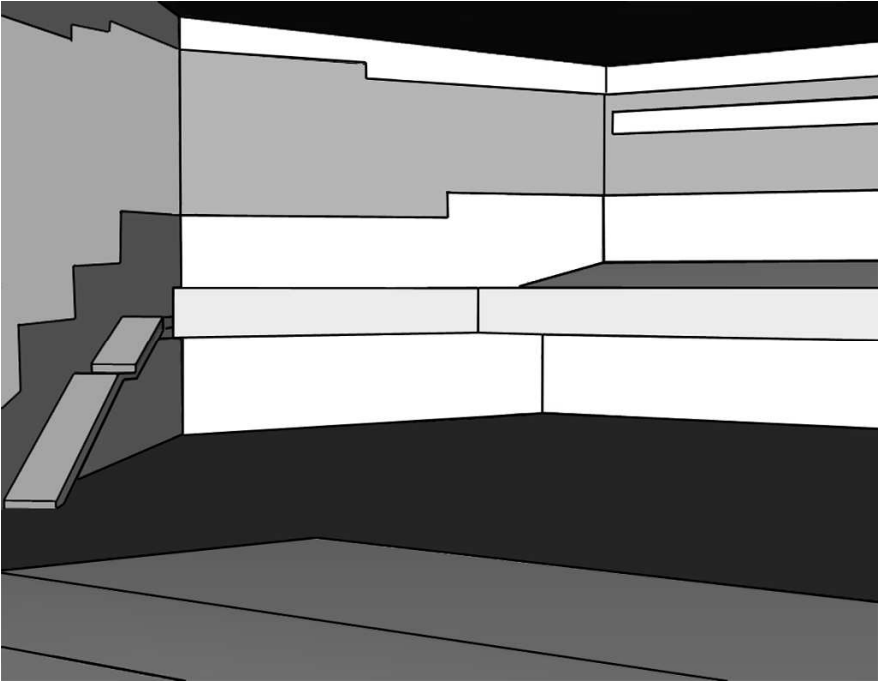
Fig. 11.3. From architectural to acoustic CAD models: Acoustic surface model of a multipurpose hall. Geometric details are replaced by smooth wall polygons
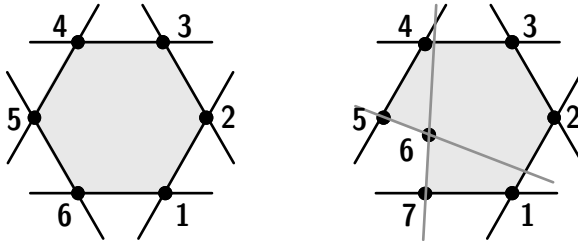

Fig. 11.4. Left: Convex polygon. Right: Concave polygon

Figure 11.5 illustrates the strategy for building a 3-D room from 2-D polygons. The polygon-to-vertex notation, the polgon orientation with reference to the normal vector and their address management must be specified clearly.
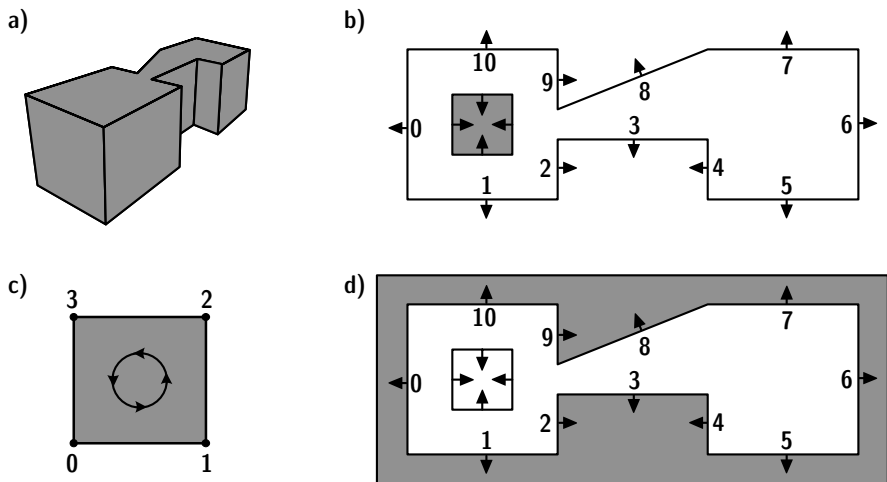
Fig. 11.5. a) Example: A simple room constructed by a set of convex polygons. b) The room's 2-D projection: the green area marks a solid object inside the room. c) Definition of the spanning direction for polygons. d) The room's 2-D projection: the green area marks the outside world (example after (Schröder 2004; Schröder and Lentz 2006)

11.1.2 Absorption coefficients

Absorption coefficients (see Annex) can be found in tables (see Annex), or they can be taken from specific test results. Most absorption coefficients correspond to diffuse sound incidence because they were obtained from measurements in reverberation chambers (ISO 354). In ray tracing, these data represent the average absorption at random-incidence. Thus, in rooms that 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 precisely. Nevertheless, the errors are small. It should be kept in mind that the difference in energy loss at a specific angle of incidence to random-incidence may be up to 40%, until the error of the reflection level exceeds 2 dB. Only at grazing incidence, may the errors become larger. But at grazing incidence, anyway, other influences come into play (see Sect. 11.5).

   If the complex wall impedance is known, the angle-dependent absorption coefficient can be calculated. This is easy for real-impedance, locally reacting walls (see Eqs. (3.3)–(3.6)). Each case of room simulation should be checked to determine if random-incidence absorption coefficients are applicable or if angle-dependent data are possibly required. The latter may apply to long or flat rooms, tunnels, corridors and coupled rooms, for example.

11.1.3   Scattering coefficients

If absorption is treated in a simplified way (assuming random incidence), this 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. 3.3). Data from free-field measurements (part 2 of (ISO 17497)) 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 that are small in high-order reflections. But we must be aware of problems which can arise in 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 ISO 17497, more information may be published in the future (see Annex).

11.2   Stochastic ray tracing

In this model, sound is radiated as bunches of particles – <u>many</u> particles (= rays). The number of particles is one of the crucial features of ray tracing. It is considered in the category of Monte Carlo methods to express the influence of probabilistic effects, such as in gambling.

Stochastic ray tracing in short

The sound source radiates an impulse. To simulate this, particles are started at the same initial time in various directions. Each ray carries certain energy, propagates at the speed of sound and once in a while hits the room boundary, for simplicity called a "wall" throughout. It is reflected from the wall, hits another wall and so on. Each wall absorption reduces the particle's energy.
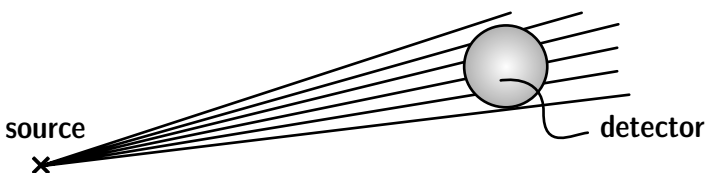
Fig. 11.6. Free field propagation and distance law "by counting"
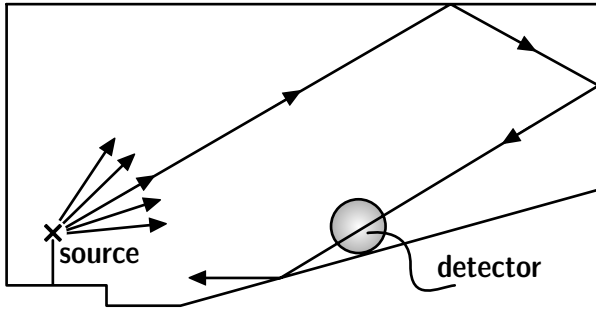
Fig. 11.7. Tracing a ray from the source to the detector
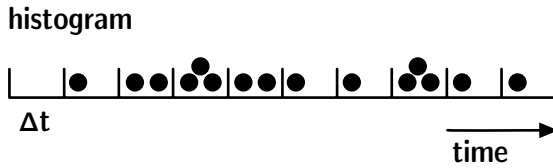
Fig. 11.8. Creating an impulse response by counting events

By means of particle detectors, the particle's energy and the time elapsed since radiation from the source are registered (see Fig. 11.8). The number of counts represents the energy detected at the receiver.

Source modelling

Sound sources are characterized by sound power and directivity (Sect. 2.4). Both depend on frequency. For ray tracing, we need just 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 particles, as illustrated in Fig. 11.9 in an area-accurate plot of the unit sphere.

To create spherically uniform radiation, we proceed as follows: The azimuthal angles are distributed evenly around the perimeter. The distribution of polar angles is to be chosen so 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

$$\theta = \begin{cases} \arccos z \\ \arccos z + \pi/2 \end{cases}, \qquad (11.1)$$

the upper choice for the northern and the lower for the southern hemisphere. $\theta = 0$ at the north pole. Other weighting functions create polar
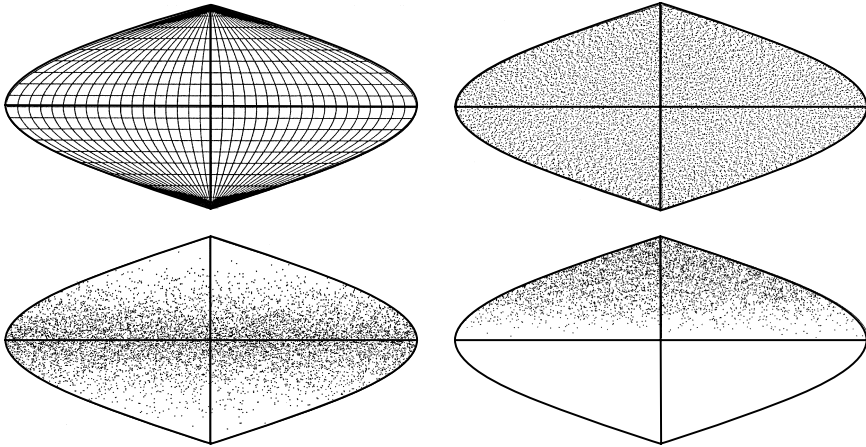
Fig. 11.9. Top left: Variants of ray sources, spherical coordinate system and regular grid. Top right: Uniform random distribut ion. Bottom left: Concentration of rays at the equator (line-array-like). Bottom right: Concentration of rays at the north pole (baffled-piston-like)

distributions with concentrations 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.[42] 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 that the energy can be reduced by multiplying the incident energy by the 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.

---

[42] Under special circumstances, parallel processing of frequency bands is possible; see below.
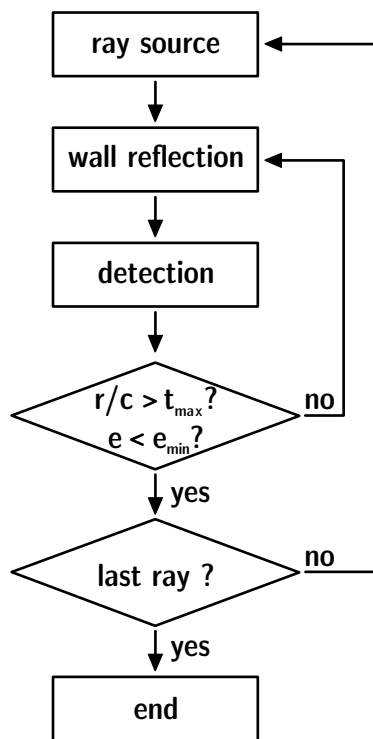
Fig. 11.10. Flow diagram of ray tracing

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, such as t_{max} or e_{min}.

11.2.1 Point-in-polygon test

The most inner loop of ray tracing algorithms is the test of whether a ray (particle), represented as a line segment, a) is hitting a plane and b) if 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 are known. Also known is the set of polygons.

The polygons qualified (those located in forward direction of the ray) are transformed in 2-D 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 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 once in the opposite direction.

If the intersection point is inside the polygon, the new travel line will be input variable for the next plane hit. Before reflection takes place, wall materials affect sound by absorption, scattering and diffraction. If a particle hits a wall, it will lose energy or be annihilated,[43] and it will change its travel direction. For specular reflection, a new direction at a specular angle with reference to the normal incidence will be calculated.

When scattering occurs, which applies when a random number exceeds the scattering coefficient, the new direction is obtained from two more random numbers. Mostly, a directional distribution is assumed according to Lambert's law. On average, the polar angles of the new flight direction (with reference to the wall normal vector) must be distributed according to

$$w(\theta)\mathrm{d}\Omega = \frac{1}{\pi}\cos\theta \,\mathrm{d}\Omega \,. \tag{11.2}$$

The azimuthal angle is evenly distributed in $(0, 2\pi)$. Hence, two random numbers z_1 and z_2 in $(0,1)$ and transformations

$$\theta = \arccos\sqrt{z_1} \tag{11.3}$$

and

$$\phi = 2\pi \, z_2 \tag{11.4}$$

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.

11.2.2 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

$$M_{\mathrm{sphere}} = \pi \, r_{\mathrm{d}}^2 \,, \tag{11.5}$$

where r_{d} denotes the detector radius. Thus, the detector position (centre point) is sufficient information. Each line segment representing the ray has to be checked to determine whether it "hits" the detector. This test requires

---

[43] depending on the way absorption is implemented.

calculating 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 method 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 heads. Including both (wave) effects in models of detectors was tried, but all attempts suffer from inconsistencies with geometrical acoustics. Anyway, audience effects can be modelled as a 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, but in deterministic models (see below).

11.2.3 Presentation of results

Up to this point, ray tracing has been described abstractly. 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 a histogram of the energy impulse response. The array must have a sampling rate that allows a sufficiently high temporal resolution, on the one hand, and sufficiently large integration intervals, on the other (see below, Sect. 11.2.5). A good compromise is a resolution of the order of magnitude of milliseconds. Psychoacoustic post-masking can be taken into account properly, but binaural attributes such as ITD cannot[44] (see Fig. 11.11). At the same time, it is clear that we cannot meet the requirement of a sampling rate adequate for audio signal processing and auralization.[45]

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 in accuracy and temporal resolution would only mimic a real gain in physical accuracy. Ray tracing still remains an approximation of wave propagation.

---

[44] It will be shown in Sect. 11.2.5 that the required number of rays would be unacceptably high.

[45] The temporal resolution for a sampling frequency of 44.1 kHz is 22.7 µs.
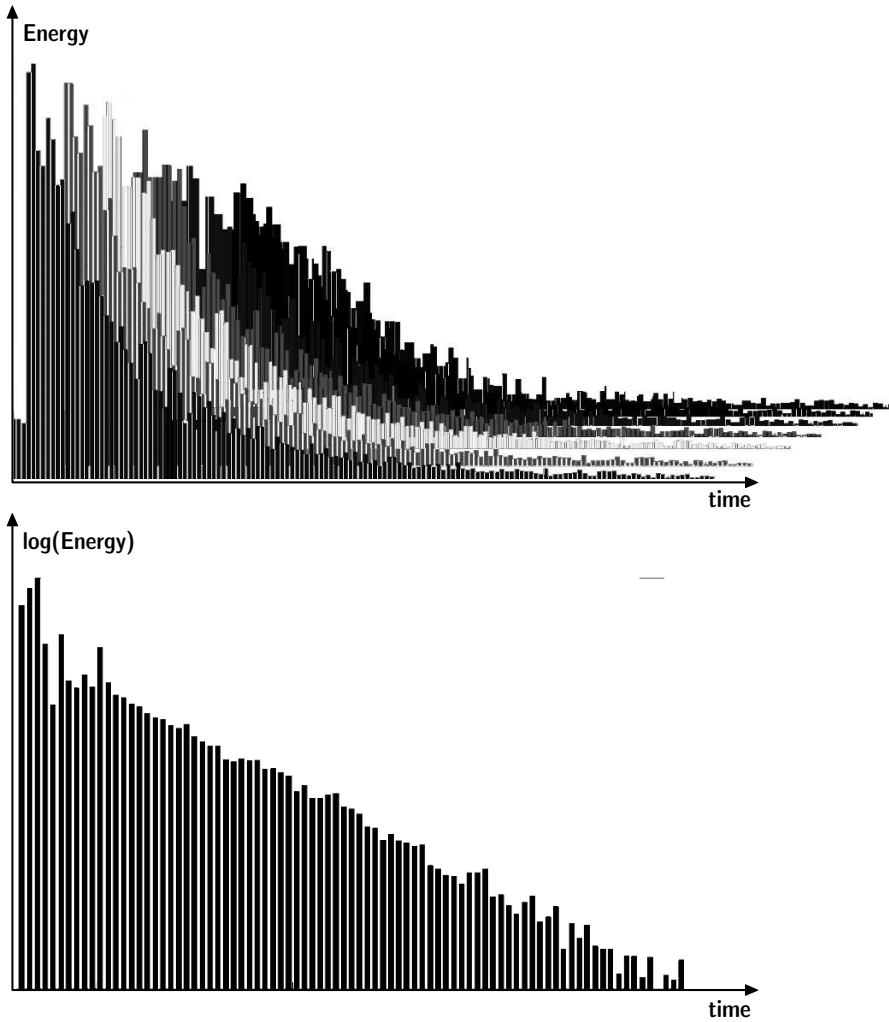
Fig. 11.11. Ray tracing energy impulse response. Top: Linear plot for several frequency bands. Bottom: Logarithmic plot for one frequency band
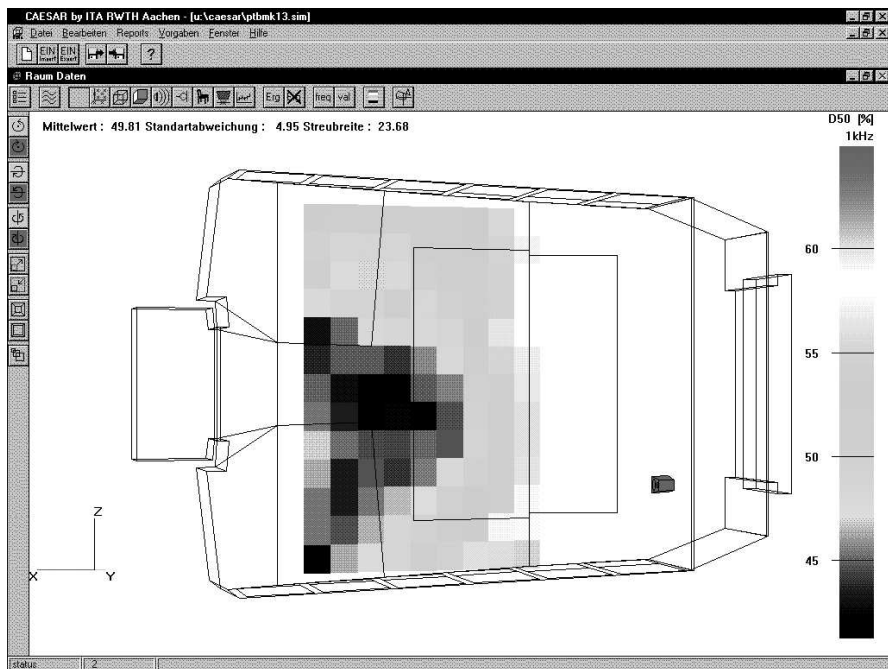
Fig. 11.12. Examples of room acoustic parameters plotted in the CAD model (audience area); see Sect. 6.4 and particularly Sect. 6.4.6

11.2.4 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 uniform distribution of good acoustic quality. In rooms with cylindrical side walls or with dome-shape ceilings, we have to expect focusing. If room acoustic computer simulation shall be a tool for design, it must be ensured that occurrences of focusing and echoes are clearly visible in the results.

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 (Kuttruff 1995) that deterministic approaches with coherent image source contributions can be used. This is obvious since the sound pressure in the focal region cannot be obtained by energetic models.
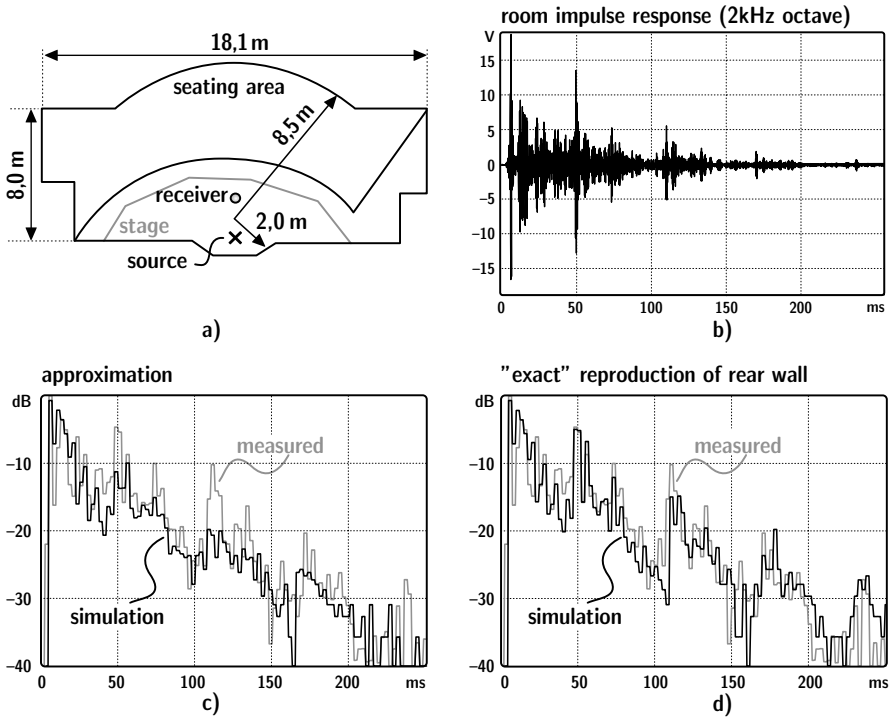
Fig. 11.13. (a) Ground plan of a small theatre (volume 900 m³) with source and receiver positions, (b) Measured impulse response (2 kHz octave) and comparison with simulated impulse responses. (c) Approximation of five plane surfaces, (d) Mathematically exact model of the cylindrical back wall, after (Mommertz 1996)

The approximation of a cylinder with radius a by planes of width b is sufficient if

$$f\, b^2 = \frac{1}{2} c\, a \tag{11.6}$$

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 as frequency dependent. For example, a cylindrical shape with radius 10 m must be modelled by panels 50 cm wide, if the frequency range reaches 8 kHz This means that the cylinder should be subdivided into 136 plane elements.

11.2.5 Reproducibility in stochastic ray tracing

Systematic uncertainties are present in all physical measurement and simulation methods. The approximations inherent in the input data (room geometry, absorption and scattering coefficients) and in the model itself (approximation of wave physics by a geometric model) are causes for systematic errors. These errors cannot be reduced by increased computational effort.

Independently, in Monte Carlo methods such as ray tracing, we must consider stochastic fluctuations in the results. These errors can be reduced by increasing the number of rays or by averaging repeated simulations. The number of rays launched, N, is of crucial importance because it affects both the reproducibility and the computation time t_{calc}.

The goal of this chapter is predicting the standard deviation in the simulation results in dependence on N and t_{calc}. To obtain an expectation value of the error, we 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 $<k>$, thus, is constant and follows a Poisson distribution.

$$P(k) = \frac{\langle k \rangle}{k!} e^{-\langle k \rangle} \tag{11.7}$$

The variance of the Poisson distribution, σ^2, describes the fluctuations around the expectation value. It is

$$\sigma_k^2 = \langle k \rangle = \langle r \rangle \Delta t . \tag{11.8}$$

The number of hits is thus related to the following expected error:

$$\sigma_k / \langle k \rangle = \frac{1}{\sqrt{\langle k \rangle}} . \tag{11.9}$$

The expectation value of the count rate $<r>$ is calculated from the number of wall hits per second, $\bar{n}N$, the ratio of the detector surface, $4\pi r_d^2$, and the room surface, S:

$$\langle r \rangle = \bar{n}N \frac{4\pi r_d^2}{S} . \tag{11.10}$$

Herein, $\bar{n}$ is the mean reflection rate according to Eq. (4.22). The mean count rate is, thus,

$$\langle r \rangle = N \frac{\pi r_{\mathrm{d}}^2 c}{V},$$

(11.11)

and the mean number of hits is

$$\langle k \rangle = N \frac{\pi r_{\mathrm{d}}^2 c \Delta t}{V}.$$

(11.12)

Now we calculate the expectation value of the energy, $<E>$, 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 a best guess for a room sound field. This approach is, by the way, exactly the same as in statistical reverberation theory (Sect. 4.4). Of course, an exponential decaying $<E>$ will be found, which matches the late decay better in a real room. Thus, early reflections are not discussed here. The theory is identical to that discussed in Sect. 4.4. Therefore,

$$E(t) = e_0 N (1 - \alpha)^{\bar{n}t} = e_0 N e^{-\frac{13.8t}{T}},$$

(11.13)

where e_0 denotes the starting 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. (11.12)). The mean energy is, thus,

$$\langle E \rangle = \langle k \rangle e_0 e^{-\frac{13.8t}{T}}$$

(11.14)

and

$$\frac{\sigma_E}{\langle E \rangle} = \frac{\sigma_k}{\langle k \rangle} = \sqrt{\frac{N\,V}{\pi r_{\mathrm{d}}^2 c \Delta t}}.$$

(11.15)

The relative error $\sigma_E / <E>$ is easier to interpret when expressed in decibels. Now, we can use Eq. (11.16) which states that the time-dependent sound pressure level is

$$L_{\mathrm{ETC}} = 10 \log(E / E_0)$$

(11.16)

with an arbitrary reference E_0, and

$$\sigma_{\text{ETC}} \approx \frac{\mathrm{d}L}{\mathrm{d}E}\sigma_E = \frac{10}{\ln 10}\frac{\sigma_E}{\langle E\rangle} \, . \tag{11.17}$$

Thus,

$$\sigma_{\text{ETC}} = 4.34\sqrt{\frac{V}{N\pi r_\mathrm{d}^2 c\Delta t}} \, . \tag{11.18}$$

This error is constant with respect to the time variable in the impulse re-sponse. It depends on the ratio of the room volume to the number of rays, the detector size and the width of the time intervals, $V/Nr_\mathrm{d}^2\Delta t$.

Furthermore, on the basis of E, integral parameters can be obtained, such as clarity or strength. The strength, for instance (see Eq. (6.16)), is derived from the integral[46]

$$\langle w\rangle = \sum_{\Delta t}\langle E\rangle \approx e_0 N\frac{\pi r_\mathrm{d}^2 c}{V}\int_0^\infty e^{-\frac{13.8t}{T}}\,\mathrm{d}t$$
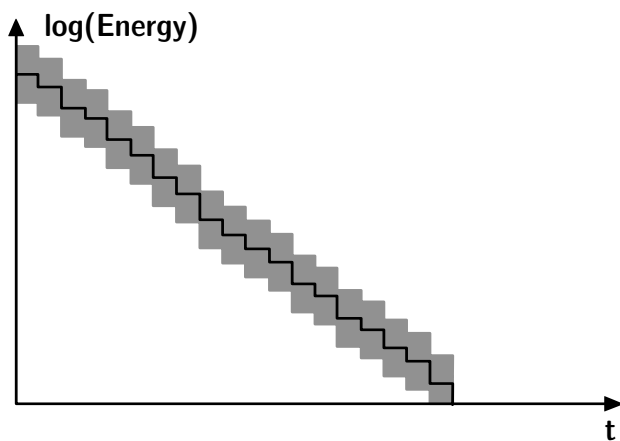$$= \frac{4\pi r_\mathrm{d}^2 e_0 N}{A} \tag{11.19}$$

Fig. 11.14. Histogram on a logarithmic scale of stochastic ray tracing (absorption by energy multiplication), expectation values and uncertainty range according to Eq. (11.18) (example)

---

[46] Accordingly, the sound power of the source, P, is equivalent to $\pi c r_\mathrm{d}^2 e_0 N$.

The expectation value of this energy integral, $<w>$, 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

$$\sigma_w^2 = \sum_{\Delta t} \sigma_E^2 .$$

(11.20)

and, thus, again expressed in decibels (applying the principle of Eq. (11.17) again),

$$\sigma_L = 4.34 \sqrt{\frac{A}{8\pi N r_d^2}} .$$

(11.21)

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, such as in measurements, to overcome the limit of the background noise floor. The noise, here, is represented by numerical noise given by the discrete ray formulation.

Finally, the computation time shall be estimated. To obtain the number of operations, we have a look at 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, the number of detectors, n_d. The computation time required for the point-in-polygon test is t_w, the time for
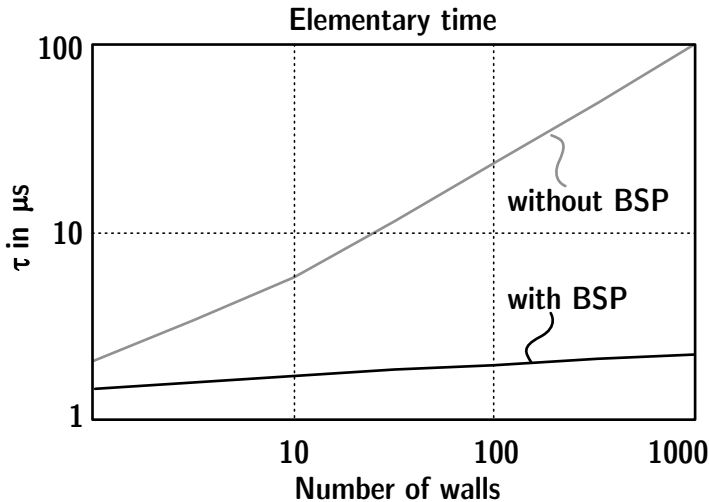
Fig. 11.15. Order of magnitude of elementary computation time per reflection as a function of the number of wall polygons, without and with binary space partitioning (example for PC with 2 GHz CPU)

checking detectors is t_d. Without spatial data structures such as the voxel technique or binary space partitioning (BSP) (see Sect. 11.3.5), the complete set of walls and detectors must be tested. The computation time for one reflection is considered elementary time, τ:

$$\tau \approx n_w t_w + n_d t_d + t_c , \tag{11.22}$$

where t_c denotes a constant offset for data management independent of the number of walls and detectors.

With BSP or similar spatial data structures, the elementary time reduces to

$$\tau \approx \log_2 n_w \ t_w + \log_2 n_d \ t_d + t_c . \tag{11.23}$$

The elementary time is to be multiplied by the events of reflections, $N\bar{n}t_{max}$:

$$t_{calc} \approx N\bar{n}t_{max}\tau \tag{11.24}$$

As a rough estimate, the following elementary computation times can be expected (see Figs. 11.15 and 11.16). 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 at $n_d = 1$.
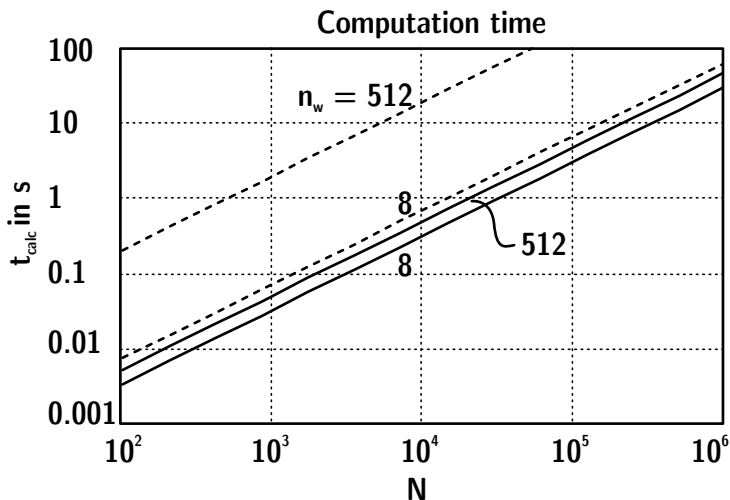
Fig. 11.16. Order of magnitude of the total computation time as a function of the number of rays, N (example for PC with 2 GHz CPU). The upper two broken lines represent classical ray tracing, the lower curves ray tracing with room subdivision strategy (example BSP). The room volume is the same in all four cases

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 case of absorption, the next ray is started.

The expectation value of the energy decreases with the running time in the impulse response since the number of hits decreases exponentially,

$$\langle k \rangle = N \frac{\pi r_{\mathrm{d}}^2 c \Delta t}{V} e^{-\frac{13.9t}{T}} , \qquad (11.25)$$

and because of

$$\langle E \rangle = \langle k \rangle e_0 , \qquad (11.26)$$

```
        ┌──────────────┐◄──────┐
        │  ray source  │       │
        └──────┬───────┘       │
               ▼               │
        ┌──────────────┐◄────┐ │
        │wall reflection│    │ │
        └──────┬───────┘     │ │
               ▼             │ │
        ┌──────────────┐     │ │
        │  detection   │     │ │
        └──────┬───────┘     │ │
               ▼             │ │
          ◄───────►  no      │ │
           z < α? ───────────┘ │
          ◄───────►            │
               │ yes           │
               ▼               │
          ◄───────►  no        │
          last ray ? ──────────┘
          ◄───────►
               │ yes
               ▼
        ┌──────────────┐
        │     end      │
        └──────────────┘
```

Fig. 11.17. Flow diagram of ray tracing (annihilation)

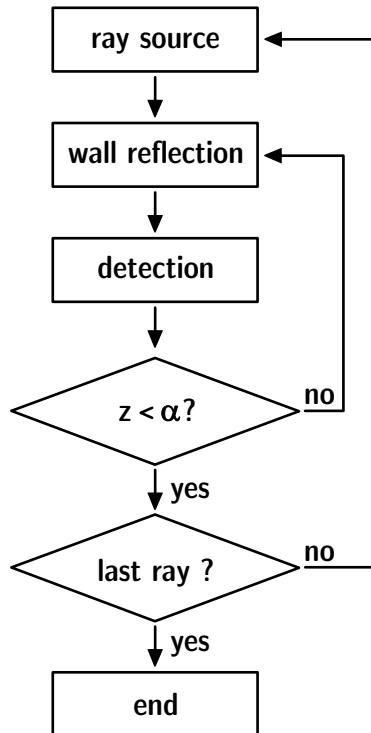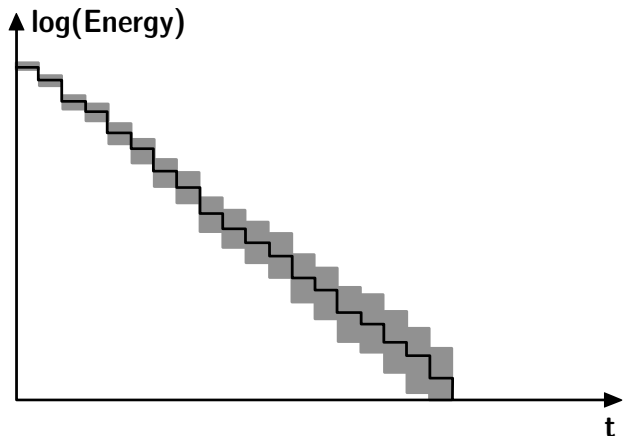**Fig. 11.18.** Histogram of stochastic ray tracing (absorption by annihilation), expectation values and uncertainty range according to Eq. (11.27) (example)

there results an increasing relative variation in the energy impulse response of

$$\frac{\sigma_E}{\langle E \rangle} = \frac{\sigma_k}{\langle k \rangle} = \sqrt{\frac{V \cdot e^{\frac{13.8t}{T}}}{N\pi r_d^2 c\Delta t}} = \sqrt{\frac{V}{N\pi r_d^2 c\Delta t}} \cdot e^{\frac{6.9t}{T}} . \tag{11.27}$$

The increase in 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

$$\sigma_L = 4.34 \sqrt{\frac{A}{4\pi N r_d^2}} . \tag{11.28}$$

When comparing the equation with Eq. (11.21), it is observed that the difference in the integral energy error is a factor of 0.7 in advantage of the energy multiplication. This is mainly due to the fact that rays disappear and leave 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 $\nu - 1$ reflections and is annihilated at the νth reflection, is

$$w(\nu) = (1 - \overline{\alpha})^{\nu-1} \overline{\alpha} , \tag{11.29}$$

where $\overline{\alpha}$ denotes the mean absorption coefficient (Eq. (4.30)). Accordingly,

$$\langle v \rangle = \sum_{v=0}^{\infty} v\, w(v) = \overline{\alpha} \sum_{v=0}^{\infty} (1 - \overline{\alpha})^{v-1} ,$$
$$= \frac{1}{\overline{\alpha}}$$

(11.30)

and

$$t_{\text{calc}} = \frac{N\tau}{\overline{\alpha}} .$$

(11.31)

11.2.6 Computation times versus uncertainties – case studies

In the following, we consider four examples of rooms as typical cases for categories of rooms used for different purposes: A living room ($V = 100\,\text{m}^3$), a lecture room (classroom) ($V = 1{,}000\,\text{m}^3$), a concert hall ($V = 10{,}000\,\text{m}^3$), and a large church ($V = 100{,}000\,\text{m}^3$). The rooms are modelled with constant average absorption of $\overline{\alpha} = 0.15$. The increasing number of walls, n_{w}, is due to the larger complexity with increasing volume. With constant average absorption, the reverberation time also increases.

Table 11.2. Examples of room situations

	Room 1 "living"	Room 2 "lecture"	Room 3 "concert"	Room 4 "church"
Volume in m^3	100	1,000	10,000	100,000
Surface area in m^2	136	906	4,725	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

In these categories of rooms, we compare the computation times which are required to achieve a certain quality of results (one receiver, one frequency band). At first, impulse responses are discussed. They may be intended either for visual inspection (see Fig. 11.11) of the quality of the energy impulse response or as a basis for further processing with the aim of auralization (see Sect. 11.7). 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 illustrated in Fig. 11.18, 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 of 1 dB,[47] the required number of rays, N, is defined and the computation times, t_{calc}, in the following table can be expected. The data, of course, represent only an order of magnitude, since they strongly depend on the hardware used. A standard PC with 2 GHz and 1 Gbyte RAM serves as an example. The programming language is C++.

Table 11.3. Order of magnitude of computation times in s for an expected error of 1 dB in each time interval of the energy impulse response (energy multiplication method), $t_{max} = T$.

Computation time in s (Eq. (11.24))	Room 1 "living"	Room 2 "lecture"	Room 3 "concert"	Room 4 "church"
N	200	2,000	20,000	200,000
t_{calc}	0.03	0.5	10	180
t_{calc} (BSP)	0.02	0.2	2.7	29

Next, an integral parameter such as the sound level (strength G) is the goal of the simulation.[48] 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 sufficient for determining the total level (Eq. (11.39)). The number of rays is calculated according to Eq. (11.21).

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. (11.24))	Room 1 "living"	Room 2 "lecture"	Room 3 "concert"	Room 4 "church"
N	16	100	500	2,500
t_{calc}	2	15	150	1,300
t_{calc} (BSP)	1	7	40	200

Now, the sound level (strength G) is the goal of the simulation again, 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. (11.28). No time limit such as t_{max} applies since the rays are traced until their annihilation.

---

[47] inspired by the limit given by the jnd for a sound level of roughly 1 decibel.
[48] It is assumed that the specific impulse is not relevant in detail. Instead, the goal of the simulation is a quick estimate of the level, clarity, definition or other integral parameters.

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. (11.31))	Room 1 "living"	Room 2 "lecture"	Room 3 "concert"	Room 4 "church"
N	31	200	1,000	5,000
t_{calc}	0.06	0.6	6	54
t_{calc} (BSP)	0.04	0.3	1.7	8.6

Note that computation times for detailed analysis of the impulse responses are much greater than those required for integral results. In other words, integral results can be estimated very quickly. For auralization, without reference to the specific room acoustic field, therefore, it might be of interest to estimate the room acoustic parameters quickly and to adjust artificial room impulse processors of early reflections and exponential late reverberation with reference to the integral simulation results (see Sect. 15.2.3). Note that this type of quick parameter estimation is best done by stochastic ray tracing with the annihilation method, since it offers real-time capability (time limit about 50 ms; see Chap. 15), even for the large church (Room 4).

11.3   Image source model

On the basis of the image source principle (Sect. 4.3) and extension to 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 when $|R| = 1$. For absorbing walls, it is a good approximation, as long as the angle of sound incidence, ϑ_0, is small, far from grazing incidence (see Sect. 3.2).

11.3.1   Classical model

The reflection factor is assumed to be angle-independent, corresponding to a quasi-plane wave:[49]

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

(11.32)

---

[49] The plane-wave assumption is equivalent to the prerequisite of nongrazing sound incidence.
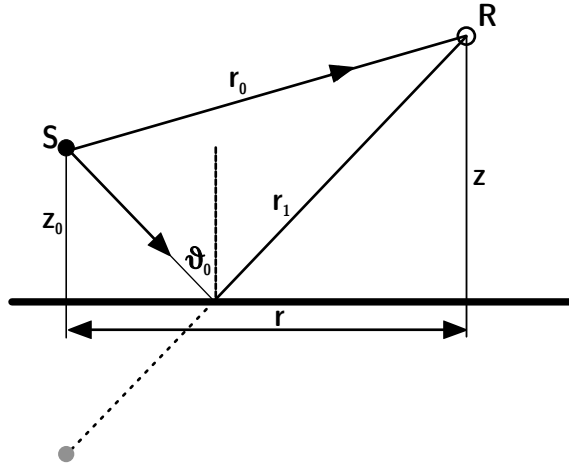
Fig. 11.19. Image source model

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 follows. For specific geometries (rectangular, triangular), it served as model for analyzing basic features of room impulse responses since the midddle 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.

If the room reflections are purely specular, the sound paths (rays) can be backtraced 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, again, mirrored at wall planes, to create image sources of higher order. All permutations of the walls must be considered, except a constellation in-volving 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 $\vec{n}$ and $\vec{r}$ yields the distance between the source and the wall, d

$$d = \vec{r} \cdot \vec{n} = |\vec{r}| \cos \alpha. \tag{11.33}$$

With this distance, we get the position of the image source,

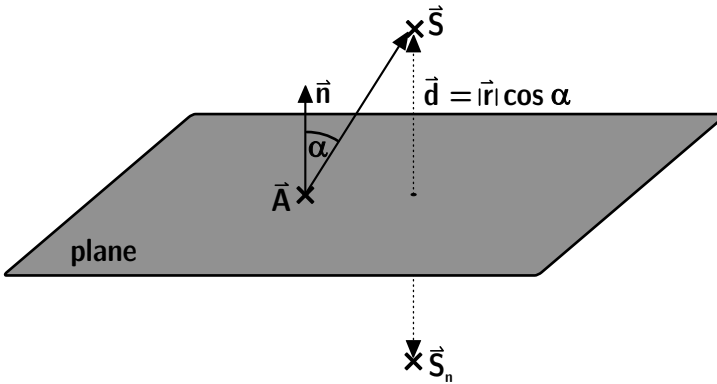$$\vec{S}_n = \vec{S} - 2d\vec{n}. \tag{11.34}$$

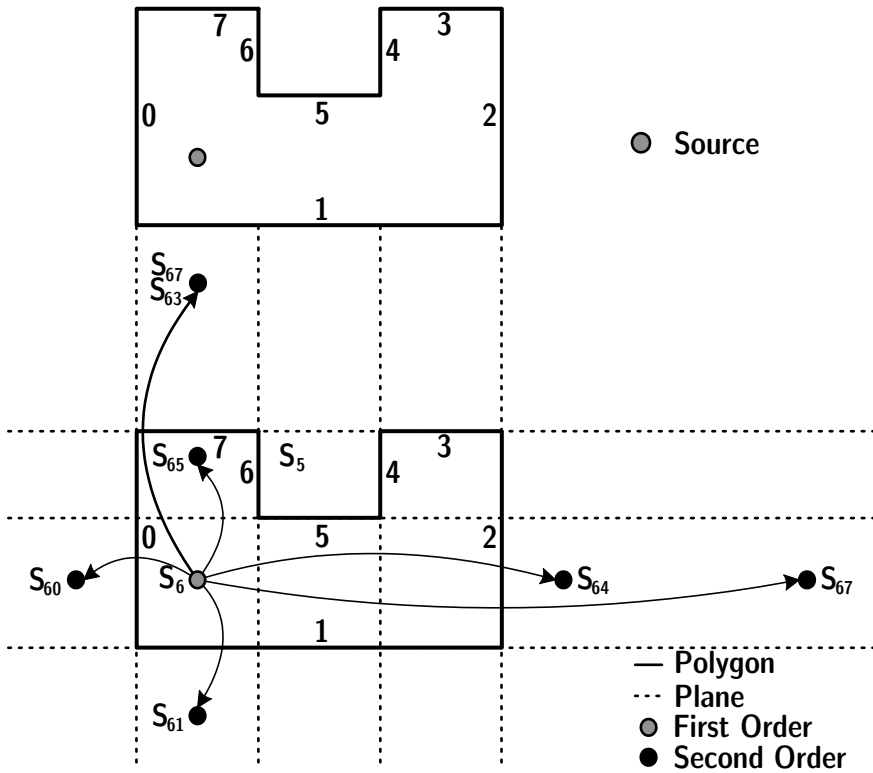**Fig. 11.20.** Construction of an image source

Fig. 11.21. Image sources constructed for a room (example in 2-D)

With this procedure applied to all walls we construct an image source of the first order. Image source of higher order are constructed in the same way by considering 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 nth order correspond to rays hitting n walls.

11.3.2 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 ith order is

$$S \rightarrow S_{n_1} \rightarrow S_{n_1 n_2} \rightarrow \ldots \rightarrow S_{n_1 n_2 \ldots n_{i-1}} \rightarrow S_{n_1 n_2 \ldots n_i} \,, \qquad (11.35)$$

with $n_k \neq n_{k\pm1}$ counting the walls hit, $n_k \in (1, n_w)$. The audibility test is started at the receiver point and the ray is backtraced from this point to the source along the chain of image sources, as shown in Fig. 11.22.
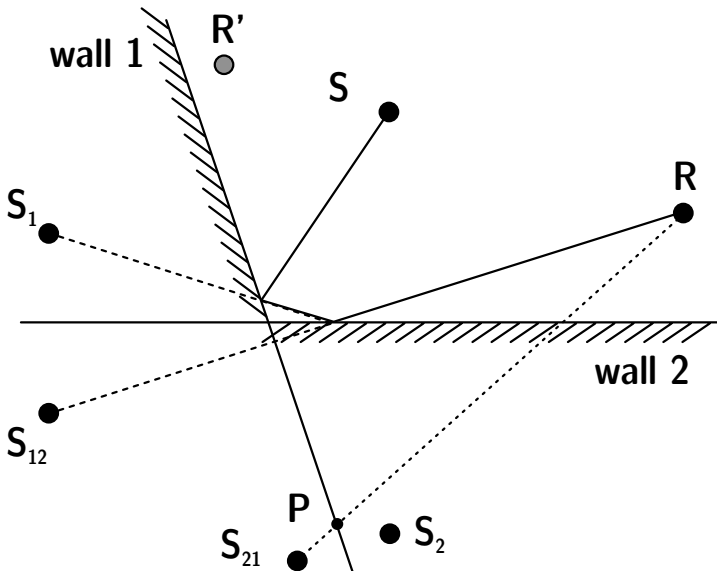

Fig. 11.22. Image source audibility test

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 hit. If the intersection of the straight line with the wall polygon 2, $\overline{RS_{12}}$, is inside the polygon, the result is preliminarily positive. We then continue by drawing a line from the intersection point to the mother source[50] of S_{12}, S_1. This procedure is repeated until the original source is reached. If all intersections hit inside the 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, such as in ray tracing. All strategies, therefore, to accelerate the backtracing procedure by spatial substructures may 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}.

$$p_{IS} = \frac{j\omega\rho_0\hat{Q}\Gamma e^{-j\omega t_{IS}}}{4\pi c t_{IS}} \prod_{n=1}^{i} \underline{R}_n \qquad (11.36)$$

with $t_{IS} = r_{IS}/c$ and $\underline{R}_n$ 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, thus, can be freely defined.

An estimate of the reflection factor from the absorption coefficient can be done straightforwardly by using

$$|R| = \sqrt{1 - \alpha} \qquad (11.37)$$

and an appropriate phase. Rindel (1993) and Mommertz (1996) show that absorption coefficients in one-third octave bands serve well for reconctruction of the complex reflection factor. Mimimum or linear phases are options for the phase function, to be reconstructed from the Hilbert transformation, for instance.

---

[50] A mother source is the source located in the chain as a predecessor.

11.3.3 Limitations

Truncation

The image source model is a strictly deterministic method. Uncertainties are caused by prerequisites of geometrical acoustics and by necessary truncation of the image source at 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 that in ray tracing. This is due to the dramatic increase of the number of sources to be treated with higher orders.[51] The number of audible sources, however, increases with t^3 in the impulse response (Cremer 1948).

At higher orders of image sources, the ratio between sources constructed and sources visible is uneconomic. With i denoting the maximum order chosen, the truncation time is

$$t_{max} = i / \overline{n} \,, \tag{11.38}$$

and the energy missing is

$$\frac{\Delta w_{trunc}}{\langle w \rangle} = (1 - \alpha)^i \,. \tag{11.39}$$

The computation time, t_{calc}, is similarly estimated from the time required for the point-in-polygon test:

$$t_{calc} = n_w \cdot N_{IS} \,. \tag{11.40}$$

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).[52]

Spherical waves at grazing incidence

Another limitation is given by the validity of Eq. (11.32). The reflection coefficient $\underline{R}(\vartheta_0) \neq 1$ in the contribution of the image source implies a constant

---

[51] Without strategies for excluding sources for geometric reasons, the increase is exponential.
[52] For a rectangular room, a regular lattice of image sources applies, with multiple sources (of indices in permutated order) coinciding at the lattice points. It can be shown, however, that for every receiver point, exactly one lattice source is audible.

angle of incidence and, thus, reflection of a plane wave at one reflection point. This assumption is well suited when a spherical wave[53] is considered with a large distance between the source and the wall and the receiver at the wall.

At smaller distances and corresponding grazing incidence, Eq. (11.32) carries a nondetermined 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

$$p = \frac{j\omega\rho_0 Q e^{-jkr_0}}{4\pi r_0} + p_{\text{IS}}, \tag{11.41}$$

whereas the standard model offers an estimate:

$$p_{\text{IS, plane wave}} = \frac{j\omega\rho_0 Q e^{-jkr_1}}{4\pi r_1} R(\vartheta_0), \tag{11.42}$$

where

$$R(\vartheta_0) = \frac{\zeta\cos\vartheta_0 - 1}{\zeta\cos\vartheta_0 + 1}. \tag{11.43}$$

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 the impedance and the reflection factor have been investigated both by experiment and by field calculation (Suh and Nelson 1999). As a rule of thumb, we can remember that at grazing incidence ($\vartheta_0 > 60°$) and at too close distances, d, of receiver and source to the wall ($d \le \lambda$), systematic uncertainties, Δp, of sound pressure must be considered which are clearly audible ($\Delta p_{\text{IS}} > 20\%$, which corresponds to 1 dB). In the middle of a room, errors are smaller than at positions near the room boundaries. Small and medium size rooms ($50\,\text{m}^3 < V < 200\,\text{m}^3$) at low frequencies ($f < 200\,\text{Hz}$), however, have hardly a centre area further than a wavelength from the boundaries. Here, the validity of standard (plane-wave impedance) geometrical acoustics is at its limits.

A more specific and precise uncertainty of the result in auralization 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. That walls

---

[53] The temporal coincidence of the wave-front phases hitting the surface is relevant here.

are not locally reacting at low frequencies and that the impedance may be distributed over the wall (by distributed screws in studs, for instance) create further complications.

11.3.4 Diffraction

Diffraction is neglected in the classical image source model (and also in ray tracing). 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 surroundings of finite room boundaries. In the latter case, either the boundary forms an obstacle, such as columns or the edge of an orchestra pit, or the boundary forms 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 in acoustic ray models. But the success was quite limited because the increase in calculation time is a severe problem.

In optics and radiowave physics, however, ray tracing models were generalized in the uniform geometric diffraction theory (UDT) (Kouyoumjian and Pathak 1974; Tsingos et al. 2001). Other approaches were presented by (Svensson et al. 1999), who applied the model by Biot and Tolstoy (1957) for acoustic problems. These are very powerful for determining first- or second-order diffraction. But all methods of geometric diffraction are very time-consuming for simulating a multiple-order diffraction and corresponding reverberation, and they introduce ray-splitting and an exponential factor in the multiple-scattering algorithm. Another possibility is to apply finite element or boundary element methods, of course.
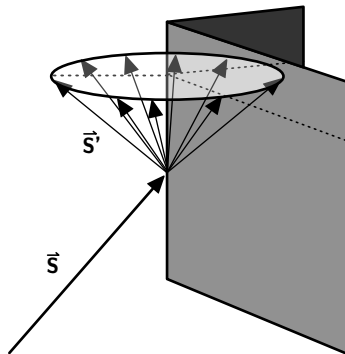

Fig. 11.23. Geometric edge diffraction using secondary sources (after (Svensson et al. 1999))
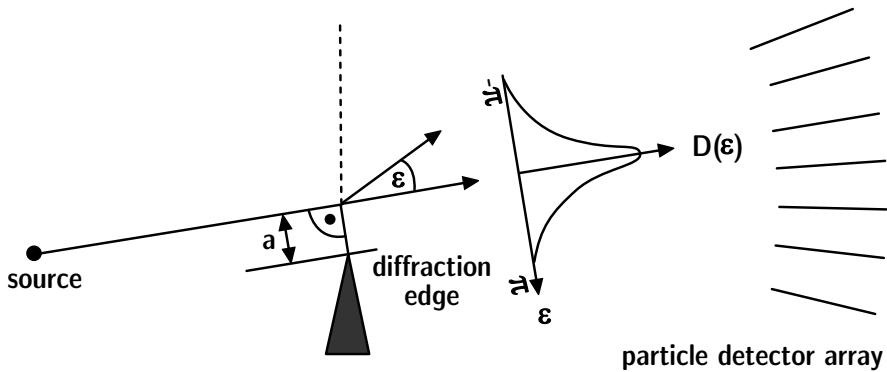
Fig. 11.24. Ray (particle) diffraction model after (Stephenson 2004b)

Stephenson (2004a, 2004b) developed two approaches for edge diffrac-
tio. One consists of the implementation of "virtual flags" mounted at the
edge, which let the edges affect rays in the room interior.

The flag width is about one wavelength representing the midband fre-
quency of the octave or one-third octave band. Rays hitting the flags can
be diffracted according to statistical distribution of diffraction angles de-
pending on the distance from the edge, the wavelength and the ray inci-
dence angle. This distribution is derived from either slit diffraction models
or Fresnel's edge diffraction models. Both methods show good results in
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 devel-
opments of geometric room acoustic simulation methods.

11.3.5   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, construc-
tion 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 used for cre-
ating the daughter source (see Sect. 11.3.1) is outside the field angle, the
mother source cannot give birth to this daughter. The example in Fig. 11.25
illustrates that the image source S_{70} (the daughter of source S_0 and grand-
daughter 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
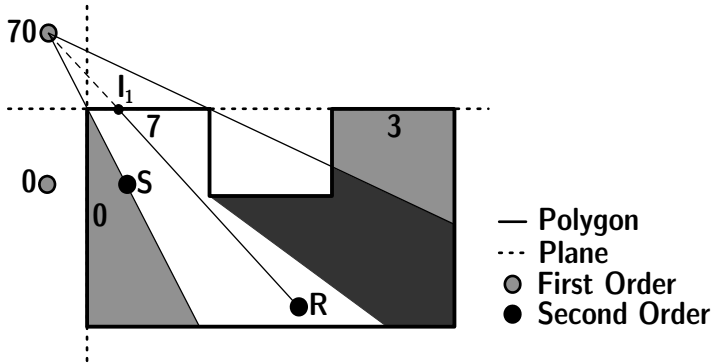
Fig. 11.25. Field angle of image souce S_{70} (reflection path with walls 0 and 7)

example, a convex corner 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 acoustical 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 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 separate the scenery into single objects enclosed by bounding volumes such as 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 arbitrarily 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 required intersection tests, a fast point-in-polygon test and an efficient data structure with respect to real-time auralization to provide a faster determination of the BRIR. This goal is illustrated further by discussing the example of BSP (Shumacker et al. 1969).
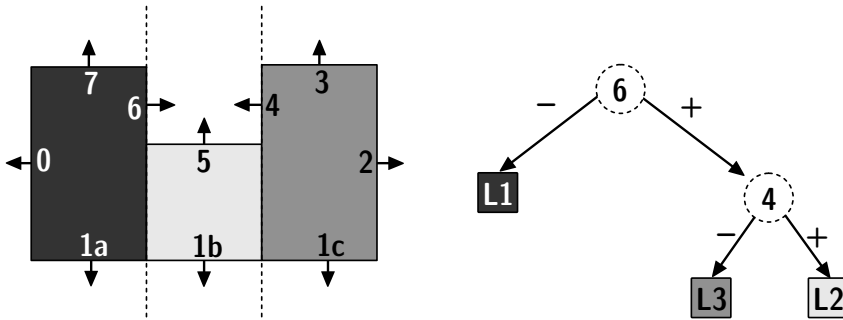
Fig. 11.26. Example of a BSP tree for a given room geometry (after (Shumacker et al. 1969; Schröder and Lentz 2006))

Binary Space Partitioning

The aim of BSP is speeding up the determination of the location of an arbitrary point relative to a geometric 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," 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 us to determine 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 minimised, which drops the complexity from $O(N)$ to $O(\log_2 N)$, where N is the number of polygons.

Voxels

Like octrees, voxels are volume boxes. They are created for subdividing the actual room volume, by using cubes, for example. In one pass of preprocessing, each voxel is tested for 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 when the complete set of polygons is not tested 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 that results in

$$t_{\text{calc}}\big|_{\text{voxel}} \propto \sqrt{n_w} \tag{11.44}$$

which was a previously linear increase without using voxels (Eq. (11.22)).

11.4 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 to obtain

– fine temporal resolution in sampling rate quality,
– inclusion of scattering, and
– faster audibility check of image sources.

This combination is called a "hybrid method." The term "ray tracing" also in use, however, is somewhat confused with the stochastic ray tracing described above. The key to combining 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. All have in common 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. (11.36). The main difference between this type of ray tracing and the previous stochastic ray tracing method is the way of detecting sound energy.

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 energetic overlap between rays and it must be ensured that each image-source 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 toward geometrically diverging objects such as cones or pyramids.

This way it is also ensured that image source contributions are counted only once. While cone or beam tracing still creates slight problems with overlap (Vian and van Maercke 1986), pyramid tracing using the wall
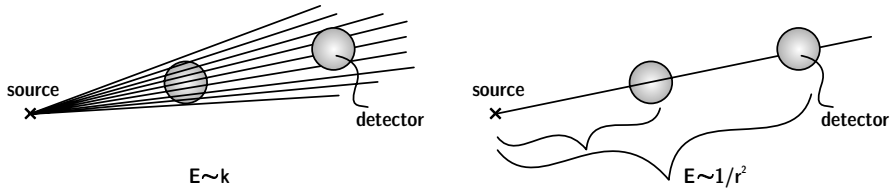
Fig. 11.27. Energy detection in stochastic (left) and deterministic (right) ray tracing
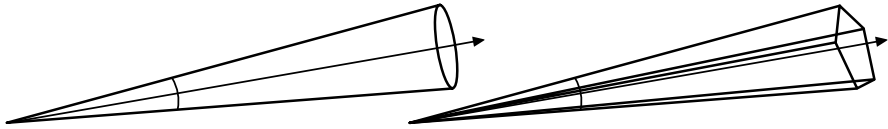

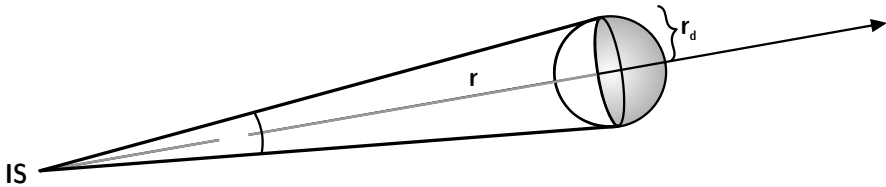**Fig. 11.28.** Cone (left) and pyramid (right) tracing

Fig. 11.29. Geometry of source, cone and receiver

polygons as pyramid base avoids double-counts (Stephenson 2004b). At the same time, with extended and diverging ray geometries, a point-like receiver can be used.

The required number of rays or cones depends on the maximum distance between image source and receiver and, thus, on the order of the image source and the corresponding average delay in the impulse response, as illustrated with the example of a cone.

The figure shows the path between image source and receiver in an expanded version.[54] N cones are started into the full space of 4π. From these,

$$k = N\Omega/4\pi \qquad (11.45)$$

will hit the receiver, with Ω denoting the spatial angle of the cone:

$$\Omega r^2 = \pi r_d^2 \qquad (11.46)$$

---

[54] The real path is folded from reflection to reflection. The total path length is identical in the expanded version.

for $r \gg r_d$. Hence, the probability of cones hitting the receiver is

$$k = \frac{Nr_d^2}{4r^2} = \frac{Nr_d^2}{4(ct)^2} .$$

(11.47)

To ensure that at least one ($k = 1$) cone hits the receiver after time t_{max},

$$N_{min} = \frac{4(ct_{max})^2}{r_d^2} .$$

(11.48)

The computation time of this kind of ray tracing (see also Eq. (11.24)),

$$t_{calc} = \frac{4c^2 \bar{n} t_{max}^3}{r_d^2} \tau ,$$

(11.49)

increases with the third power of the impulse response length.

In the last few decades, several authors have presented strategies to off-set the computational load of the image source method (Mechel 2002; Funkhouser et al. 1999; Stephenson 2004a). Other authors have 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 or extensions which do not consider wave theoretical extensions still suffer from limits set by the theory of specular reflections.

Before we focus on strategies for speeding up the simulation algorithms for auralization, the physical limits of room acoustics simulation by ray
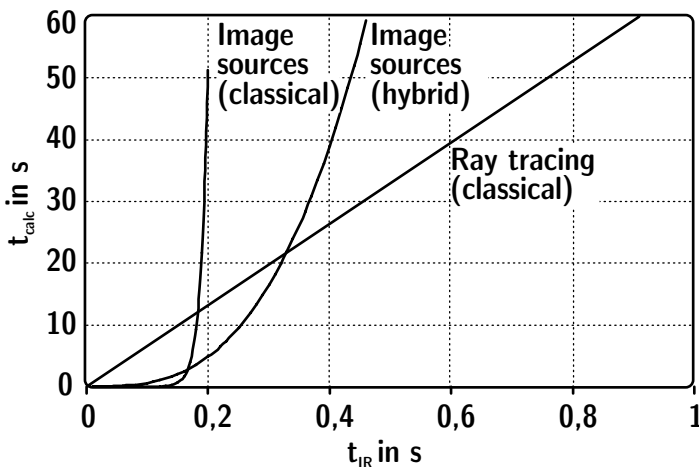

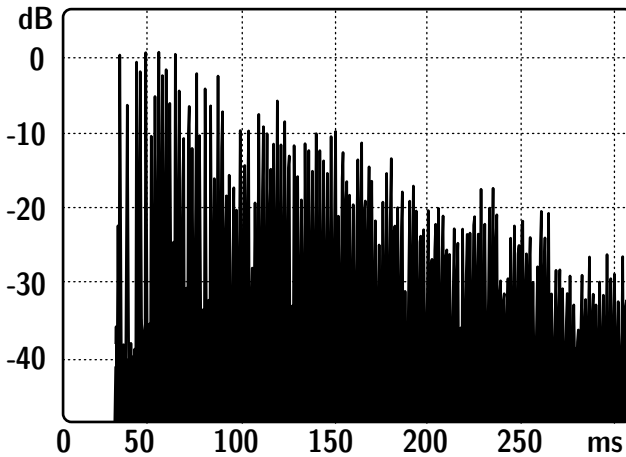**Fig. 11.30.** Computation times of room acoustic simulation algorithms

Fig. 11.31. Results with a hybrid image source algorithm

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 must also be considered.

11.5 Systematic uncertainties of geometrical acoustics

Geometrical acoustics (ray tracing, image sources) allows simulating sound fields in rooms and outdoors. These models yield correct results under specific circumstances (mostly academic cases) and fulfil the wave equation. But absolute physical correctness is not always a reasonable goal. In most cases, geometrical acoustics offers approximate solutions, solutions which are sufficient in practice. But it should be kept in mind that the limits are

- large rooms, nongrazing incidence,[55]
- low absorption coefficients, and
- broadband signals.

Room acoustical results such as 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, such as speech and music. Also from the physiological point of view, one-third octave bands are a good compromise since they are an approximation for critical bands (Sect. 6.2.1). These arguments
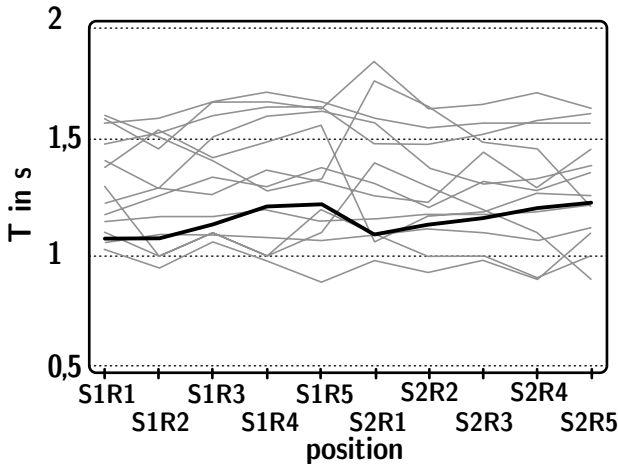
---

[55] or tangential modes

Fig. 11.32. Round-robin I, PTB auditorium (Vorländer 1995). Reverberation times calculated and measured (thick line) for source/receiver combinations

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 to find optimum 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.[56] At first, the results were compared with measurement results in only one frequency band.

The simulated reverberation times 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 software was run again with harmonized input data. The agreement was much better, of course, but now the efficiency of the software could 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 10 source and receiver combinations. The average of these location-dependent differences is then related to the jnd

---

[56] Auditorium of the Physikalisch-Technische Bundesanstalt, Braunschweig, Germany.
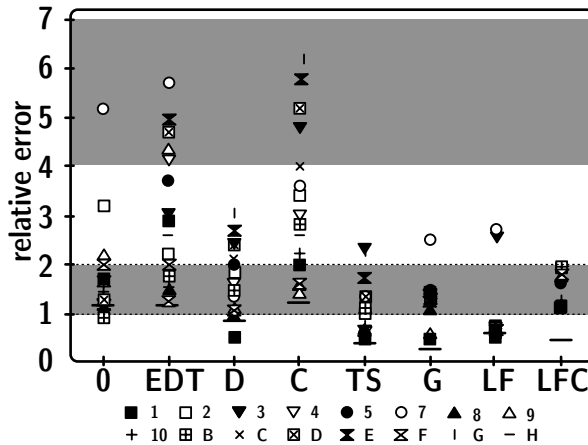
Fig. 11.33. Overall results of the "Round-Robin I" (Vorländer 1995). The ordinate scale refers to a normalized deviation from the jnd[57]

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 within similar accuracy as given by the standard deviation of the measurements and by the just noticeable differences (Table 6.1).

In the second intercomparison project, Bork (2000) 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, that simulation algorithms require a module for treatment of scattering. Also, in the second project, one CAD room model was provided 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. Programmes are available for achieving this accuracy, so that a certain quality of he results can be guaranteed. The operator of the software, however, still has crucial influence on the choice of absorption and scattering coefficients.

---

[57] At that time, the jnd for clarity was assumed to be around 0.5 dB. Later, it was shown that 1 dB is a better approximation; see also Table 6.1.

After all, it has been proven that room acoustical computer simulation yields 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 experimental results by the same order of magnitude as the uncertainty of the experiment or the jnd.

11.6   Hybrid models in room acoustics

Deterministic models such as the image source model suffer from inherent systematic errors and from limitations in 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 a certain temporal resolution. However, in the intercomparions of simulation programs (see Sect. 11.5), it soon became clear that pure image source modelling would 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 dominates the reverberation process after a few reflections (order of 3 or 4), even in rooms with rather smooth surfaces (see Fig. 11.34 and (Kuttruff 1995)). Also the audibility of the 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 to be Lambert scattering. Solutions to the problem of surface scattering are given by either
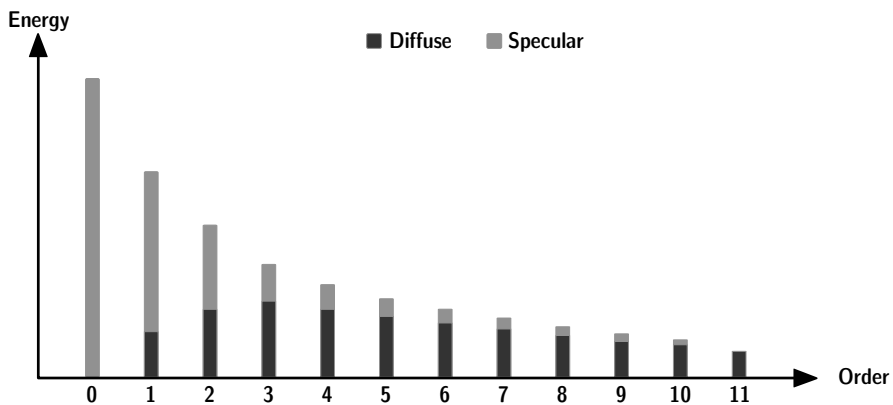

Fig. 11.34. Conversion of specularly into diffusely reflected sound energy, illustrated by an example (after (Kuttruff 1995))

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). In applications other than outdoor noise propagation,[58] diffraction models have not yet been implemented. New algorithmic schemes, such as presented by Stephenson (2004b), have not yet been implemented, too. It should be mentioned here that numerous algorithmic details have been published in the field of wave field simulation, until today. Sound transmission and diffraction, too, must be implemented in cases of coupled rooms or in room-to-room transmission; see Chap. 12.

11.6.1   Hybrid deterministic-stochastic models

The general principles of room acoustic modelling were discussed in the previous sections. They yield results of high temporal resolution (on scales of µs), typically represented by reflections of directly used or implicitly involved image sources. Due to numerical constraints and also because it is unnecessary to create such a high resolution for the late reverberation (> 100 ms), stochastic methods such as ray tracing and radiosity move into the focus of our interest. It is obvious to combine the tedious but exact deterministic methods with a quick estimate of late reverberation.

In the following, some well-known software solutions are briefly discussed. This overview is not intended to cover all kinds of software used today. The examples are particularly listed since they 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)). Implementation of the software already began in the late 1980s and has been continuously improved and extended. The main model consists of 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 using deterministic methods and image source sources and are thus created 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

---

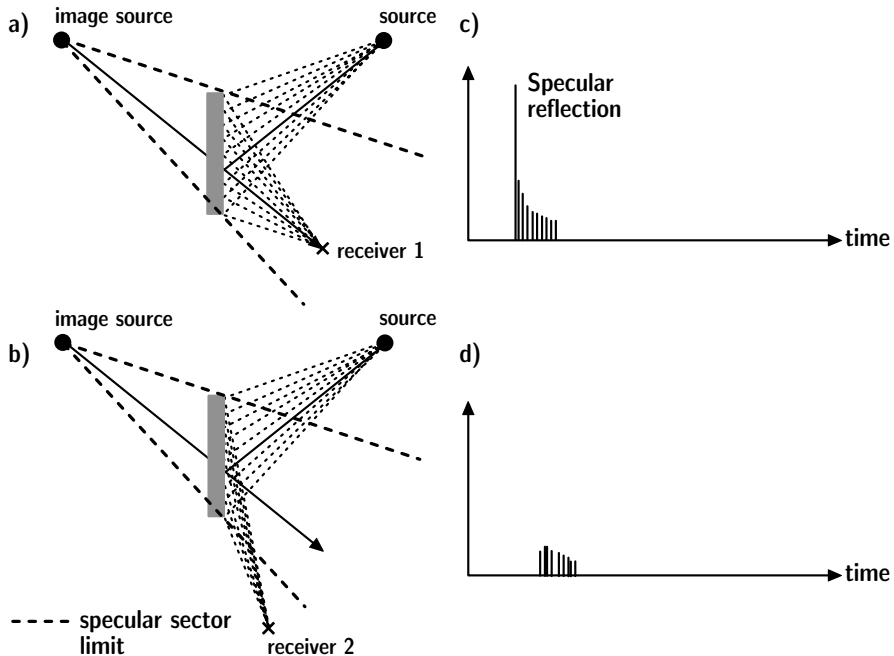[58] almost free-field propagation with at most one diffraction edge.

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)

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 in 1995.[59]

ODEON

The first version of ODEON was published in Naylor (1993). It is a hybrid image source model with a stochastic scattering process using secondary sources. The 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) into one final scattering direction for each ray (Rindel and Christensen 2003).[60] The process of creation of secondary sources is shown in Fig. 11.36 (top) for

---

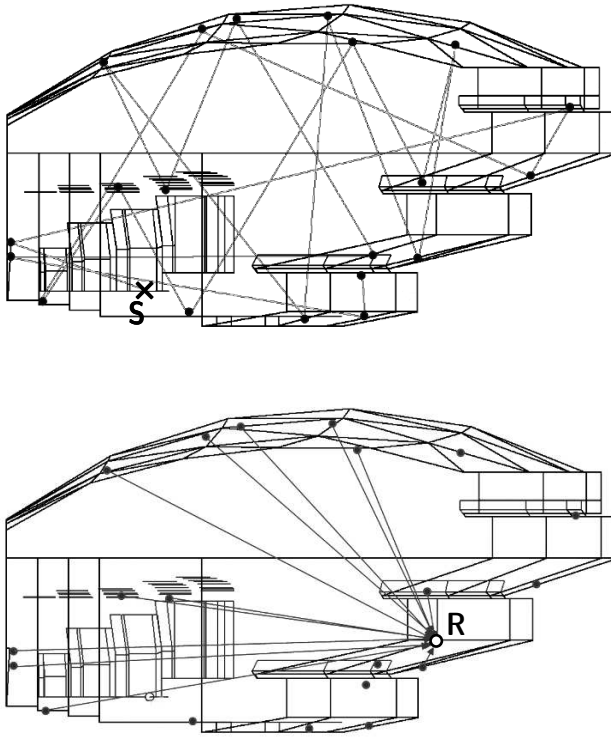[59] http://www.catt.se
[60] http://www.odeon.dk

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

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, 2004). The geometric principle of the method is pyramidal beam tracing based on the wall polygons that form the pyramidal base area. Successive reflections are created by extending and splitting the pyramids toward the other wall polygons. The problem of exponential increase of pyramids is inherent in the geometric imaging method. In QPBT, this is solved by unifying spatially closely located pyramids. Furthermore, by
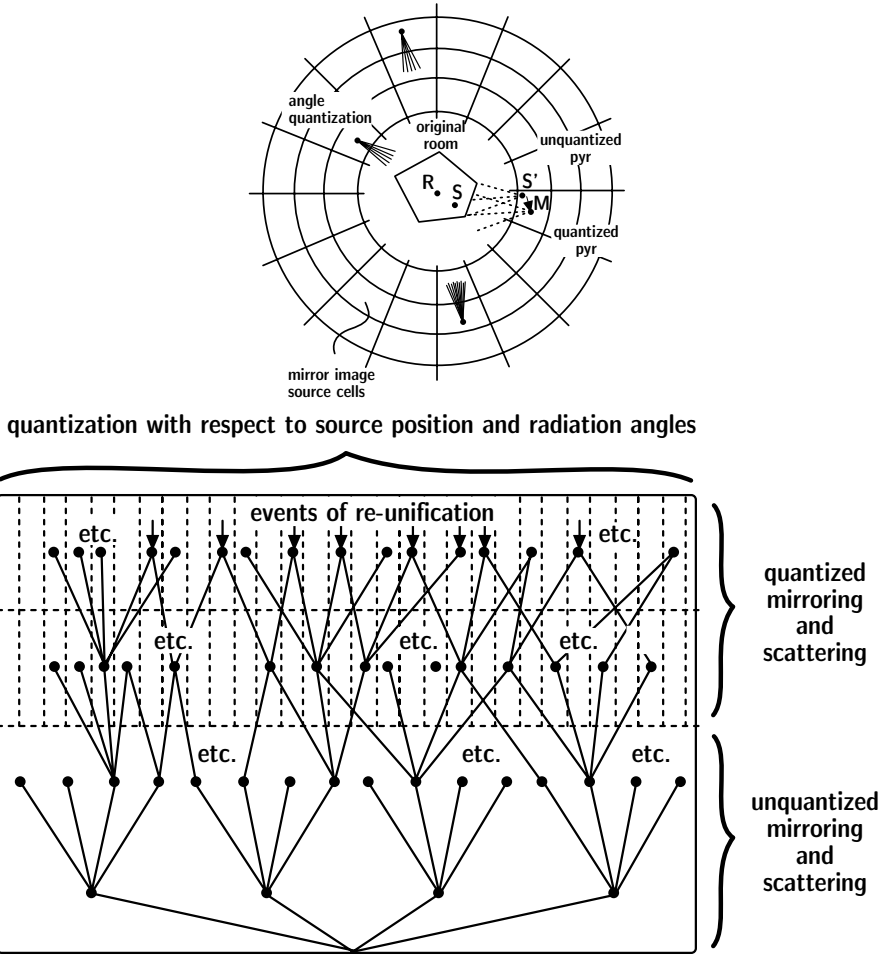
Fig. 11.37. Quantized pyramidal beam tracing and corresponding pyramid tree (after (Stephenson 1996))

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 was actually tested in Stephenson and Svensson (2007).

EASE

The software EASE stems from the application of simulating installations of professional audio and sound reinforcement systems (Ahnert and Feistel
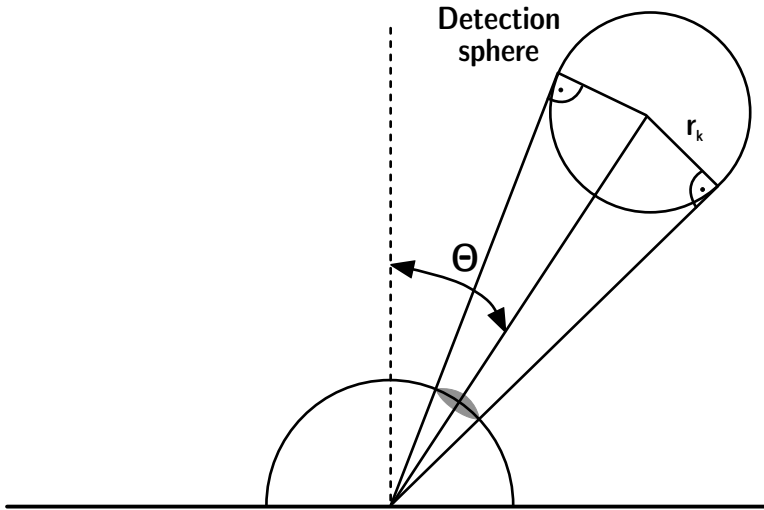
Fig. 11.38. Diffuse rain implemented for calculating the portion of energy scattered to a detector sphere (see (Heinz 1993) or the recently extended version in (Schröder et al. 2007); see also Fig. 15.5)

1991).[61] 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 implemented at first in an image source model capable of simulating particularly the direct sound and the early response.

In a later version of EASE, the room acoustic simulation was extended by integrating elements from a hybrid image source/ray tracing model called "CAESAR" (Vorländer 1989; Schmitz et al. 2001). The key to the latter method is parallel operation of image source and ray tracing algorithms. This delivers the specular reflections precisely and the diffuse reflections via the stochastic approach ("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.

State of the art

Still today, the programmes listed above are subject to permanent development of the components of impulse response computation and other elements such as source and material databases. Most of them also offer

---

[61] http://www.ada-acousticdesign.de/set_en/setsoft.html

signal processing tools, visualization tools and, auralization tools in various reproduction formats.

More recently, algorithms of geometric room acoustics modelling were also published by (Farina 1995; Alarcão and Bento Coelho 2003; Lokki 2002; Camilo et al. 2002), besides others. Many of these programmes were inspired by (Kulowski 1985; Stephenson 1985; Vian and van Maercke 1986; Vorländer 1989; Lewers 1993). In some methods. building room impulse responses from reflection statistics was tried (Vorländer 2000; Bento Coelho et al. 2001). Due to the rapid development in acoustic room simulation also, this list cannot be complete.

Now, the basics of hybrid room acoustic modelling techniques are known, so that we can discuss the next step, creation of impulse responses for auralization. This is reasonable, including binaural technology.

11.7 Construction of binaural 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. 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 must be based on binaural hearing. Otherwise, the spatial information would be lost. By using the HRTF (see Sect. 6.3.1), the directional information of a human listener is taken into account.

During the historical development of auralization of rooms, the first studies were based on purely specular reflections. After the principles were 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.[62] The model was further extended by (Lehnert and Blauert
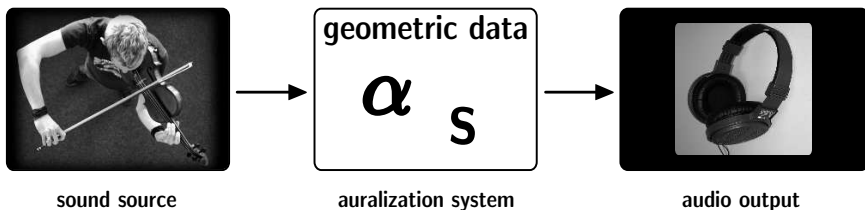

Fig. 11.39. Auralization of sound in rooms

---

[62] The reason for this was clearly the limitation by processing time in the late 1980s.

1989, 1992; Lehnert 1992; 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 binaural room impulse response based on spatial direct sound and reflections. With the deeper understanding of the details of room impulse responses and their relevance for auralization available, extended model approaches, including hybrid models and scattering models in particular, have been developed.

Finally, a binaural impulse response is created from direct sound, early reflections and scattered components by using the concept of binaural synthesis (Sect. 9.3). All components are added. When formulating in the frequency domain,[63] this process is described by multiplication of the transfer and filter functions representing sound travelling from the source to the receiver:

$$\underline{H}_j\Big|_{\text{left,right}} = \frac{\mathrm{e}^{-j\omega t_j}}{c t_j} \cdot \underline{H}_{\text{source}}(\theta,\phi) \cdot \underline{H}_{\text{air}} \cdot HRTF(\vartheta,\varphi)\Big|_{\text{left,right}} \cdot \prod_{i=1}^{n_j} \underline{R}_i \quad (11.50)$$

where H_j denotes the spectrum of the jth 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 sound incidence in listener coordinates at a specified orientation.[64]

Since data for air attenuation and particularly absorption coefficients of walls are available in frequency bands, the spectral representation required for Eq. (11.50) must be created by interpolation.

Interpolation

Several interpolation algorithms are possible candidates. The cubic spline interpolation is one example of a well-qualified method to create interpolated data. For interpolation of reflection factors from octave band data (see also Sect. 11.1.2), the effects of interpolation in the frequency and time domains must be discussed. Too rough interpolation may cause audible ringing. Onset and decay of the interpolation filter creates a kind of tonal reverberation.

---

[63] In the time domain, the same can be formulated, of course, by convolution, but the spectral product terms are easier to interpret.

[64] Usually the listeners are oriented toward the original source or a similar reference point.
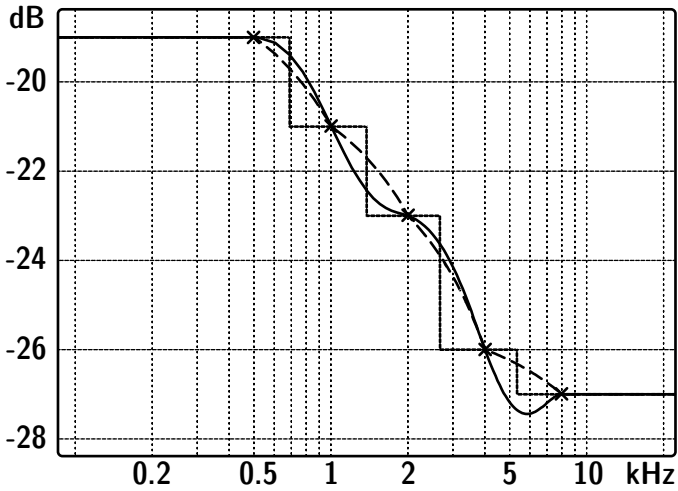
Fig. 11.40. Step (dotted line), linear (broken line) and spline interpolation of spectra based on octave band data (example)

Fig. 11.41. 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)

Mommertz (1996) showed that absorption coefficients in one-third octave bands serve well for reconstructing the complex reflection factor. 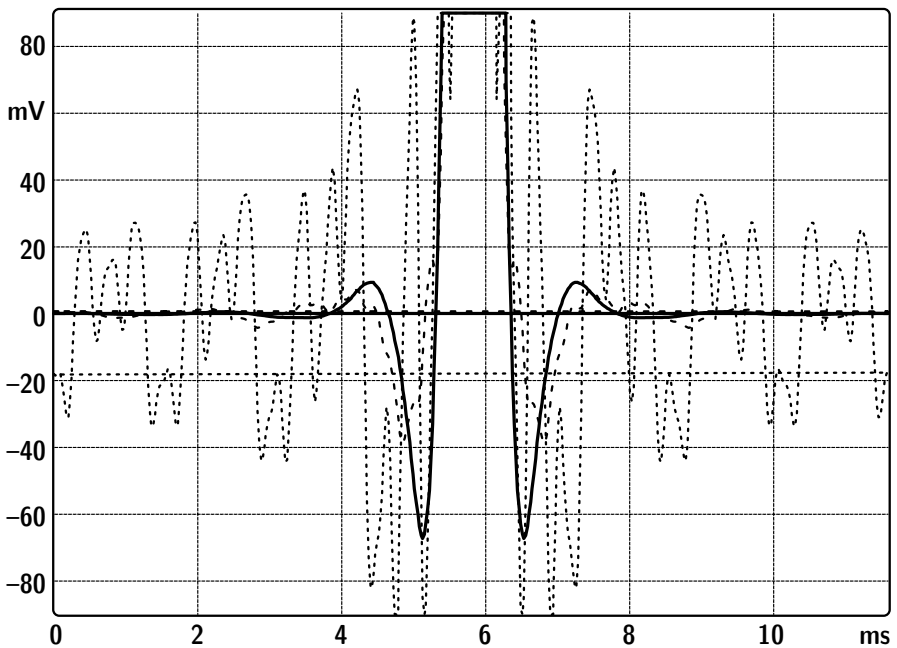While the modulus of the reflection factor can be obtained directly, a plausible phase must be added. Minimum or linear phases are options. In only very few cases such as for focuses, the specific choice will be important.

After inverse Fourier transformation, a spectrum consisting of M complex coefficients corresponds to a signal of M samples. Convolution of signal and transfer function of M lines each in the frequency domain creates a result of M lines, too. In contrast, in the time domain convolution, the result contains $2M$ samples.

Late reverberation

As explained above, scattering plays an important role particularly in the late response. Apparently, Eq. (11.50) deals only with image sources. But it can be generalized easily, if the contributions of the scattered and late part of the impulse response are represented by a set of equivalent reflections $\underline{H}_j$, whose arrangement must be constructed. 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[65]. By adding an adequate fine structure which represents the actual reflections statistics, the binaural impulse responses can be created (Heinz 1993).

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

$$M^2(\Delta t) = \left|\underline{H}_j\right|^2 \cdot n_j \qquad (11.51)$$

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 subjects 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.

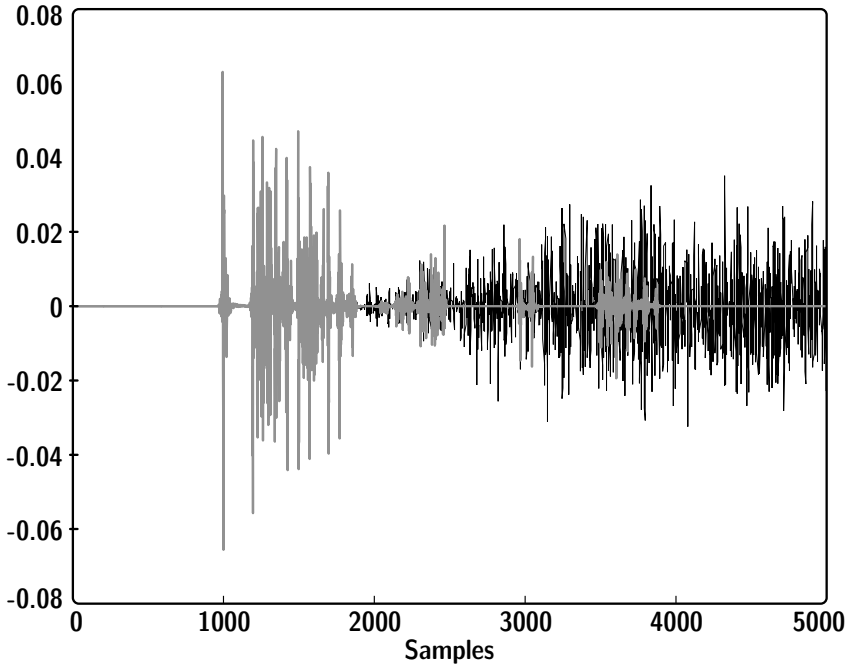[65] similar to a modulation function.

Fig. 11.42. One channel of a binaural room impulse response created from specular (grey) and diffuse (black) reflections (example after (Schröder et al. 2007))

Many more details of signal processing are found in the specific programmes listed in Sect. 11.6.1 above. For further information on these details, refer to the Internet links listed in Sect. 11.6.1.

12 Simulation and auralization of airborne sound insulation

After dealing with virtual environments in examples of sound in free fields and in enclosures (room acoustics, Sect. 11.6), the variety of virtual worlds is now extended to situations with partition walls and to complete buildings. This field is classical acoustic engineering in buildings as part of architectural acoustics.

Building acoustics is the discipline of sound and vibration transmission in buildings. Unlike 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 such as vehicle acoustics or noise control, as will be seen.

A typical example of building acoustics is the problem of sound transmitted into a so-called "receiving room" from the room next door (source room), such as sound from a stereo set into the neighbouring sleeping room. Another example is impact noise created by walking on the floor in the room above, a typical example of a structure-borne source. Furthermore, sound generated by building service equipment such as heating or air conditioning systems is of interest. Noise from outside the building, such as traffic noise or industrial noise, involves the sound insulation of facades.

In typical room-to-room situations, the signal received sounds quiet (fortunately) and dull. An auralization must reproduce these properties (level and colouration) in the first place.

> 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 here. We will therefore focus on the basic quantities used in building acoustics today. From 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. Standard building acoustics, thus, must be introduced in detail.

12.1   Definitions of airborne sound transmission

The basic quantity to describe the performance of construction elements[66] 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:

$$R = -10\log\tau = 10\log\frac{I_0}{I_t}, \tag{12.1}$$

where τ denotes the transmission coefficient, $\tau = I_t/I_0$.

In a free space with the partition separating two domains, R is identical to the sound pressure level difference, $D = L_S - L_R$. In buildings, however, the sound reduction index is related not to plane waves at one specific angle of incidence, but to diffuse and reverberant sound fields. From this fact, two consequences arise: a) the angle of incidence must be averaged and b) reverberation changes the level difference between the source and receiving room.

In the receiving room, the sound power, P, will create a diffuse sound field with energy density[67]

$$w_S = \frac{4P}{cA_S}, \tag{12.2}$$
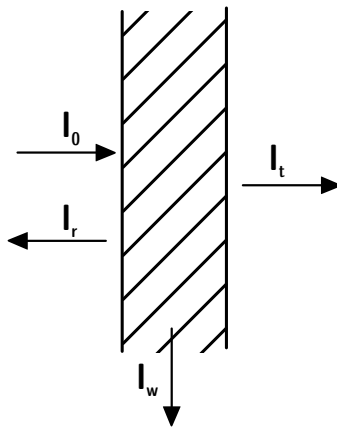

Fig. 12.1. Sound transmission through a wall

---

[66] Construction elements are typically walls, floors, doors, windows, etc.
[67] The direct field is neglected; thus we assume a distance larger than the reverberation distance.

where A_S denotes 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:

$$B = I_S = \frac{c}{4} w_S .$$
(12.3)

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

$$P_R = S \cdot I_R = \frac{c A_R w_R}{4} .$$
(12.4)

From Eqs. (12.3) and (12.4), we can derive the sound reduction index, R, for two adjacent rooms with diffuse sound fields:

$$R = -10 \log \tau = 10 \log \frac{I_S}{I_R} = 10 \log \frac{w_S}{w_R} + 10 \log \frac{S}{A_R}$$
$$= L_S - L_R + 10 \log \frac{S}{A_R} .$$
(12.5)

These definitions are most simple if the sound is transmitted via exactly one element, to which the transmission coefficient is addressed. In building practice, however, the situation is much more complicated. Sound and vibration and their coupling create energy transmission over multiple paths. Therefore, we must separate the problem into the elements involved. Building elements have finite size. They are coupled in s specific way, rigidly coupled or mounted with elastic connections. All these conditions have specific effects on the total sound insulation. Before we discuss the details of element coupling, we will study the basic effects of single partition sound transmission.

12.2   Sound insulation of building elements

Partition walls react to excitation by airborne sound waves. An adequate parameter to describe the amount of movement caused by sound pressure (or force) is wall impedance. Generally, the impedance may contain effects of inertia, of stiffness and of damping. Also, excitation and radiation depend on the size of the building element, as explained in Sect. 5.2.1.
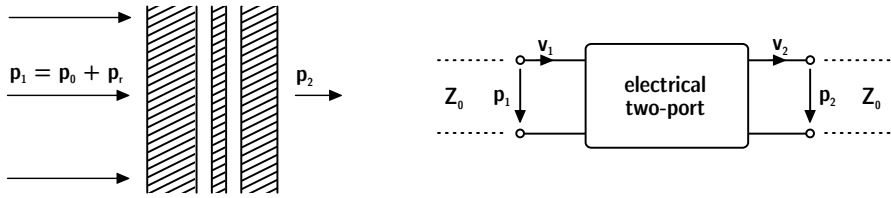
Fig. 12.2. Two-port model of partitions

A two-port model (Sect. 10.2) 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, $\underline{Z}_p$.

The ratio of the transmitted pressure side to the sound pressure of the incident wave is

$$H_{trans} = \frac{p_0}{p_2} = \frac{1}{2}\left[a_{11} + a_{22} - \rho_0 c\, a_{21} - \frac{a_{12}}{\rho_0 c}\right], \tag{12.6}$$

where a_{ij} denotes the matrix elements of the two-port (see Sect. 10.2),

$$\underline{a} = \begin{pmatrix} \dfrac{p_1}{p_2} & \dfrac{\det \underline{Z}}{p_2/v_1} \\[2mm] \dfrac{v_1}{p_2} & \dfrac{v_1}{v_2} \end{pmatrix}. \tag{12.7}$$

In this formulation, the sound reduction index of the partition is

$$R = 10\log\left|H_{trans}\right|^2. \tag{12.8}$$

Detailed information about the content of the two-port yields the matrix coefficients and the transmission function.

Single layers

With only an effect of inertia, the sound insulation of a single layer can be calculated easily. The impedance $\underline{Z}_p$ is given by $j\omega m''$, where m'' denotes the mass per unit area (see also Sect. 3.1.1). These results are, at least for buildings, a good estimate of the trend of insulation.

$$a_{11} = a_{22} = 1, \; a_{12} = -Z_p; \; a_{21} = 0, \tag{12.9}$$
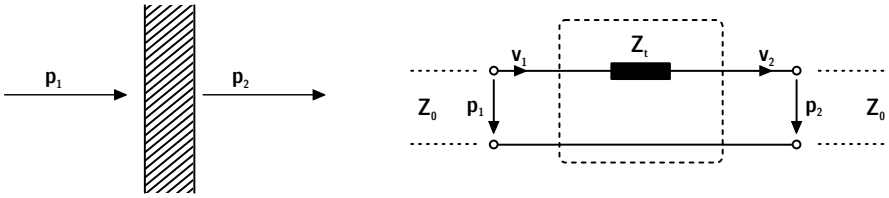
Fig. 12.3. Sound transmission through a partition with mass impedance

and according to Eq. (12.6)

$$R = 10 \log \left| 1 + \frac{Z_p}{2\rho_0 c} \right|^2 \tag{12.10}$$

and

$$R = 10 \log \left[1 + \left(\frac{\omega m'' \cos \vartheta}{2\rho_0 c} \right)^2 \right] \approx 20 \log \left(\frac{\omega m'' \cos \vartheta}{2\rho_0 c} \right), \tag{12.11}$$

the latter an approximation for $\omega m'' \cos \vartheta \gg 2\rho_0 c$.

Fig. 12.4. Typical results of measured sound insulation of massive walls. Range of variation found in intercomparison tests (after (Meier 2000))

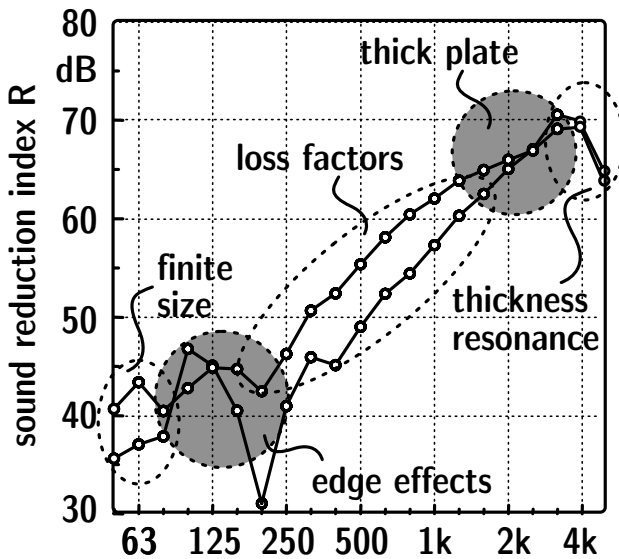Stiffness effects such as bending waves (Sect. 5.2), size effects (Sect. 5.2.1) or losses (Sect. 5.2.2) are accounted for by adding more detailed terms into Z_p:

$$Z_{\mathrm{p}} = \mathrm{j}\left(\omega m'' - \frac{B d \omega^3 \sin^4 \vartheta}{c^4} \right).$$ (12.12)

For an incident diffuse field, the effective sound reduction index is about 5 dB lower than the result of Eq. (12.11). The resulting sound reduction index is

$$R = 20\log\left(\frac{\omega m''}{2\rho_0 c} \right) - 10\log\left(\frac{1}{2\eta_{tot}}\sqrt{\frac{f_c}{f}} \right).$$ (12.13)

Double layers

The principle of the double wall is well known in lightweight constructions and for glazing. In many applications outside building acoustics, too, installation of a double or multiple wall system 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 to 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 a spring (Sect. 3.1.1, Eq. (3.13)).

For the frequency range above resonance ($\omega \gg \omega_0$), the movement in the system (vibrational 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 at rest. The sound reduction index can thus be approximated by

$$R = 10\log\left(1 + \left(\frac{\omega(m''_1 + m''_2)\cos\vartheta}{2Z_0}\left(1 - \frac{f^2}{f_{\mathrm{r}}^2} \right) \right)^2 \right).$$ (12.14)
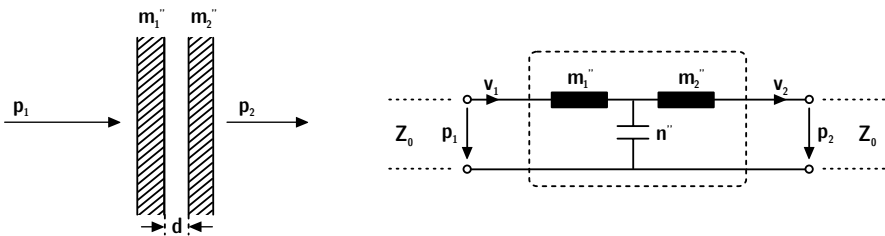

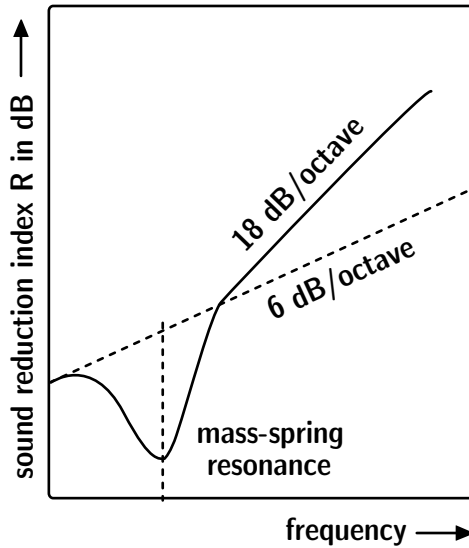**Fig. 12.5.** Sound transmission through a double wall

Fig. 12.6. Sound reduction index of a double wall

At frequencies less than 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 to 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

$$f_r = \frac{1}{2\pi} \sqrt{\frac{1}{M'' n''}} \, , \tag{12.15}$$

with the reduced mass M''

$$M'' = \frac{m''_1 m''_2}{m''_1 + m''_2} \, . \tag{12.16}$$

12.3   Sound insulation of buildings

The most important difference between academic studies of sound insulation and the effective sound insulation in a real building is given by 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 facades.

As in a strong flood of water, any dam will be as efficient as the bypass water flow is prohibited. The characterization of sound energy flow, therefore, is not sufficiently covered by the incident and transmitted sound waves for adjacent rooms separated by the partition. Also 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. Simulation and auralization in 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 in effective sound insulation, and this might even be increasing due to more low-frequency sound sources in residential buildings (TV and Hifi equipment).

By definition, the resulting sound reduction is denoted by R', the so-called "apparent sound reduction index":

$$R' = -10\log\tau' = L_S - L_R + 10\log\frac{S}{A_R}, \qquad (12.17)$$

where τ' denotes 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:

$$D_{nT} = L_S - L_R + 10\log\frac{T_R}{T_0}, \qquad (12.18)$$

which is called the standardized sound level difference. The difference between R' and D_{nT} is a constant geometric term ($\approx V_R/3S$) introducing the receiving room volume, V_R:

$$D_{nT} = R' + 10\log\frac{0.32V_R}{S}. \qquad (12.19)$$

Note that this equation allows predicting receiving room sound from source room sound and data from the transmission in the building structure. To achieve this form, we rewrite Eq. (12.19) as

$$L_R = L_S - D_{nT} + 10\log T_R + 3 \qquad (12.20)$$

with all L and D in decibels and T in s. It will be used later in Sect. 12.5.
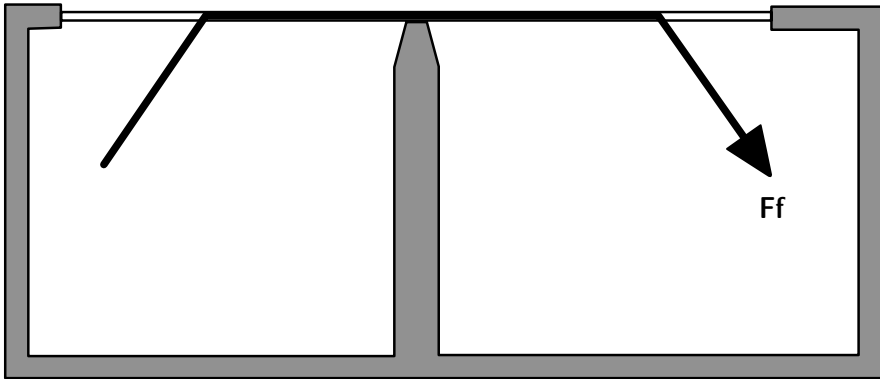
Fig. 12.7. Test facility for determining flanking transmission

12.3.1   Flanking transmission

The energy flow via flanking paths can be addressed similarly to direct sound transmission. The conditions for characterizing flanking paths are defined in measurement standards. Now, 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 for the specific energy flow considered, from the building element in the source room, i, to the element in the receiving room, j. Its definition, therefore, is equivalent to that of direct transmission:

$$R_{ij} = -10\log\tau_{ij} = 10\log\frac{P_i}{P_j},\qquad(12.21)$$

where P_i and P_j denote 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.

12.4   Sound transmission prediction models

To combine all relevant transmission paths, the specific transmission coefficients obtained in separated measurement conditions or calculated must be combined into a resulting sound transmission coefficient. This

calculation consists of the addition of energy transmission with weighing factors proportional to the area of sound irradiation and radiation in the receiving room. Thus, it can be considered an energy balance approach between subsystems. In a consequent and theoretically valid approach, the statistical energy analysis, (SEA) is a method of interest. It required a high modal density and is, therefore, applicable in many building situations with low critical frequency, to allow SEA application in a wide frequency range (see also Sect. 10.1.2). 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 the 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.

The total portion of the sound power transmitted is (see also Eq. (12.17))

$$\tau' = \sum_{i=1}^{N} \tau_i \tag{12.22}$$

with

$$\tau' = \tau_{\mathrm{d}} + \sum_{f=1}^{n} \tau_{\mathrm{f}} , \quad R' = -10 \log \tau' \tag{12.23}$$

and

$$\tau_{\mathrm{d}} = \tau_{\mathrm{Dd}} + \sum_{F=1}^{n} \tau_{\mathrm{Fd}}, \quad \tau_{\mathrm{f}} = \tau_{\mathrm{Df}} + \tau_{\mathrm{Ff}}, \quad \tau_{\mathrm{Dd}} = 10^{-R_{\mathrm{Dd}}/10} . \tag{12.24}$$

Generally, all specific transmission coefficients are related to their corresponding flanking sound reduction indices, R_{ij}.

$$\tau_{ij} = 10^{-R_{ij}/10} \tag{12.25}$$

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:

$$R_{ij} = -10 \log \tau_{ij} = \frac{R_i}{2} + \Delta R_i + \frac{R_j}{2} + \Delta R_j + K_{ij} + 10 \log \frac{S_S}{l_0 l_{ij}} . \tag{12.26}$$
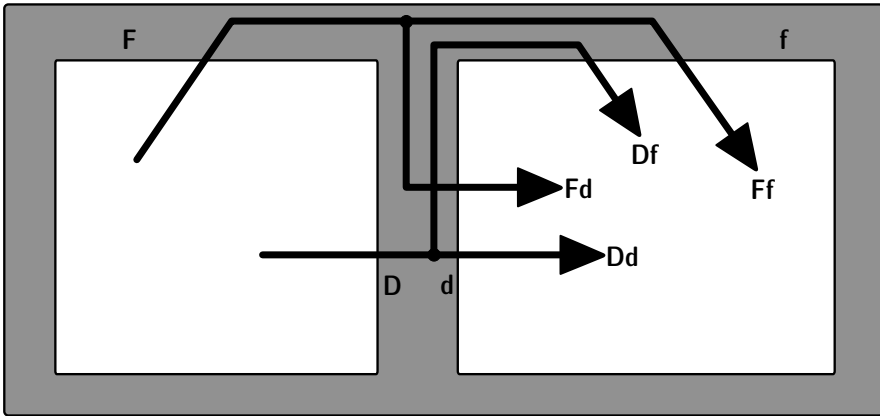
Fig. 12.8. Notation of flanking paths in sound transmission between rooms

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. (12.17)),

$$K_{ij} = \frac{D_{v,ij} + D_{v,ji}}{2} + 10\log\frac{l_{ij}}{\sqrt{a_i a_j}} \;. \tag{12.27}$$

The latter term represents the total losses in the plates (see also Sect. 5.2.2). a is the equivalent absorption length of the plates' perimeters and l_{ij} the length of the common junctions in decaying sound fields in rooms, the equivalent absorption length can be related to the decay time, here called 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

$$a = \frac{2.2\pi^2 S}{cT_s} \sqrt{\frac{1\text{kHz}}{f}} \;. \tag{12.28}$$

Finally, the structural reverberation and the loss factor can be separated into three parts, according to the basic information in Sect. 5.2.2. In 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 grounding electric current).

The final adjustment of absorption must be done 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.

12.5   Auralization of airborne sound insulation

Before discussing the procedure for creating an auralization filter, we separate the problem into the possible variants of source signal recording. The interface between the source signal and the filter must 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 a free field. In the latter case, the directivity in notation of a polar pattern or as coded data can also be taken into account.
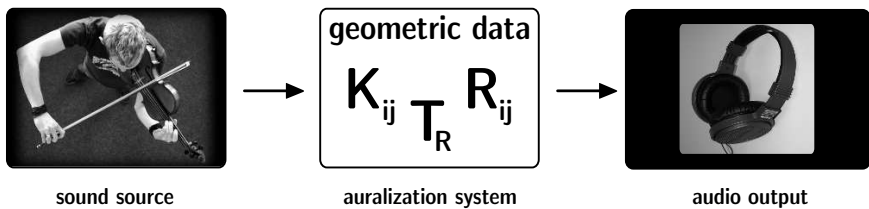

Fig 12.9. From input data to sound: Auralization of buildings

Source signal recording in the source room

Starting with the input data of sound insulation in frequency bands which stem from prediction or from 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. (12.22)), the sound level reduction for each path is obtained. Since the energy model is using 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

$$D_{nT} = L_S - L_R + 10\log\frac{T}{0.5\text{ s}}$$

$$= -10\log\tau' + 10\log\frac{0.32\,V}{S} = -10\log\tau_{nT}$$

(12.29)

where V denotes the receiving room volume in m^3 and S the wall surface of the partitions in m^2.

The clue for 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. (12.29) as

$$p_R^2 = p_S^2\,\frac{\tau_{nT}\,T}{0.5\text{ s}}$$

(12.30)

and

$$p_R(\omega) = p_S(\omega)\cdot F_{total}(\omega)$$

$$= p_S(\omega)\sum_{i=1}^{N} F_{\tau,i}(\omega)e^{-j\omega\Delta\tau_i}F_{rev,i}(\omega)$$

(12.31)

where $F_{\tau,i}$ denote 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. $F_{rev,i}$ 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
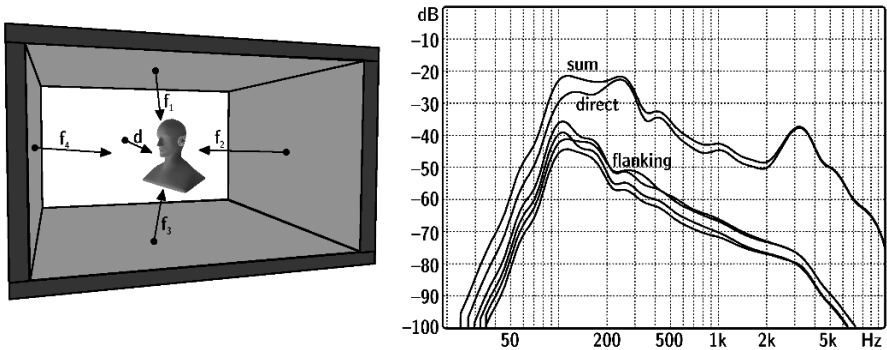

Fig. 12.10. Geometric situation and interpolated filters of the sound transmission paths involved
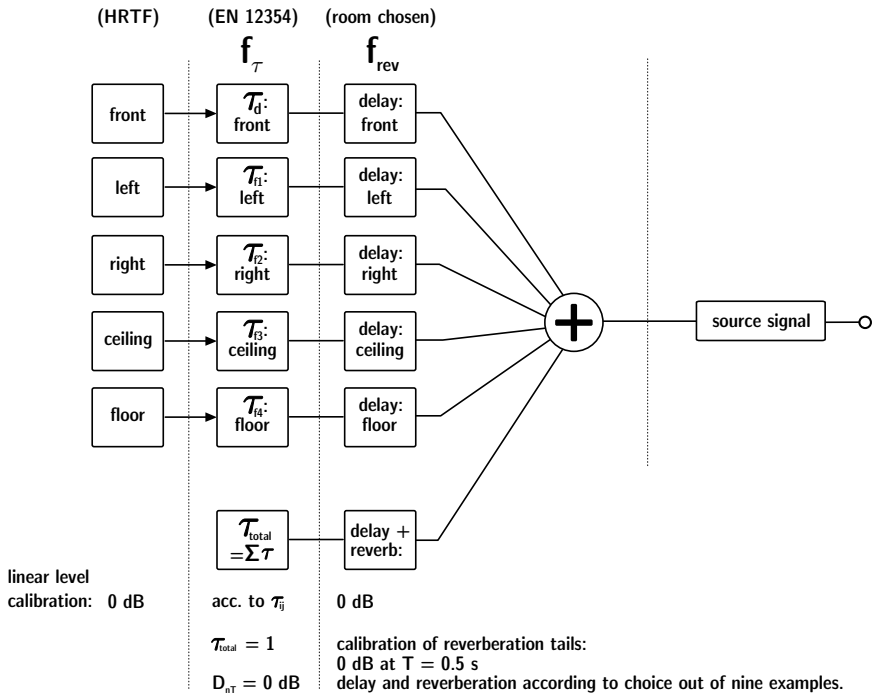
Fig. 12.11. Flow chart for auralization of airborne sound insulation

correct if the input sound pressure signal in the source room is calibrated with reference to $2 \cdot 0^{-5}$ Pa. Note that $p_S(\omega)$ is nothing but the complex spectrum of the recorded source time signal. Equation (12.31) can hence be expressed in the time domain by

$$p_R(t) = p_S(t) * f_{\text{total}}(t), \qquad (12.32)$$

where $f_{\text{total}}(t)$ denotes the transmission impulse response; see Fig. 12.12.

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.[68] This situation is acceptable since phases in reverberant sound fields cannot be recognised by human hearing. This does not apply, however, for 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 $\Delta \tau_i$.

---

[68] If the interpolation would be too rough, statistical energy analysis would not be an appropriate model. Thus, the model has its limits at low frequencies (at low modal densities).
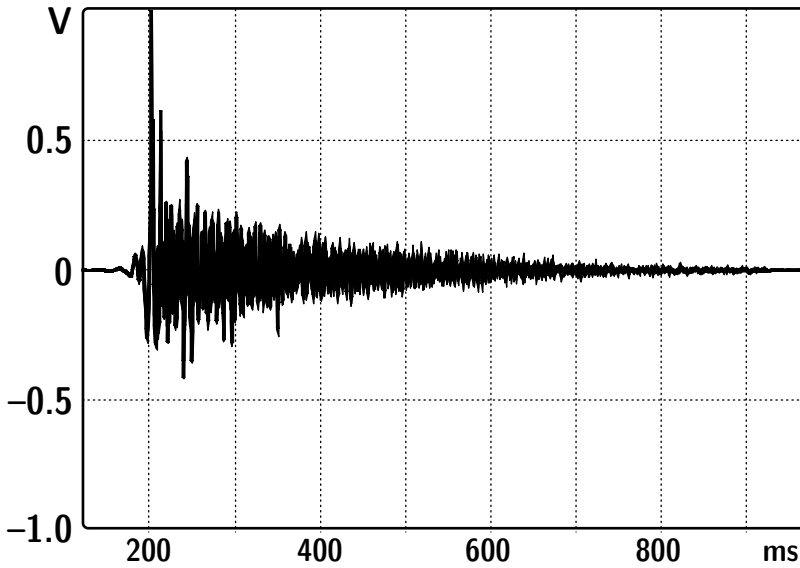
Fig. 12.12. Impulse response of the transmission source room–receiving room
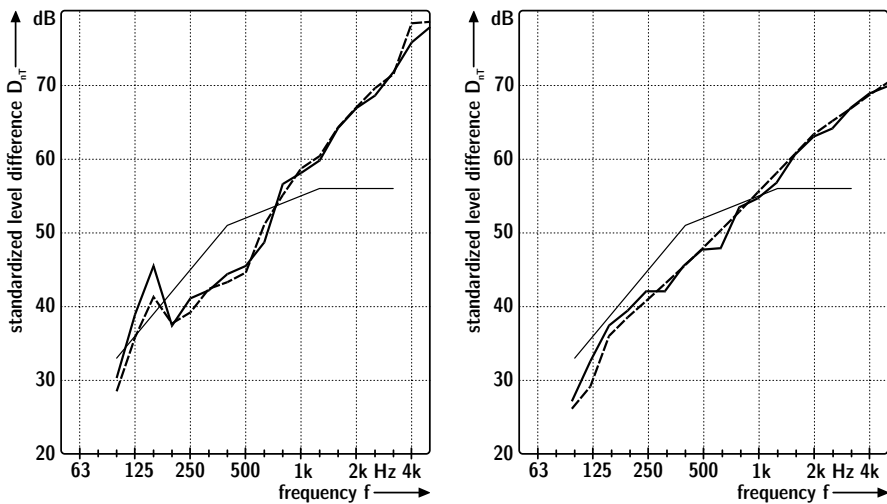
Fig. 12.13. Comparison of the input data of D_{nT} and D_{nT} measured from the auralized signals. The auralization stimuli were created by using pink noise and a measurement by feeding the signals to headphones on an artificial ear

The deviation of the auralized level differences from the measured level differences is mostly less than 1 dB. A deviation of more than 2 dB occurs only at low frequencies. These uncertainties are acceptable since they occur anyway in measurements at these low frequencies due to modal effects. For the auralization "measurement," only one point in the receiving

room was used. With appropriate spatial averaging, the agreement would be even better.

Source signal recording in free field

In 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))

$$D_{nT} = L_w - L_R + 10\log T_S + 10\log T_R - 10\log V_S + 17 , \qquad (12.33)$$

where $D_{nT} = -10\log\tau_{nT}$ denotes 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 rearranging to obtain the sound pressure in the receiving room, this reads

$$p_R^2 = \left(2.0\cdot 10^4\,\frac{\text{Pa}}{\text{s}}\right)\cdot P\cdot\frac{\tau_{nT}\,T_S T_R}{V_S} . \qquad (12.34)$$

As expected, the sound power of the source determines the squared sound pressure level in the receiving room. The factor $2\cdot 10^4$ Pa/s stems from combination of various constants such as the reference level for sound pressure ($p_0 = 2\cdot 10^{-5}$ Pa), the reference sound power level ($P_0 = 10^{-12}$ W) and from the reference reverberation time ($T_0 = 0.5$ s) in the definition of D_{nT}.

Equation (12.34), however, is not applicable straightforwardly to creating a source-filter model, since the source pressure is somehow hidden in the sound power. For source characterization, however, the far-field pressure in a well-defined direction and distance and the directivity pattern can be used (see Sect. 8.1). Omnidirectional sources and signals recorded in a free field at 1 m distance yield[69]

$$p_R^2(\omega) = \left(2.0\cdot 10^4\,\frac{\text{Pa}}{\text{s}}\right)\frac{4\pi}{\rho_0 c}\cdot p_S^2(\omega)\Bigg|_{1m} \cdot\frac{T_S T_R \tau_{nT}}{V_S} , \qquad (12.35)$$

---

[69] All quantities are to be expressed by using their numerical values in SI units.

which can further be expanded into

$$
\begin{aligned}
p_{\mathrm{R}}(\omega) &= p_{\mathrm{S}}(\omega)\big|_{1\mathrm{m}} \cdot 24.6 \cdot \sqrt{\frac{T_S T_{\mathrm{R}} \tau_{nT}}{V_S}} \\
&= p_{\mathrm{S}}(\omega)\big|_{1\mathrm{m}} \cdot F'_{\mathrm{total}}(\omega) \\
&= p_{\mathrm{S}}(\omega)\big|_{1\mathrm{m}} \cdot \frac{24.6}{\sqrt{V_S}} \cdot F_{\mathrm{rev},S}(\omega) \cdot \sum_{i=1}^{N} F_{\tau,i}(\omega) e^{-j\omega\Delta t_i} \cdot F_{\mathrm{rev},i,\mathrm{R}}(\omega)
\end{aligned}
\qquad (12.36)
$$

The filters to be created in this situation must be related to a free-field calibrated sound source, $p_{\mathrm{S}}(\omega)|_{1\mathrm{m}}$[70], to a reverberation filter, obtained by simulation or measurement in the source room, $F_{\mathrm{rev},S}$, reverberation determination by simulation or measurement in the receiving room, $F_{\mathrm{rev},i,\mathrm{R}}$, and to the standardized sound insulation quantity D_{nT}, as described above. The reverberation time filters must be calibrated with regard to a total energy of 1 when $T = 1\,\mathrm{s}$, respectively. As above, the same can be expressed in the time domain by

$$
p_R(t) = p_S(t)\big|_{1\mathrm{m}} * f'_{\mathrm{total}}(t) , \qquad (12.37)
$$

with $f'_{\mathrm{total}}(t)$ denoting the transmission impulse response.

In contrast to the example of source recording in the source room, a directional source must be treated in a specific way. The direction used in the recording situation must match the directional reference in the determination of the source room impulse response or the corresponding transfer function, $F_{\mathrm{rev},S}(\omega)$. This situation can be understood by considering the differences in the intensities radiated in various directions into the source room. This process feeds the source room reverberation with distributed sound with a total power of $P\,|\Gamma|^2$, which can also be introduced in Eq. (12.34).

The methods described in this chapter are explained with the example of airborne sound transmission in buildings. Nevertheless, they are applicable to 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. 14.

---

[70] Note the different reference for the recorded signal compared with Eq. (12.31).

13 Simulation and auralization of structure-borne sound

So far, the prediction and auralization was focused on airborne sound propagation in a free field (Sect. 9.3), in a room (Sect. 11.7), through a system of partition walls (Sect. 12.5). 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. With more thorough consideration of the source mechanism, it might be necessary to take the acoustic load on the source into account. This, however, hardly applies to the applications listed above.

The straightforward general extension of the propagation and auralization model is discussed now, and this is related to the excitation of solid structures by forces. The source, therefore, is to be characterized by its force output or by its velocity injected into the medium. With the velocity in the structure 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 predicting and evaluating impact sound are given in the next section.

13.1 Definitions of impact sound transmission

The characterization of sound insulation against the force impact of walking persons 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 a weight of 500 g each (see Sect. 8.2). They are lifted 4 cm and released to a freely falling condition. Their hammer speed at impact, hence, is fixed at 0.886 m/s. The total repetition rate is 10 Hz. In elastic impact, the impulse force will be transmitted completely into the contact surface, provided that the input impedance is very high compared with the impedance of the hammer. This

kind of force source represents an ideal case, obviously. The contact conditions are

– point contact and
– large contact impedance of the transmitting system.

These conditions are not valid for lightweight or resilient transmitting systems such as wooden floors or soft floor coverings. Nevertheless, we continue to discuss impact based on this assumption, before introducing in Sect. 13.4 some modifications that extend the validity to smalller 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 established today in international measurement standards and rating schemes (see also Sect. 8.2).

Due to direct excitation by a standardized force, the source components need not be determined. The result of impact sound insulation is, thus, given by the receiving room level, normalized to a reference condition of reverberation:

$$L_{\mathrm{n}} = L_{\mathrm{R}} + 10\log\frac{A}{A_0}, \quad A_0 = 10\mathrm{m}^2, \tag{13.1}$$

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 used:

$$L_{\mathrm{nT}} = L_{\mathrm{R}} - 10\log\frac{T}{T_0}, \quad T_0 = 0.5\mathrm{s}. \tag{13.2}$$

The above-mentioned problems of flanking transmission do not show up to 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 can be observed. A quantity denoting a transmission loss in terms of an intensity ratio, however, does not exist here.

13.2   Impact sound model

Compared with that described for airborne sound insulation, modelling and auralization of impact sound generated by walking on a floor is more difficult. The methodology listed here, therefore, is based on preliminary results of ongoing research. At first, it must be noted that all data of the

impact noise levels of floors are defined on the basis of the tapping machine. This prerequisite is not quite appropriate for lightweight floor constructions or coverings such as wooden floating floors (for dealing with these problems; see Sect. 13.4). Furthermore, nonlinearities have to be taken into account in case of timber floor constructions excited by low-frequency heavy impacts.

The estimation presented in this model is again following the approach of SEA, such as in airborne sound insulation. 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 is 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,

$$L'_n = 10\log\left(10^{L_{n,d}/10} + \sum_{j=1}^{N} 10^{L_{n,j}/10} \right). \tag{13.3}$$

Such 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).

$$L_{n,situ} = L_{n,lab} + 10\log\left(\frac{T_{s,situ}}{T_{s,lab}} \right), \tag{13.4}$$

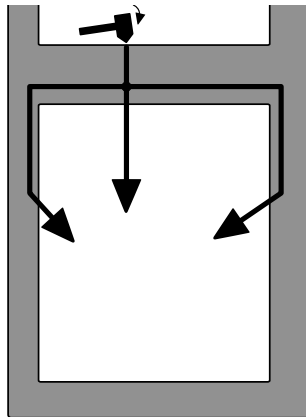with $L_{n,lab}$ denoting the reference laboratory condition.

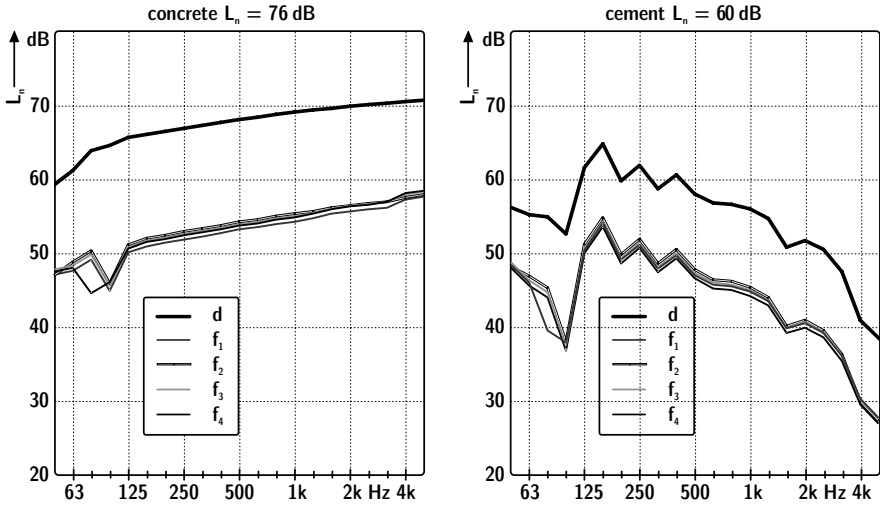**Fig. 13.1.** Sound transfer path from force source to receiving room

Fig. 13.2. 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.

$$L_{n,d} = L_{n,\mathrm{situ}} - \Delta L_{\mathrm{situ}} - \Delta L_{d,\,\mathrm{situ}} \, . \tag{13.5}$$

For the flanking paths,

$$L_{n,ij} = L_{n,\mathrm{situ}} - \Delta L_{n,\mathrm{situ}} + \frac{R_{i,\mathrm{situ}} - R_{j,\mathrm{situ}}}{2} - \Delta R_{j,\mathrm{situ}} - \overline{D_{v,ij,\mathrm{situ}}} - 10\log\sqrt{\frac{S_i}{S_j}} \, . \tag{13.6}$$

Equation (13.6) also includes influences of junction level differences in the actual field situation. Which component is most relevant, must be identified in each specific case. 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. 13.2 accordingly 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 when just a rough estimate of the floor construction is required, the normalised 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,

$$L_n \approx 155 - 30\log m'' + 10\log T_s + 10\log\sigma + 10\log\frac{f}{1\,\text{kHz}}.$$
(13.7)

13.3 Impact sound auralization

The interaction between the impact source and the floor construction will be discussed in Sect. 13.4. 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 L_n, for instance, we obtain

$$p_R^2 = \sum_i 10^{\frac{L_{n,i}}{10}} p_0^2 \frac{T}{0.5s} \frac{V}{30},$$
(13.8)

equivalent to the procedure discussed in Sect. 12.5.

If one attempts to model and auralize the noise of a person walking on the above floor on the basis of standardized impact sound levels, the tapping machine excitation must first be extracted from the data. This could 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

$$\begin{aligned}
p_R(\omega) &= F_{\text{walker}}(\omega)\frac{p_{\text{TM}}(\omega)}{F_{\text{TM}}(\omega)}\sum_{i=1}^{N}F_{\tau,i}(\omega)F_{\text{rev},i}(\omega) \\
&= F_{\text{walker}}(\omega)\cdot H_{Fp}(\omega)
\end{aligned}$$
(13.9)

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, F_{TM} the force spectrum of the tapping machine. $F_{\tau,i}$ and $F_{\text{rev},i}$ were defined above in Sect. 12.5.

As an example, the two floor constructions were auralized and analyzed regarding their sound pressure levels by (Thaden 2005). For this, 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. 13.2. The forces of the tapping machine and a rubber ball according to (Tachibana 1998) were measured.
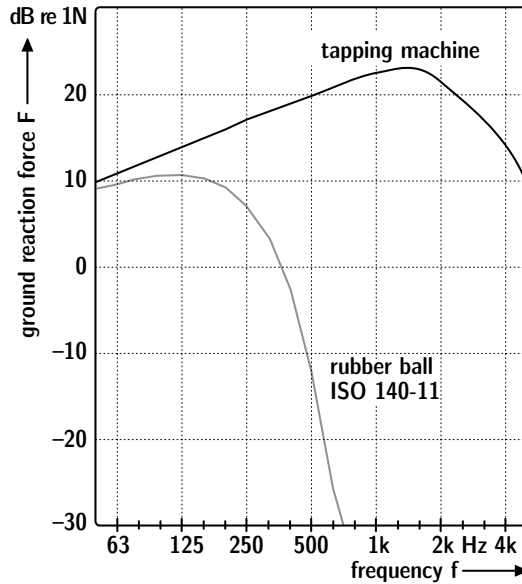

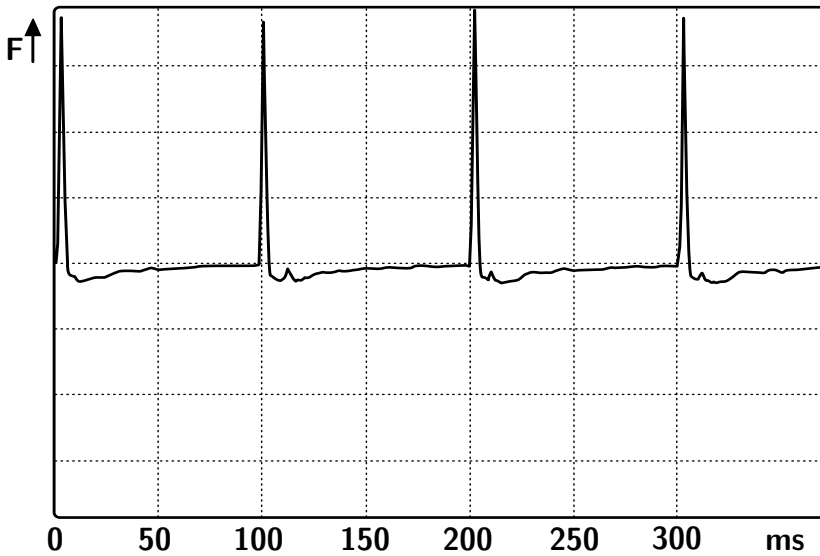**Fig. 13.3.** Force spectra of the tapping machine and rubber ball

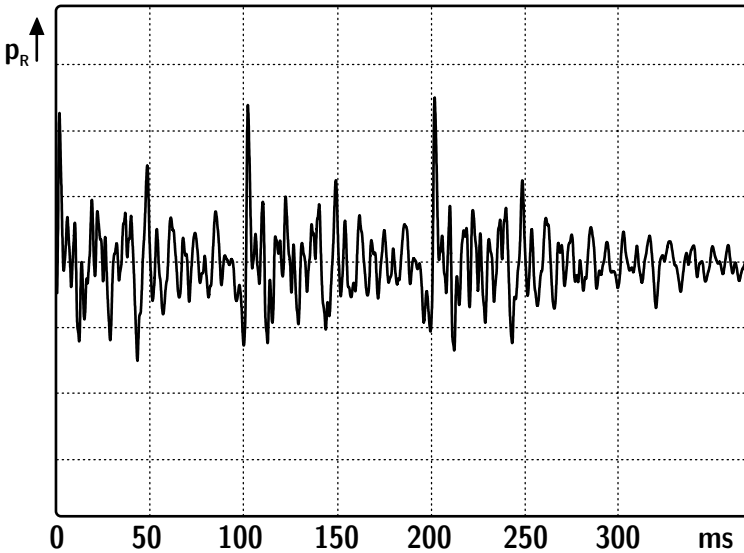**Fig. 13.4.** Force-time signal of the tapping machine

Fig. 13.5. Example of time signals of the sound pressure in the receiving room, with tapping machine excitation of a bare concrete floor

Force time signals of the tapping machine are constructed as follows. To obtain appropriate time signals, hammer force pulses are recorded and added at 10 Hz rate. Jitter in time and amplitude was introduced to get a more natural impression. A convolution of this primary force signal with the impulse response yields the sound pressure signal (see also Sect. 8.2).

Other force sources can be treated in the same way, based on recording of the force-time function.

13.4   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 depend on the floor mobility. This problem is difficult to solve in practice, even for only linear transmission.

The open-circuit (free) or blocked force and the inner impedances of the source and the floor must be determined. Impedances and forces can be obtained according to the method described in Sect. 8.2. 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
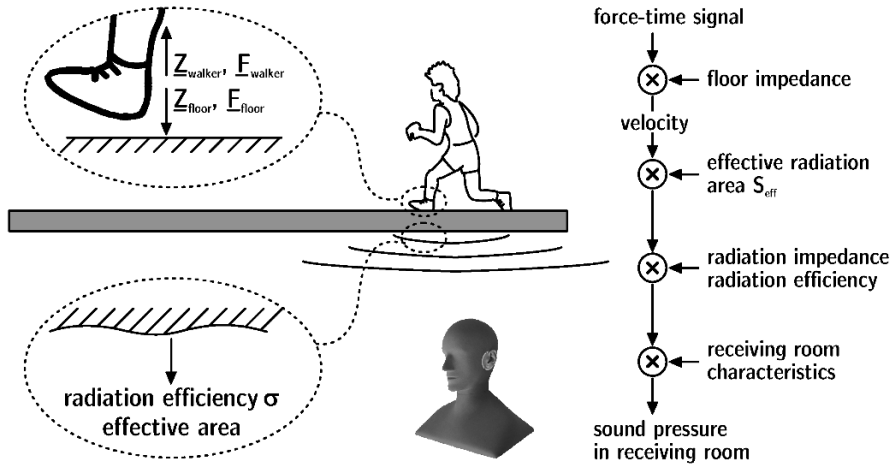
Fig. 13.6. Auralization model for impact noise

and the receiving point (receiving room) is again derived from the normalized or standardized impact sound level (according to Sect. 13.2).

$$p_R(\omega) = F_{0,\,\text{walker}}(\omega) \frac{p_{\text{TM}}(\omega)}{F_{0,\,\text{TM}}(\omega)} \cdot \frac{Z_{i,\,\text{TM}}(\omega) + Z_{a,\,\text{Floor}}(\omega)}{Z_{i,\,\text{Walker}}(\omega) + Z_{a,\,\text{Floor}}(\omega)} \sum_{i=1}^{N} F_{\tau,i}(\omega) F_{\text{rev},i}(\omega)$$

$$= F_{0,\,\text{walker}}(\omega) \cdot H_{\text{Filter}}(\omega)$$

(13.10)

where $F_{0,\text{walker}}$ denotes the blocked-force spectrum of the footstep source, the rubber ball of the tapping machine; p_{TM} is the sound pressure spectrum of the tapping machine (derived from L_n); and $F_{0,\text{TM}}$ is the blocked-force spectrum of the tapping machine. Z_i and Z_a are the impedances of sources and construction, respectively.

Impedances of sources and of floor systems are under investigation. At the time being, no standardized definition or measurement method is available. Preliminary results were presented by (Scholl 2001; Brunskog and Hammer 2003; Lievens and Brunskog 2007), among others. The curves show agreement in principle but they differ in some details. Examples are shown in Fig. 13.7.

For foootsteps, the model of point contact, however, is just a first approach. To what extent it is important to consider the actual contact area and the temporal dependence of contact forces on the area must be studied. And after all, the problem of nonlinear contact due to changes in stiffness caused by weight balancing on the heel and toe during walking are factors that create uncertainties. It is the task of further research to determine whether these effects are noticeable in auralized signals.
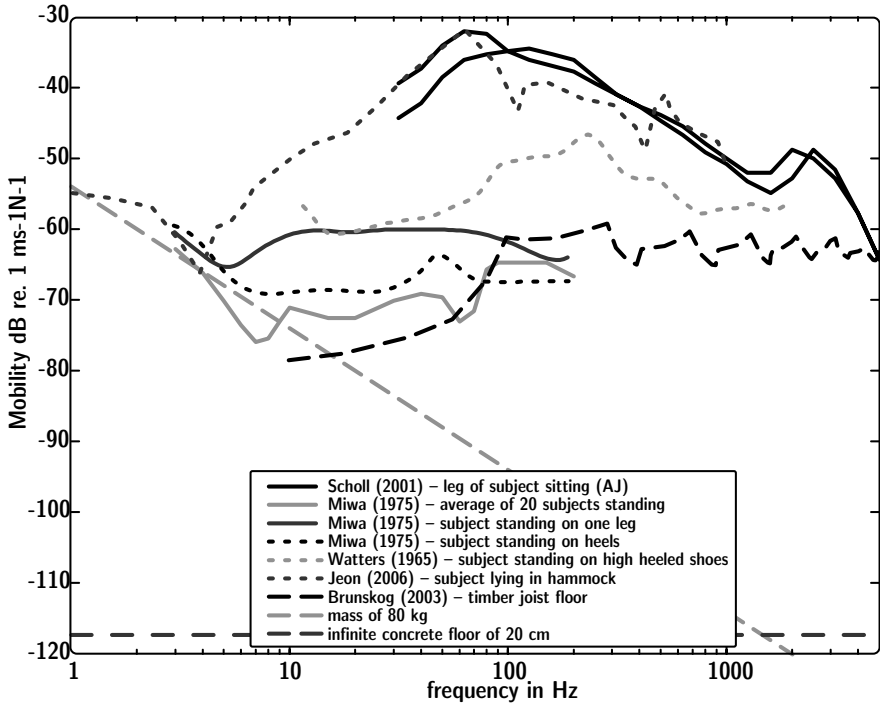
Fig. 13.7. Impedances of feet and floor constructions (after (Watters 1965; Miwa 1975; Scholl 2001; Brunskog and Hammer 2003; Jeon et al. 2006))

14 Binaural transfer path synthesis

In this chapter, the acoustic model of airborne and impact sound transmission and the subsequent auralization technique is generalized to other applications in acoustics and noise control engineering. 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 to achieve the correct level and colouration. Phase aspects can often be neglected in the 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.

Binaural transfer path synthesis (BTPS) is closely related to binaural transfer path analysis.[71] The analysis part is an indispensable tool of acoustic engineering. The applications are manifold in identification and characterization of sources and weak elements of sound transmission and of 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. This is possible not only in a simple reproduction of the situation analyzed, but with a variation of the complete system and with exchange of system components. In the example of sound insulation in buildings, the paths of direct and flanking transmission were treated exactly in this way. Adding linings or elastic interlayers in one of the paths is a typical variation of the system which can well be studied with regard to effective sound insulation performance, and it can be auralized for comparison of systems and their variations.

The concept of binaural synthesis, an important component of the concept of BTPS, was introduced in Sect. 9.3. Any monaural sound can be linked to a specific direction of incidence. In addition, sound that carries the characteristics of transmission paths can be modified in loudness and colouration. With these elements, the binaural transfer path synthesis is defined.
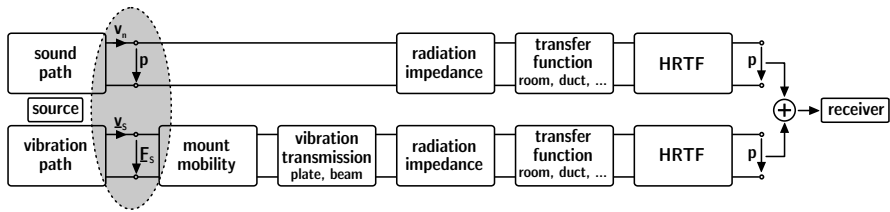
---

[71] so-called BTPA

Fig. 14.1. Block diagram of BTPS

Tools for transfer path analysis and synthesis are standard methods in the automotive industry. Early in the design and development of an automobile, the engine and the car body and particularly combinations of both are studied with regard to the structural and acoustical performance. We therefore take the discipline of vehicle acoustics as a 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 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 goal 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. 14.1 are two-ports. The upper path represents airborne sound generation and transmission. The source may be modelled as a volume flow source which acts on the radiation impedance. By coupling source and radiation impedance, the effective velocity is calculated. 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. 10.2.1. Sound radiation and propagation in a room can as well be simulated by using geometric models and binaural filters such as introduced in Sect. 11.6. In general, the lower path contains three degrees of freedom (x, y, z) or up to six, when rotational states are included.

The two-ports include circuits of concentrated components such as springs and masses. For higher frequencies, this approach might be no more adequate. The system might require field quantities distributed in the geometry (for instance, the velocity pattern on a plate or membrane), but

still then the velocity distribution may serve as an input quantity if it is derived on a modal functional basis.

The boxes on the right side represent signal processing from the model of physical wave propagation into the 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 Sect. 9.3.

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 for analyzing and optimizing acoustic signals. In noise control engineering for trains and aircrafts, it is used increasingly. The methodology can be applied as well to any other machines, household appliances and devices of daily life (such as, for instance, personal computers, dishwashers or vacuum cleaners). But for better understanding of the concepts, the following discussion is focused on the example chosen, which is automotive engineering.

14.1   Source identification and characterization

The specific behaviour of sources in general is related to the degrees 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[72]. 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. In Fig. 14.2, it can be observed 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.

---

[72] When driving at high speed, of course, wind noise is another important sound source.
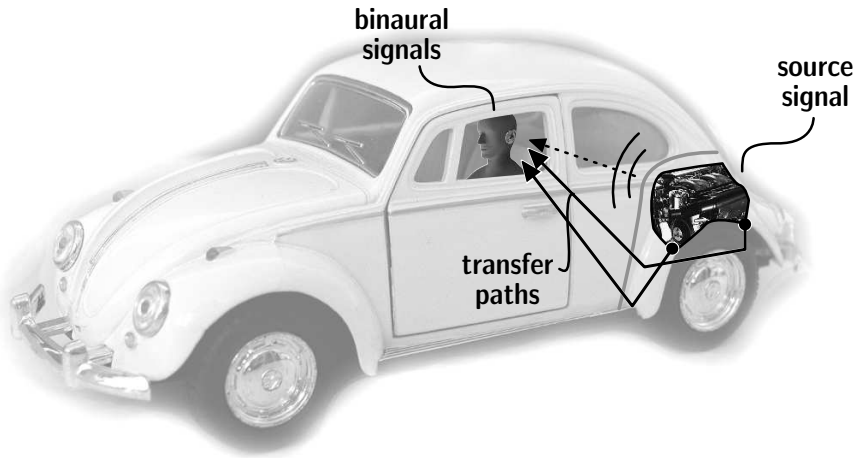
Fig. 14.2. Binaural transfer path synthesis for simulating in-cabin sound[73]

Transfer path determination and interfacing to the source is the key to creating 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 Sect. 14.1.1.
– In the component of vibrational force, feedback between the source and the transfer function is significant. This is the main problem discussed in Sect. 14.1.2.

14.1.1 Airborne sound sources

We start again by discussing vehicle sound and the combustion engine. For airborne sound transmission, the motor must be characterized as a primary source. The motor block is heavy and stiff. Vibration of the engine body induced by the engine operation is not at all affected by the acoustic load of the fluid medium (air). Thus, the source acts as an ideal velocity source. We don't have to take feedback at the interface between source and transfer function into account. Instead, severe difficulties must be faced when the specific pattern of the source vibration must 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 such as FEM or a detailed experimental

---

[73] This example illustrates BTPS for a vehicle with a rear engine. It is also applicable, of course, to cars with front engines.
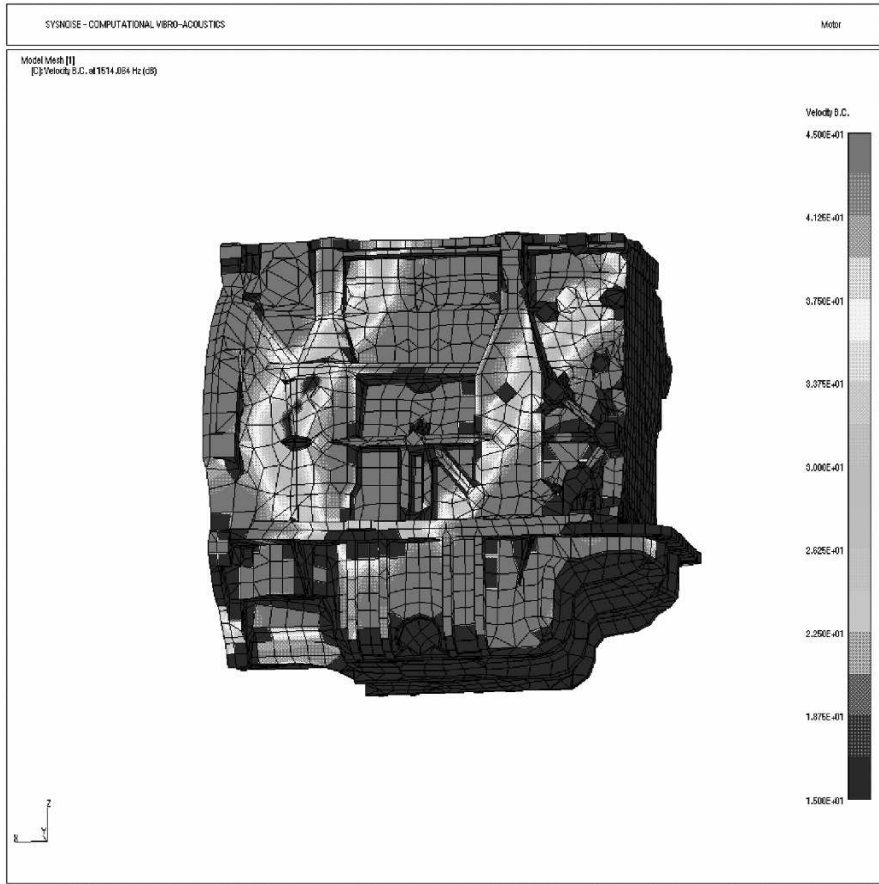
Fig. 14.3. Vibrational distribution on a combustion engine. Measurement result by using a scanning laser vibrometer

modal analysis.[74] These methods can be used directly to determine the surface velocity distribution.

The actual vibrational pattern can also be approximated by using parameter sets of multipoles (Sect. 2.4), spherical harmonics (Sect. 2.5) 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 at 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 such as an anechoic environment are required. For further processing in BTPS, the parameter

---

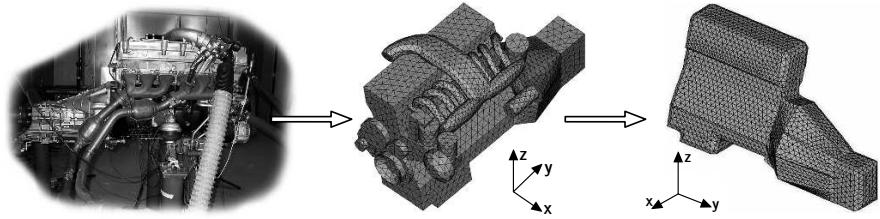[74] by using laser Doppler vibrometry, for example.

Fig. 14.4. Volume-conservative steps to create an engine model for BTPS (Römer 2004)

sets of multipoles, spherical harmonics or pressure maps are used as intermediate source data. But BTPS for airborne sound sources in vehicles now suffers from the problem that 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, therefore, the method applied is not consistent physically, but is an approximation for acoustic engineering, whose validity must be checked in each application.

When engine data are obtained in situ, i. e., in the engine compartment at the engine mounted inthe 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 us to model the source by estimating the total volume flow, while the actual vibrational pattern is ignored. However, the validity of this approach must be checked carefully, too.

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 in a way that all relevant acoustic effects are captured. An overall system modelling solution, of course, is the only approach with high precision. Here, 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. The problems there are given by the necessity of describing the source by an air flow, such as in the intake and exhaust system, or tonal components such as 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, which cannot be discussed in detail here, must be considered in specific problems.

One important task to be explained in more detail is that of measuring sources and transfer functions in situ. In the 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, the 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 cross talk between the measurement points and due to a somewhat arbitrary distance and near-field effect.

Nevertheless, it is possible to apply BTPA and BTPS based on in situ measurements, if the calibration is adjusted between the test facility and in situ conditions. These calibration spectra can be obtained by experiment, too.

14.1.2 Structure-borne sound sources

Structural paths are not as multidimensional as airborne sound fields.[75] 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 another 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 Sect. 10.2.1).
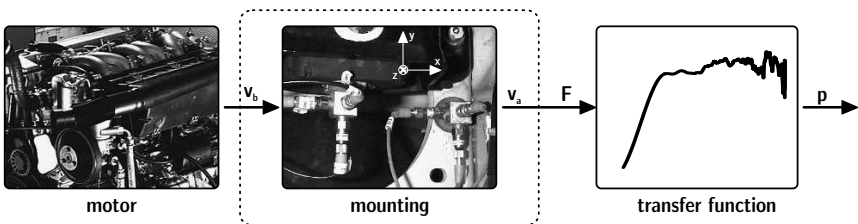

Fig. 14.5. Primary vibrational source connected to mount and transfer function (Sottek et al. 2005)

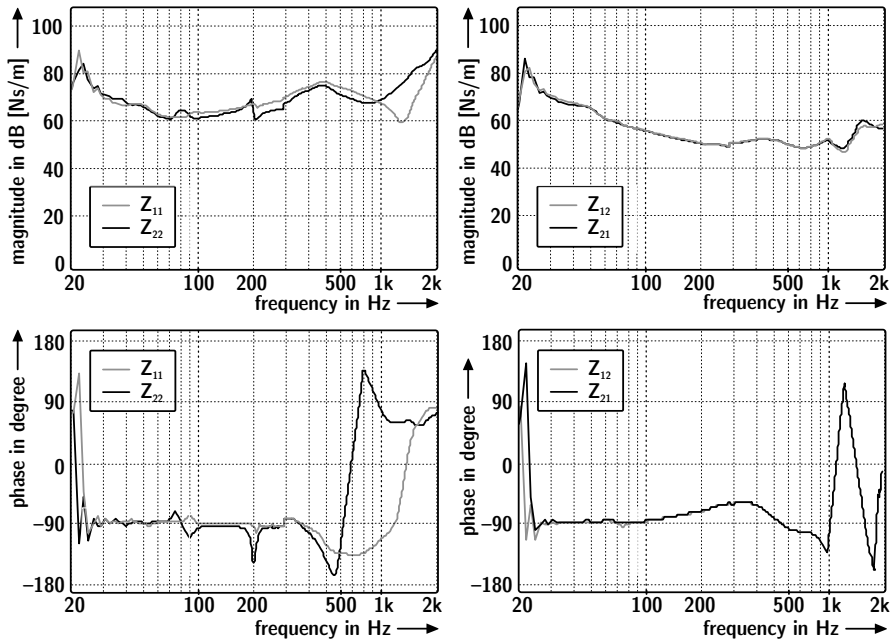[75] which is true at least for structures with point or line contacts.

Fig. 14.6. Transfer impedance of engine mounts (Dohm 2004)

Well-defined load impedances are a prerequisite for 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. 14.6.

14.2   Transfer path characterization

In BTPS, sound and vibration propagation from the source to the receiver are 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 Sect. 10.1); for vibration components, two-port matrices are used. Binaural filters are introduced in the end to account for the direction of sound incidence.

In a straightforward approach, transfer function filters can also be measured between the source and the receiver's ears directly. The source arrangement (see Sects. 14.1.1 and 14.1.2) is thus connected to the receiver's ear by filters, whose number coincides with the number of source signals recorded.

For each source point, this measurement must be performed separately and independently to avoid cross talk between paths. In the example of a transfer function between the vibration source velocity and the ear sound pressure, the direct measurement required 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 by using a dummy head for all paths, one by one.

In a reciprocal arrangement, however, all paths can be measured in parallel. Due to vibroacoustic reciprocity (see also Sects. 10.1 and 10.2), the ratio

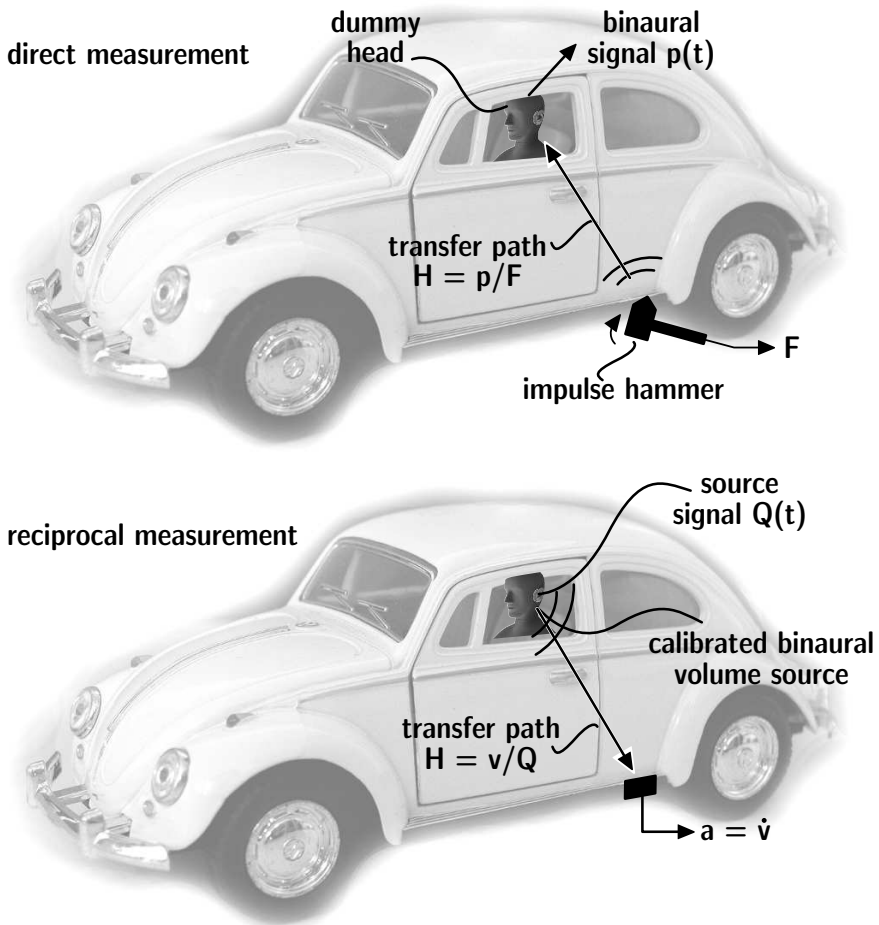$$H = \frac{p_{\text{receiver}}}{F_{\text{source}}}\bigg|_{v=0} \tag{14.1}$$

Fig. 14.7. Measurement of vibroacousic transfer functions (see also (Sottek 2004; Sellerbeck 2003))

is equivalent to

$$H' = \left. \frac{v_{\text{eq,source}}}{Q_{\text{eq,receiver}}} \right|_{F=0} = H \ . \tag{14.2}$$

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. 14.7).

14.3 Auralization in BTPS

When the primary source is recorded in operation, such as an engine running at increasing speed (rpm), this signal serves as an input for binaural
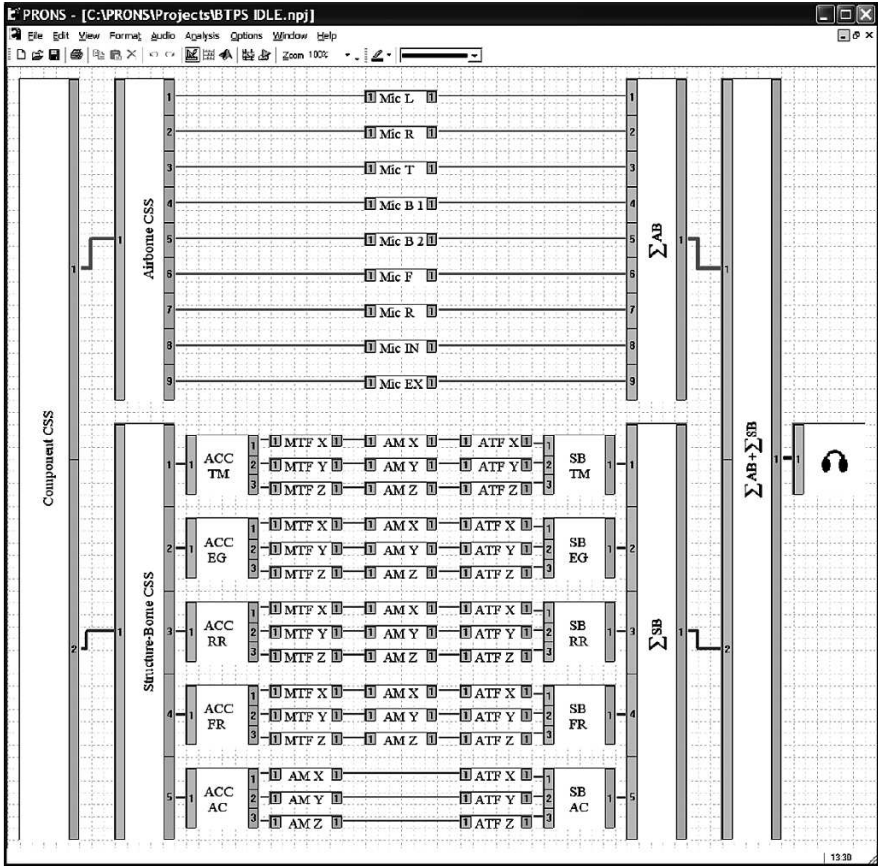

Fig. 14.8. Example of a BTPS software user interface (after (Sottek et al. 2004))

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, such as wind noise or tyre noise. Integrated into a driving simulator, the binaural transfer path auralization allows detailed psychometric tests on sound character and sound design targets.

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 development of new engines, car prototypes are not yet available. Therefore, 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 different frequency ranges. Also of interest is the variability of input data on material properties and the variability of junctions, related to the variation of those parameters in the production process.
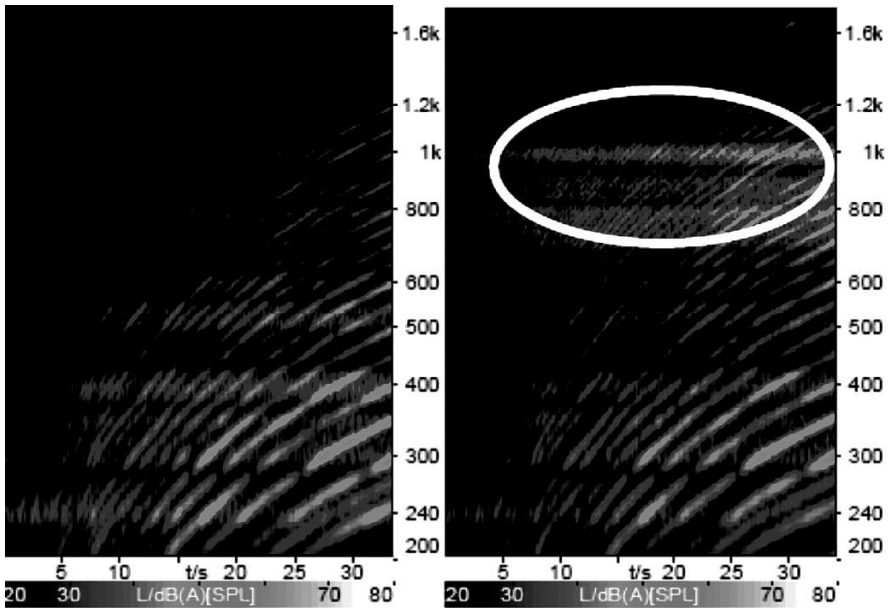

Fig. 14.9. Example of spectrograms showing the effect of changing engine mounts by using BTPS (after (Behler et al. 2006))

The increasingly sloped curves in Fig. 14.9 illustrate an increasing rpm. A combustion engine is mounted in the car body by using two different engine mounts (different stiffness). The time and the frequency, respectively, are shown on the abscissa and the ordinate. The sound level in dB(A) is presented in a grey scale. On the right-hand side, the resonance effect illustrates a suboptimal impedance match between engine mount and car structure. An rpm-independent spectral maximum appears between 800 Hz and 1 kHz 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.

15 Aspects of real-time processing

Virtual environments shall be multimodal and interactive. The technology for simulation and reproduction must include the aspect of several sensory inputs in real time and of the interaction with the person working in the virtual environment. To achieve real-time performance, specific run-time conditions must be taken into account to stay within acceptable limits for latency and update rates. For the acoustic component of VR, we can use information from psychoacoustics about these limits.

It should be noted that the requirements may be contradictory. To illustrate this fact, we assume a sound source on a circular rotational path around a listener. The source is permanently moving. A high update rate will create a smooth, continuous transition from location to location on the circular path. A high update rate, however, allows only little calculation and processing time for simulating sound propagation and binaural synthesis. The new geometric situation (new source position) may be available in one step of simulation, and delays are included between the actual position and the reproduced binaural sound pressure. Interaction will be disturbed since the system's feedback is delayed. Vice versa, if the computation time available for simulation is larger, the update rate is slower, from which 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 are introducing latency already. Accordingly there is only little time left for acoustic simulation (acoustic rendering and reproduction). For this reason, limited computer performance prohibits the acoustic component of VR from reaching full authenticity, and this can be expected to be the case for some 10 years. For the moment, however, the performance is capable of creating 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, good perceived coincidence of the stimuli is still found. This result might be caused by faster neural processing for the visual path, compared with the auditory path. In other words, acoustics rendering may involve some 30 ms longer latency than visual rendering (Kohlrausch et al. 2006).

What is "real-time processing?" An important factor to be discussed here is the previously mentioned interactivity. This means that the dynamic changes introduced by the active user must have an immediate effect on perception. The latency, accordingly, must be sufficiently small so that the listener is not disturbed or irritated (Brungart et al. 2004). In the ideal case, it is below the just noticeable differences (see Sect. 6.2). Furthermore, the frame rate (update rate), defined as the time the acoustic input is changed, the simulation recalculated and the output presented must be sufficiently high to ensure a smooth, continuous scene. These requirements are associated with the parameters for responsiveness and smoothness. If the update rate of the simulation and the perception of continuous movements are set just 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 at 50 Hz, the maximum processing time for simulation amounts to 20 ms.

The key technology of real-time systems can be divided into the components of sound rendering (creating of auditory cues by software algorithms) and of sound reproduction (3-D 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 of crucial importance.

15.1 Real-time binaural synthesis

Before we consider a full interactive simulation in real time, an intermediate step is introduced. Binaural sound can be preprocessed or recorded for some locations and orientation 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,[76] dry sound is convolved with the valid binaural impulse response for the actual position and orientation of the listener. Listener movement is tracked. Accordingly, the best matching binaural filter for the position and orientation is 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 is used as a basis for simulation or measurement.

---

[76] also referred to as "walkthrough" (Dalenbäck 2006).

15.1.1 HRTF in multiple degrees of freedom

Advanced developments allow dummy heads also to introduce multiple degrees of freedom (DOF) for the head-to-shoulder orientation, which is best illustrated by 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. The result for plausibility and naturalness was excellent. The test subjects, on average, could not tell if a stimulus was presented via auralization (head-tracked binaural real-time convolution) or if it was the original. This test was performed in situ in the real hall (see (Lindau et al. 2007) and Sect. 16.3).

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. An extension to room–related coordinates creates a very big effect by enhancing plausibility and immersion, even for the use of nonindividualized HRTF. This finding illustrates the importance of dynamic cues of localization[77] and the necessity of implementing this feature in binaural systems (Mackensen 2003).
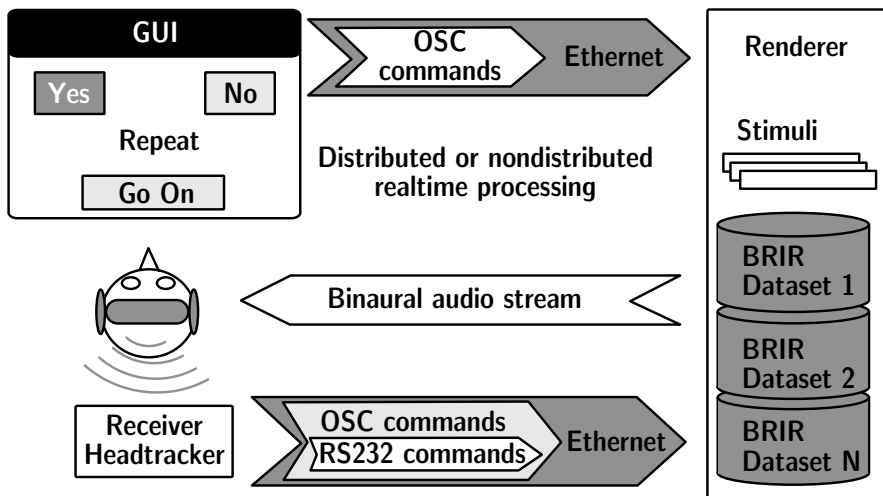

Fig. 15.1. System for dynamic real-time auralization of precalculated or measured binaural impulse responses (after (Moldrzyk et al. 2004; Lindau et al. 2007))

---

[77] Like humans also use for normal listening, when they move their heads slightly to solve the problem of front–back confusion.

To account for dynamic cues, interaction and adaptation of filters, the position, orientation and motion of the listener must be known. Systems that deliver 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 calculating the reference point for stereoscopic projection. As for stereoscopic projection, the room environment is the invariant coordinate system.

Head Tracker

Several tracking devices are available for listener-centred 3-D video projection and binaural reproduction. 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 on the head. He precision and latency of the head tracking system are the most relevant parameters for qualification of applicability in VR technology.

Fast convolution

Real-time convolution is a specific problem, especially in binaural synthesis systems where many channels must be processed in parallel. In dynamic room auralization, binaural impulse responses must be changed rapidly without audible transitions or "clicks" (McGrath 1996). Segmented convolution (see Sect. 9.2.1) is efficient in this respect. 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 in multiprocessor systems which follows the trend in PC hardware platforms. Furthermore, fading between impulse responses must also be optimized to avoid colouration (comb filter effects), clicks and abrupt changes in the distance or the direction of the direct and early sound.

15.2   Room acoustical real-time auralization

When the sound event is related to a virtual room, the geometric situation is characterized by a number of polygons, whose acoustic data represent various coefficients of absorption and scattering. A "wall" is considered an 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 efficiently used for ray tracing and image source simulation.

Valid room boundaries are defined by polygons that include the acoustic data of absorption and scattering. The polygons are defined by a list of vertices. It is advantageous to separate the polygons into convex parts to speed up the ray intersection test.[78]

15.2.1 Source and receiver

The source and receiver characteristics are determined by their position, orientation and directional pattern. 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 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. The HRTF is implemented in the same way as described in Sect. 9.4.
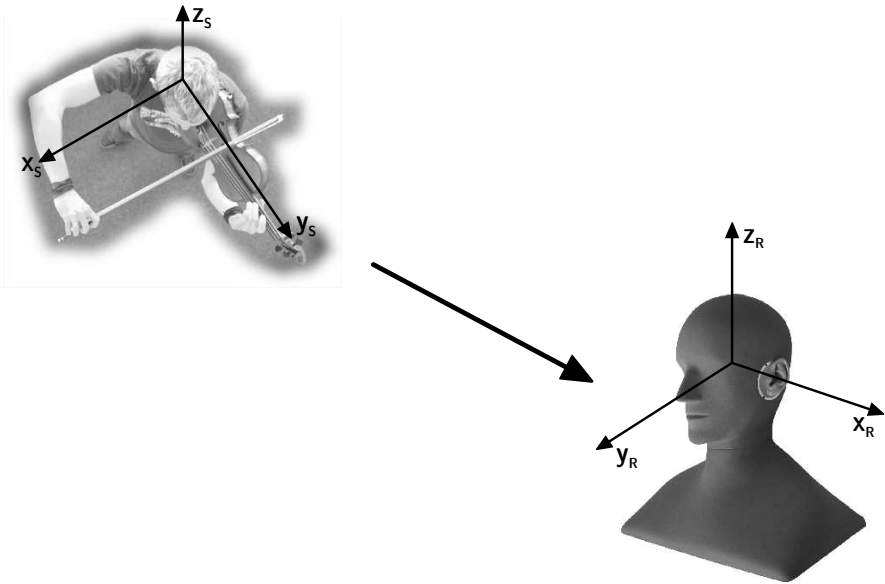

Fig. 15.2. Algorithmic characterization of source and receiver

---

[78] Test of "point in polygon."

15.2.2   Real-time processing of image sources

Image sources are given by their position in 3-D 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. Also important is a source-related coordinate system for translational or rotational transformation; see below.

The specular part of the binaural room impulse response is composed of direct sound and specular reflections (see Sect. 11.3). The image source contributions can be determined best in the frequency domain. The spectral resolution of input data such as absorption coefficients, however, is hardly useful for signal processing. HRTF data are stored in discrete spectra[79] with a resolution of $\approx 300\,\text{Hz}$. The spectral density of a HRTF must be identical to 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. In higher octave bands, this requires more and more interpolation finally resulting in a linear resolution with constant frequency spacing.

The same problem occurred if a modal response of a transmission system was calculated with high frequency resolution (possibly pure-tone excitation makes this high resolution necessary). other spectra such as 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 result from simulation (see Sect. 11.7). In the frequency domain, the superposition of the components reads:

$$\underline{H}\big|_{\text{left,right}} = \sum_{i=1}^{N} \frac{\text{e}^{-j\omega t_i}}{ct_i} \cdot \underline{H}_i(\theta,\phi) \cdot \underline{H}_{i,\text{air}} \cdot HRTF(\vartheta,\varphi)\big|_{\text{left,right}} , \qquad (15.1)$$

or converted into the corresponding binaural impulse response:

$$h(t)\big|_{\text{left,right}} = IFT\left\{\underline{H}\big|_{\text{left,right}}\right\}. \qquad (15.2)$$

Where $h(t)$ denotes the binaural room impulse response, $j\omega t_i$ the phase lags due to retardation of the ith reflection, $1/(ct_i)$ the distance law of spherical waves, $\underline{H}_i$ the source directivity (Sect. 2.4) in source coordinates (θ, ϕ), $\underline{H}_{\text{air}}$ the low pass of air attenuation (Sect. 3.6), and HRTF the head-related transfer function (Sect. 6.3.1) of the sound incidence in listener coordinates at a specified orientation (ϑ, φ).

---

[79] HRIR in FIR filters of 128 taps represent discrete spectra of $\Delta f = 344\,\text{Hz}$ (at a sampling rate of 44.1 kHz).

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 the parameters are changed in dynamic scenes. The binaural impulse response is to be changed accordingly.

Room subdivision

It is essential 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. For this, approaches for reducing 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 Sect. 11.3.5).

Dynamic interaction

When the cloud of image sources is 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. At changes in this position, which is quite obvious in dynamic situations in VR environments, all image sources must be recalculated. However, we can use geometric facts. It is, thus, not required to start the imaging process from the beginning, instead it is only necessary to rotate and translate the image sources given.

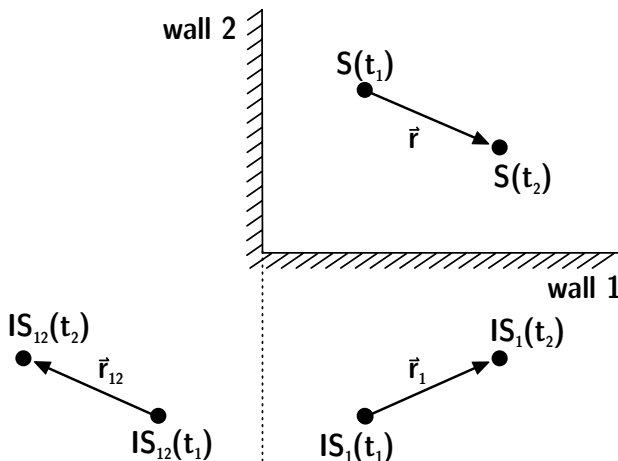The difference vector between source and receiver[80] is used as a translation vector.

Fig. 15.3. Translation of image sources

---

[80] Representing the direct sound path.

Similar to mirroring of vertices at walls (see Sect. 11.3.1), we can calculate the mirrored translation vectors. As illustrated in the example in Fig. 15.3, 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. Such a matrix is to be found which provides the first-order image source being placed in the position corresponding to the actual movement of the original source.

$$T = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix}. \tag{15.3}$$

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.

From the figure, the vector elements $\vec{u}, \vec{v}$ and $\vec{w}$ can be used to calculate the coordinates of the image space:

$$\vec{u}' = \vec{u} \cdot T = \vec{u} ,$$
$$\vec{v}' = \vec{v} \cdot T = \vec{v} , \tag{15.4}$$
$$\vec{w}' = \vec{w} \cdot T = -\vec{w} .$$

Hence, vectors parallel to the wall plane are invariant, while the normal vector is mirrored. With these nine equations, the transformation matrix of any shift vector can be determined. It is also observed in Fig. 15.4 that the shift vector of image source 12 is obtained by mirroring the shift vector of
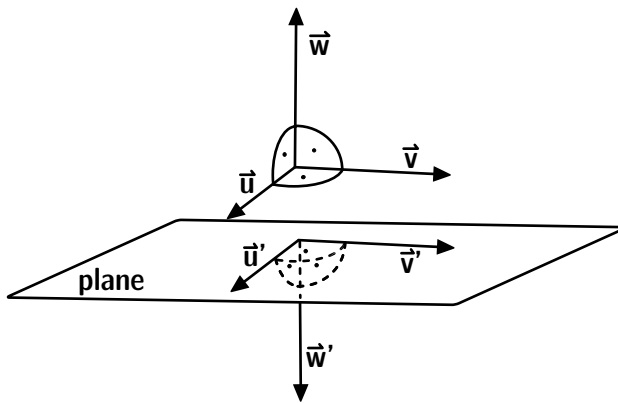

Fig. 15.4. Mirroring the wall orthonormal system

image source 1. The total shift vector is thus the result of matrix multiplication of all wall matrices involved:

$$T_{total} = \prod T_{walls} \ ,$$
(15.5)

and accordingly,

$$\vec{r}_{i,new} = \vec{r}_i + T_{total} \cdot \Delta \vec{r}_{source} .$$
(15.6)

As long as the geometry of the environment is invariant, the transformation matrices are also invariant. Calculation of new image source positions is thus faster by using the matrix.

15.2.3 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 must then be added by using other techniques, such as reverberation processors. Reverberation processors are well known in audio engineering (Griesinger 1989; Blesser 2001). To get a first effect and a proper reverberation time of the scene, these reverberation machines can well be used.

The first digital[81] reverberation processors (Schroeder 1961) were based on all-pass 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 all-pass filter approach involved integrating 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 the sense of "non-Sabinean" behaviour.[82] In these cases, artificial reverberation is not sufficient and a more detailed estimate 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 of computational time. In the 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 $10 \times 10°$ of the spatial angle.

---

[81] Historically, also implemented in analogue hardware.
[82] This would apply to coupled rooms or rooms with a flat or long shape.

These resolutions are actually sufficient to describe the late reverberation (see Sect. 15.3 and Fig. 15.5).

Methods of ray tracing, radiosity, waveguides, randomized tail-correction or other stochastic methods are well qualified for obtaining late reverberation spectral 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 created in post-processing; see Sect. 15.3. 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 15.5 illustrates the late reverberation energy decay plotted in time and frequency domains 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.
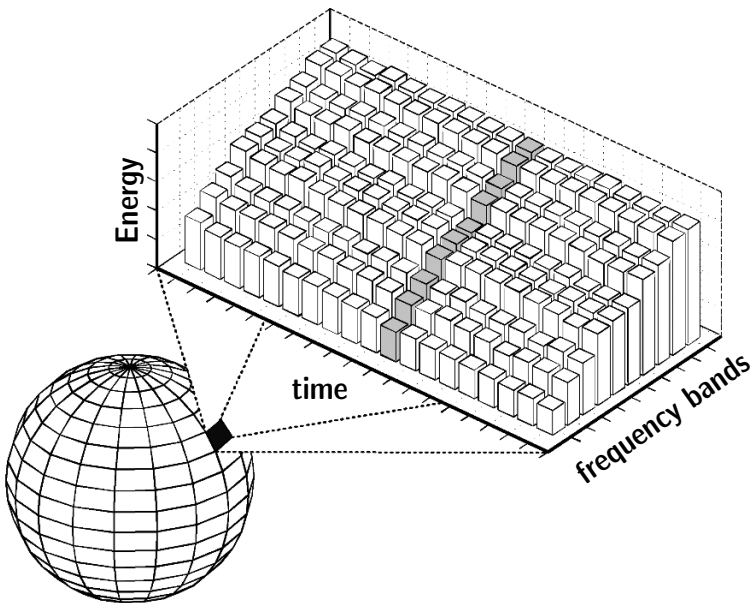

Fig. 15.5. Receiver-related sphere divided into directional groups with energy decay histograms for every group and frequency band (after (Schröder et al. 2007))

15.3   Hybrid real-time room auralization

To start with the discussion on hybrid real-time models, we consider the psychoacoustically relevant temporal resolution in the impulse response. At least two regions can be separated: the early part and the late part. 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 late reverberation modelling by a time-invariant recursive digital filter structure. The evaluation by auditory models and by listening tests is an important part of this work, comparing also auralization with the real room under test. This and other studies are oriented toward 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 a 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. By investigating the user's behaviour in the VR system, several aspects can be noted (LaViola 2003). These results are not only interesting for determining the absolute speed of the movements and permitted latency, but also for predictive algorithms of filter updates.

Head rotation is the quickest and most abrupt movement. Then, translational movements are considered that take place at the user or at the source. Translational movements are in the speed range of 30–50 cm/s. Head rotations may be as quick as some 45° per second. If we use these data as a limit, the filter update rate and the necessary resolution can be estimated. The azimuthal resolution is then about 1°, and the time available for filter update is 20 ms (Lentz et al. 2007). Although the movements are limited in speed, required update rates can be estimated. Table 15.1 shows an example of a set of update rates in the specific temporal regions of the binaural room impulse response.

These data are, of course, subject to future development, which will allow faster dynamic scenes. Current research, therefore, deals with aspects of faster convolution processors and the limits of time-invariant system theory. Finally, the goal will be to create *any* dynamic scene including not only time-variant systems but also nonlinearities such as the Doppler effect.

Table 15.1. Proposal for update intervals for some interactions of listener and source movement (after (Lentz et al. 2007)).

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

We conclude the discussion of real-time processing and real-time rendering here. 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 be evaluated only when rendering and reproduction are discussed as a unit. In the next chapter, therefore, sound reproduction systems are introduced, and examples of complete VR system architectures are presented.

16 3-D sound reproduction and virtual reality systems

Computer simulation of acoustic scenes is an important prerequisite of rendering. Technology for 3-D 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 involving psychoacoustically relevant cues. These cues may differ from one VR application to the next. Some applications require an exact localization, while for others, monaural spectral features such as reproduction with exact loudness and timbre are more important.
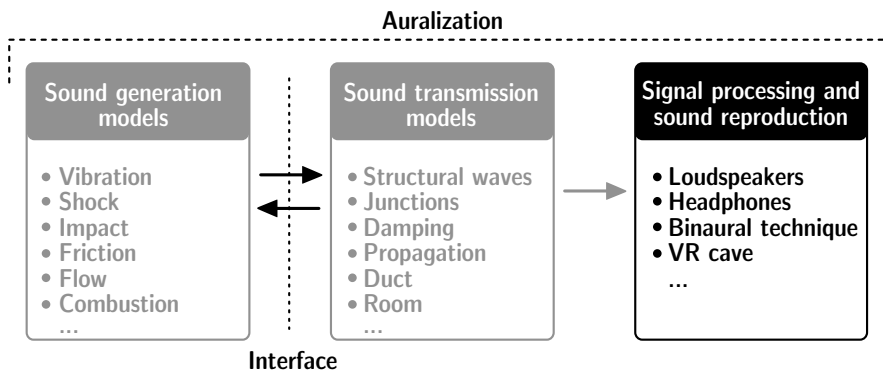
Fig. 16.1. The components of sound reproduction

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. A 3-D 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 physical models and a high degree of realism in the components of sound and vibration generation, transmission and reproduction. Recording engineers for classical music have a different goal. Even for live recordings, they will use recording techniques and strategies of microphone placement and mixing to obtain the best result in the final sound for

Fig. 16.2. Concert hall model in a CAVE-like environment (CAVE^{TM} Automatic Virtual Environemnt)

home environments, replayed by stereo or 5.1 equipment. Thus, the balance of instruments or instrument groups is manipulated for aesthetic effects.

First of all, for acoustic virtual reality, we accept the strict and exclusive goal of as much authenticity as possible related to three-dimensional perception. A "neutral" reproduction related to linear distortions is required anyway.

16.1 Headphone systems

Headphones or other audio systems integrated into head-mounted displays are well qualified to serve as reproduction transducers and, thus, they are widely in use. 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 listener's ear canal. Wearing comfort and unnatural ear occlusion are additional factors that affect the quality of the hearing sensation. The so-called "in-head localization" is one example of such unwanted effects. Externalization of sound sources is one of the main issues in a discussion of headphones. When externalization is insufficient, immersion in a VR system is drastically reduced. With proper equalization and special attention to high-frequency radiation into the ear canal, this problem can be partly solved. Adaptive filtering (head tracking; see Sect. 15.1.1) for taking head movements into account is also a very important tool for creating realistic localization and externalization.

Fig. 16.3. Headphones of various types: open, closed, semiopen

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 such as a "piston in free half space." Instead, the radiation impedance into the ear canal is relevant, which brings us to the first difficulty. The properties ear canals of listeners have a large variance among a population of test subjects. As concerns 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 are dependent on the individual anatomy.

Artificial ears were developed as kind of "average ear," but their applications are restricted to special headphone types.[83] Even in ideal measurement conditions for digital equalization and calibration, there remains the uncertainty introduced by mounting at the real ear. Particularly in closed headphones, uncertainties are observed due to leakage. Interindividual differences are also important there.
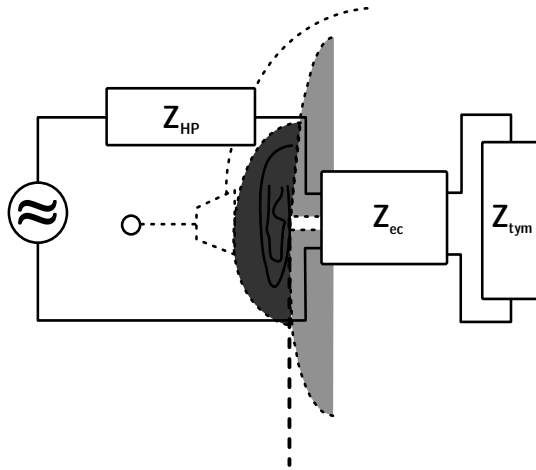
---

[83] typically audiometric headphones.

Fig. 16.4. Model of a headphone mounted at the ear

In Fig. 16.4, the headphone source impedance, $\underline{Z}_{HP}$, is coupled with the ear canal impedance, $\underline{Z}_{ec}$, and finally with the termination impedance of the eardrum. Headphones connected to ear canal and eardrum impedances
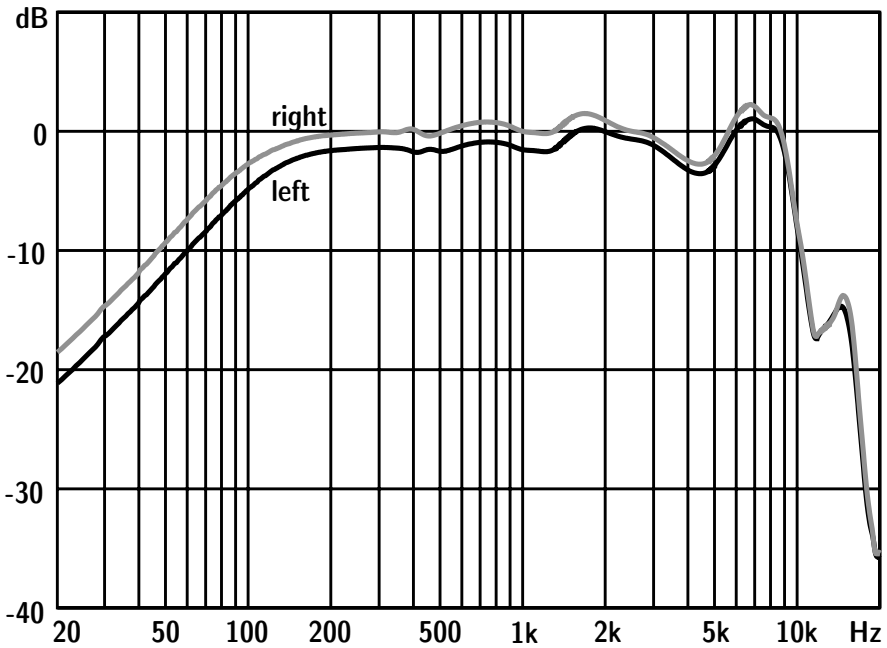

Fig. 16.5. Example of an in-ear headphone frequency response measured with an ear simulator (IEC 60711)

have been under investigation in basic research for many years, until today (Hammershøi 1995). In a purely technical sense, the eardrum sound pressure represents the complete excitation signal of the auditory system. For this assumption, other paths of sound transmission to the cochlea, such as bone conduction, are neglected. With a good model of the ear canal impedance and definition of the sound pressure at the ear canal entrance as the driving signal, the eardrum pressure is given unambiguously. The problem, however, suffers from many uncertainfactors, as described previously. Also, occlusion by a headphone affects the standing wave pattern and, thus, the transfer impedance of the ear canal. Only in a specific test scenario such as a blocked ear canal is the driving pressure measured with a clear and accurate reference.

16.1.1 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 binaural cues.

Correction (equalization) filters have twofold purposes. First, they must ensure 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, at replay, the sound path between headphone and eardrum is included twice. Thus, a correction filter must be used to extract the ear canal (see Fig. 16.4). Other filters are also used, but these are defined for 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 by the loudspeaker (plane wave) flat spectrum. Binaural recording replayed
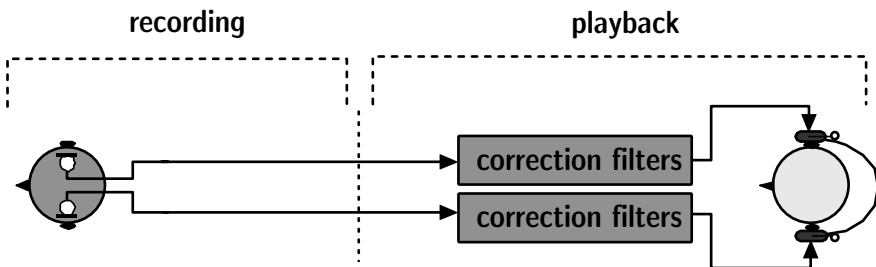

Fig. 16.6. Arrangement for reproducing binaural signals (after (Blauert 1996))
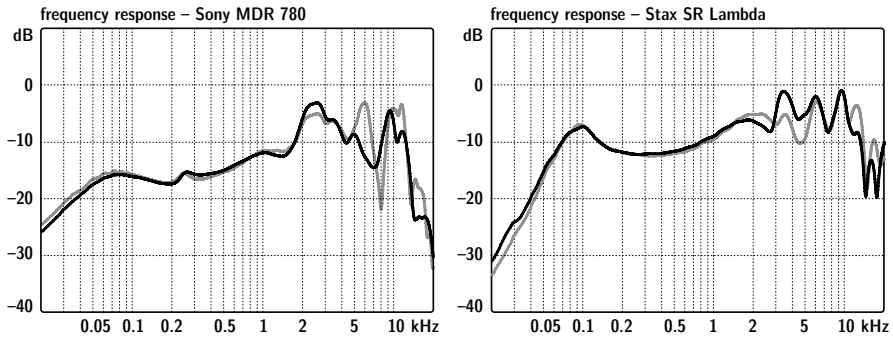
Fig. 16.7. Examples of typical free-field sensitivities of headphones

by loudspeakers, accordingly, carry the HRTF twice, one from the dummy head in the recording situation and the other 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.

But due to the need for compatibility with loudspeakers, 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; diffuse-field headphones accordingly, are used for random incidence of incoherent sound waves. A reproduction of a recorded diffuse field[84] by a diffuse-field headphone yields the same hearing sensation as that in the real (recorded) situation.

16.1.2 Individual filters

All definitions given above hold 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 can be identified in individual features in the HRTF, particularly in the frequency range above 6 kHz.

When individualized HRTFs are available 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 the electric voltage fed to a loudspeaker in a certain direction and at a large distance (>2 m) and the sound pressure measured in the ear canal of a test subject.

---

[84] for instance, created in a reverberant room far outside the reverberation distance.
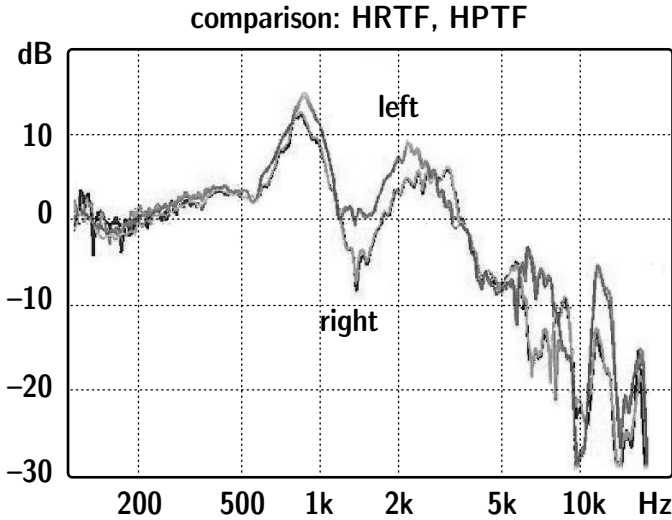
Fig. 16.8. Four equalized binaural transfer functions, two for a loudspeaker and two for headphones (HPTF: headphone transfer function) into the ear canal (individual test subject). The two curves corresponding to left and right ears are identical

This electroacoustic transfer function can be interpreted as an individual HRTF.[85] The same transfer function should be present when the headphone is applied. Figure 16.8 shows a comparison of this experiment. In this figure, four curves are shown, two of each for the right and the left ears, respectively. They are identical with respect to the resolution of the plot. The deviations are smaller than 0.5 dB. Accordingly, no audible effects should be noticed in a listening comparison. In fact, practically 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 matters. Furthermore, the psychological component in such tests which aim at very small perceptual effects, must be considered. After all, it has been found that with exact equalization of input signals into the hearing system, the auditory sensation is exactly the same (Blauert 1996).

After individual equalization, the remaining differences between loudspeaker and headphone binaural reproduction systems are less than the just noticeable differences; see Fig. 16.9. More detailed results were published by Møller (1992), also including considerations of the accuracy of binaural cues and the final performance of headphone reproduction in listening tests. Finally, one can state that headphones can be used for 3-D sound reproduction with good results, but their equalization requires much attention.

---

[85] With exact calibration of the loudspeaker, it is actually the HRTF.
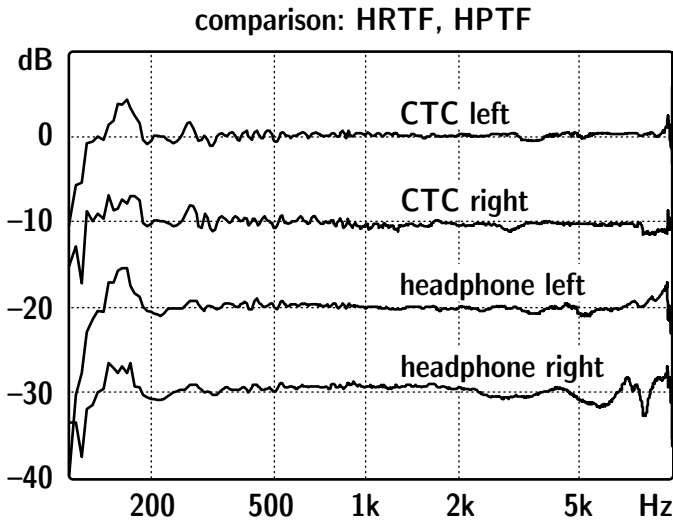
Fig. 16.9. Individually equalized transfer functions of loudspeakers (top, HRTF CTC) and headphones (HPTF bottom) from the free field to the reference point in the ear canal, divided by the individual transfer function between an electric voltage fed into a loudspeaker in a free field and the ear canal sound pressure. Curves are shifted in steps of 10 dB for better visibility
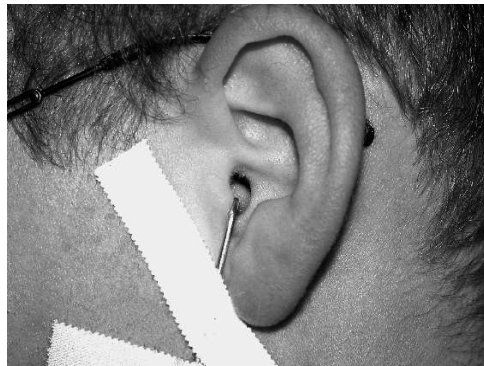

Fig. 16.10. Probe microphone in the ear canal of a test subject

Head-mounted displays

Head-mounted displays are equipped with small LCD or OLED[86] (or other) displays which create stereoscopic images for the two eyes. The resolution is rather small compared with full screen CAVE-like solutions, and the main disadvantage is the limited view angle, so that even one's

---

[86] liquid crystal device; organic light emitting diode

own hands and fingers must be rendered. The acoustic path is usually covered by integrated headphones. Due to the discomfort involved in wearing the device, immersion is affected. Advantages, however, can be seen in the feature of easily added components for haptic or tactile stimuli which do not interfere with vision and sound. All aspects of headphone technology hold for sound reproduction.

16.2   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 depends only on the two input signals to the eardrums. Also, loudspeakers arranged around a listening point ("sweet spot") may serve as 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.

Binaural technology is described in Sect. 16.2.4 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[87]. These approaches use 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 field reproduction is the method of wave field synthesis (WFS; see Sect. 16.2.4). Here, we create an approximation to the spatial sound field incident by using a microphone array too, but not at
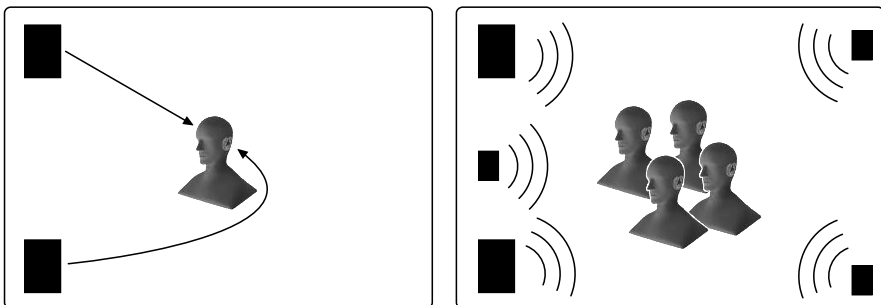

Fig. 16.11. Binaural technology (left) and sound field technology (right)

[87]   mostly in use is "Dolby Surround®"

a listener point. Instead, the microphone arrangement is larger and it is located on elementary geometric figures such as straight lines or circles around the listening area. In this way, the artificial sound covers a larger area, and more than one listener can be served with spatial sound.

16.2.1 VBAP surround sound

Starting from a home stereo setup, where the balance control between right and left channels is used for image shifting, additional pairs of loudspeakers can be placed around the sweet spot. The stereo balance controls the relative amplitude between right and left. Thus, 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 applied in home entertainment in the early 1970s with the name "quadrophony" (with no commercial success, however). Today, some virtual reality systems include surround sound systems called vector-base amplitude panning (VBAP). For ambitious integration of acoustics into VR, however, this method involves too many shortcomings such as inherent implementation of binaural processing mechanisms. Accordingly, VBAP is fine as an effect device, but not as a aplatform for unbiased studies of the human hearing and neural systems.

16.2.2 Ambisonics

Based on work by Gerzon (1976), the so-called B-format or technique of ambisonics was developed. The B-format is a four-channel recording standard that uses a sound field microphone.

Usually the channels represent the front-back (Y), up-down (Z), left-right (X) 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 Sect. 2.5).

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.[88] In spatial sound field processing, the decoding process is

---

[88] usually performed by decoding matrices.

Fig. 16.12. Sound field microphone for recordings in B-format (courtesy of Lab. of Acoustical Imaging and Sound Control, TU Delft)

nothing but a reconstruction by linear combination of zero (W) and first-order (X, Y, Z) spherical harmonics.

Apart from the strict physical approach of spherical harmonics, decoding schemes are in use, which take benefit from the localization and distance perception of human hearing. It must be kept in mind again, however, that virtual sound fields may be used for investigating the hearing system. A generally applicable reproduction system must not introduce any artificial auditory cue which is not part of the simulated sound.

16.2.3 Wave field synthesis

The technology of wave field synthesis, (WFS) is a rigorous approach to wave field reconstruction (Berkhout 1888). Its concept stems from seismic physics and source localization by array techniques. The basic mathematical algorithms are a kind of spatial Fourier transformation between the space and the wave number domains. Thus, complex wave fields are decomposed into elementary waves such as plane, spherical or cylindrical waves, by an appropriate transformation.

In plots of the position along a line versus propagation time, these wave fronts can be studied, either in a static condition or in an animated movie in motion, similar to results from FTD simulation.

The wave decomposition is achieved by analyzing the signals in microphone arrays. According to Huygens' principle, the points where the sound pressures were recorded at the microphone positions can be interpreted as
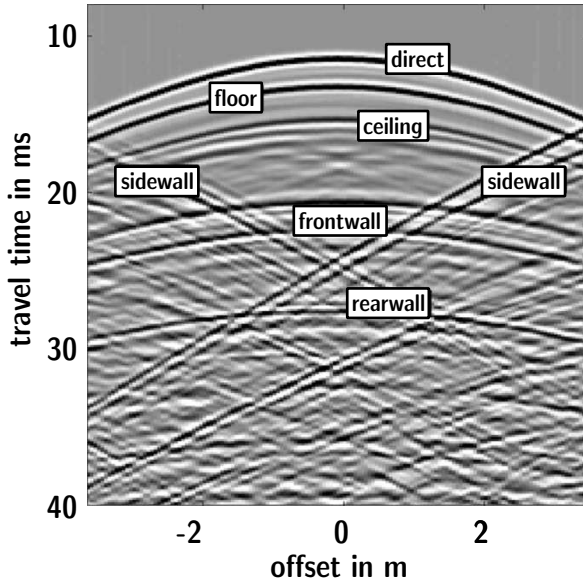
Fig. 16.13. Wave field analysis (courtesy of Lab. of Acoustical Imaging and Sound Control, TU Delft)

elementary sources. In a replay situation, the wave field is reconstructed by sending waves from these points,. This illustrates the step from wave field analysis to synthesis.

As long as the discrete spatial sampling is sufficiently high, any wave field can be reconstructed. Unfortunately, this prerequisite creates severe practical problems. A physically perfect sound field with frequency content up to $10\,\text{kHz}$ within a space of $3 \times 3 \times 3 = 30\,\text{m}^3$ required some
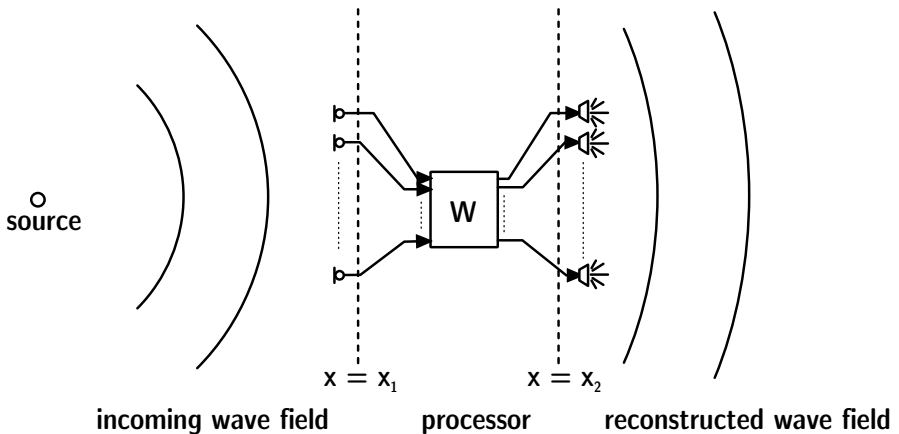

Fig. 16.14. Field reconstruction by loudspeaker arrays (after (Berkhout 1988))

50,000 loudspeakers of 3 cm size. However, in the same way as discussed in several sections of this book, physical constraints are not as relevant as psychoacoustic constraints. The existence of broadband signals and masking in higher frequency bands allows more relaxed conditions for loudspeaker size and distance. And the spatial resolution of human hearing is more sensitive in the horizontal plane. Hence, a two-dimensional array serves well for the most crucial spatial information. With these simplifications, practically installed WFS systems are in use with some 500 loudspeakers, thus, 100 times less effort than theoretically required.

The theoretical background of WFS is 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:

$$p(r) = \iint \left(g(r|r_0) \frac{\partial p(r_0)}{\partial n} - p(r_0) \frac{\partial g(r|r_0)}{\partial n} \right) dS_0, \qquad (16.1)$$

where r denotes the position of the receiver and the set of r_0 denotes the source positions in the loudspeaker array.[89] 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 consequence of the Helmholtz–Kirchhoff formulation is that sources within the surrounded volume cannot be modelled. As a compromise, solutions for near sources are in development, but these approaches require head-tracking, and this is valid only for one person. For sources from outside arrays, however, any arrangement is theoretically possible, and groups of listeners can be fed with 3-D sound (see Fig. 16.15).

For two-dimensional arrays and reproduction of the sound field in a plane instead of a volume, the functional basis is cylindrical waves. The corresponding Green's functions do not represent the "correct" distance law in 3-D 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 that sets no limitations on the way of listening such as position, orientation or movement. Spatial decoding is accordingly transferred in the WFS reproduction system, while each source[90] is recorded separately from other sources in the best possible anechoic situation. Furthermore some control parameters such as position and relative level of the source are encoded.

---

[89] The source notation by a running index is omitted here.
[90] typically a musical instrument or a voice.

Also the performance environment such as a room is not included in the recording, but is created at the end in WFS by mapping the parameters and signals into an actual, synthetic or augmented spatial situation.

Fig. 16.15. Installation of WFS (courtesy of the Fraunhofer Institute of Digital Media Technology (IDMT), Ilmenau, Germany)
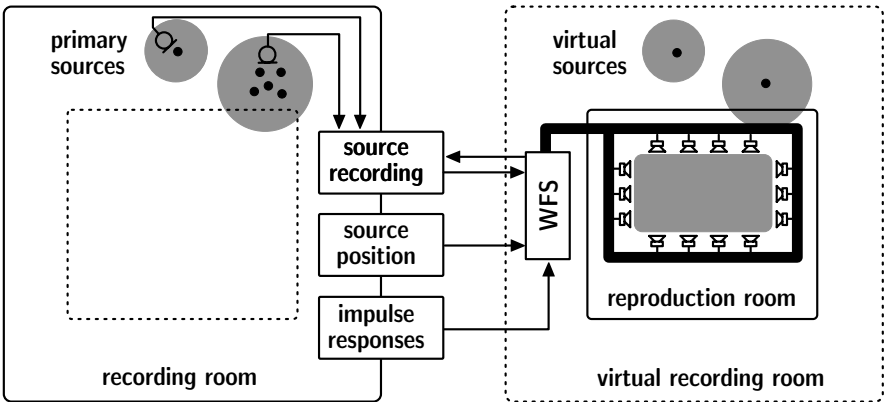

Fig. 16.16. Recording formats for WFS (after (Spors et al. 2004))

Fig. 16.17. CAVE-like environment with WFS (courtesy of the Fraunhofer Institute of Digital Media Technology (IDMT) and Technical University Ilmenau, Engineering Design Group)

For a room acoustical simulation, the room impulse responses related to all sources must be calculated for all reference signals (microphone array) from which the spatial information 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. 11.6, but excluding binaural filters.

16.2.4 Binaural loudspeaker technology

We now go from 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 given in the expression "binaural." The two ears must be fed with binaural signals, as created by binaural synthesis or dummy head recordings. Hence, a left/right stereo setup should serve well. However, as has been common knowledge for many decades, cross talk takes place which interferes with the wanted ear-related signals. The virtual headphone in its function is disturbed by insufficient channel separation.
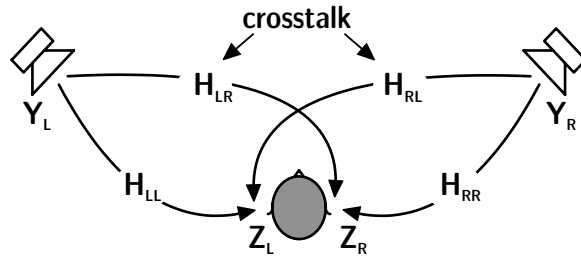
Fig. 16.18. The loudspeaker signals Y_{LL} and Y_{RR} should represent the ear signals Z_{LL} and Z_{RR}. Unwanted cross talk happens over paths H_{LR} and H_{RL}

Without a compensating filter, the aim of localization by binaural synthesis is clearly not successful. In an experiment, test subjects are asked to mark the perceived direction of sound incidence. The diameter of the circles represents the number of responses. Expecting all responses on the diagonal, we observe that the uncertainty of judgment is large, and also we identify front–back confusions near the 90° direction, as indicated by perpendicular groupings of circles.
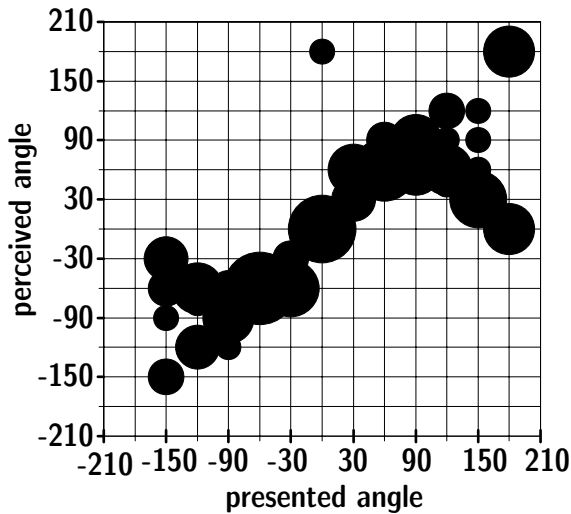

Fig. 16.19. Results of a localization experiment in an anechoic chamber. The diameters of the circles represent the numbers of the perceived direction as responses to the synthesized direction of sound incidence

Cross talk cancellation (CTC)

The required channel separations 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 implemented in this filter, we define the so-called ipsilateral ear oriented toward the direction of sound incidence and the contralateral ear in the shadow region. Cross talk happens to the contralateral ear. For example, the right ear for the right loudspeaker is the ipsilateral ear, and the left ear is contralateral. The cross talk signal component from the right loudspeaker into the left ear can be compensated for by radiating a respective signal from the left loudspeaker and vice versa.

The compensation signal, however, also creates cross talk. We must continue with the same strategy for some interactions. These iterations are stable because over the iterations, the contralateral signals have less and less energy. Five iterative steps are usually sufficient to achieve channel separation of 20 to 25 dB. Simple solutions (low-order iteration), by the way, are implemented in portable radios to create a "wide-function," however suffering from loss in bass performance. The weak bass is an indicator of the quality of the iterations.

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 small, the filter produces ringing effects due to large narrowband peaks in the transfer
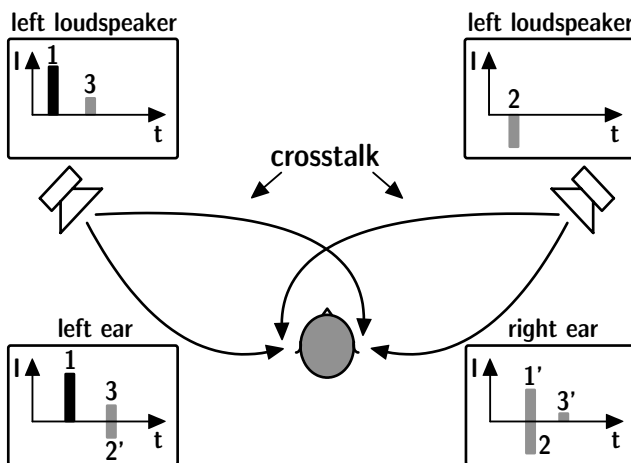

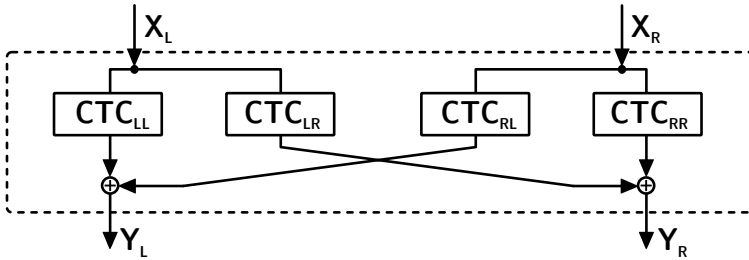**Fig. 16.20.** Iterative cross talk compensation

Fig. 16.21. Cross talk filter design

function. The angular zone which provides a natural damping of the contralateral paths is thus limited and it depends on the geometric situation between the loudspeaker pair and the listener.

Notation according to Fig. 16.18 yields

$$Y_L = \frac{1}{L} \cdot \left[\frac{H_{RR}}{H_{LL} \cdot H_{RR} - H_{LR} \cdot H_{RL}} \cdot X_L - \frac{H_{RL}}{H_{LL} \cdot H_{RR} - H_{LR} \cdot H_{RL}} \cdot X_R \right] \quad (16.2)$$

and Y_R accordingly.

The ear signals X_L and X_R are constructed from four filters as represented in the quotients in Eq. (16.2) and the corresponding equation for the right channel.

Stereo dipole

This technique is similar to the cross talk cancellation system. The loudspeaker arrangement, however, is a left/right dipole source placed in the front of the listener with at most a 10° angle span. Enhanced cross talk damping is achieved by adjusting the null axis of the dipole into the front plane. Both ears receive signals from only one lobe of the dipole. The result of the stereo dipole source is similar to a theoretical monopole and dipole combination.

The stereo dipole system performs very well in symmetric geometry between loudspeaker and listener, and it performs best at close distances, such as in PC displays. More information can be found in (Kirkeby et al. 1998).

Dynamic cross talk cancellation

The compensating method of cross talk cancellation and of the stereo dipole are 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 compensating filters are calibrated
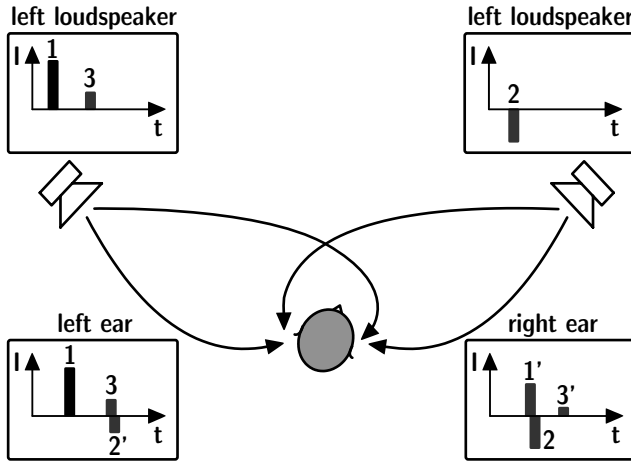
Fig. 16.22. Mismatch by head movement

for one specific position (sweet spot), offside positions of the order of magnitude of 2 cm may lead to audible effects of distorted timbre and localization.

The possible range for these movements is larger for the stereo dipole. For interactivity and free movement by the listener of the order of magnitude of metres, however, all static techniques fail.

Cross talk compensation need not 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 the listener is moving. The solution for getting this performance is adaptive filtering (Gardner 1997).

First, the position of the head must be determined by using a head tracker.

Four-speaker system

Dynamic cross talk 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 stable only within the angle spanned by the loudspeakers. To provide free rotation for the listener, the two-speaker solution is expanded to four channels, two of which run at the same time.

Between the areas, a cross-fade algorithm is implemented using two parallel working CTC systems, which are then partly superimposed. In this way, the current binaural audio signal is filtered with the correct cross talk
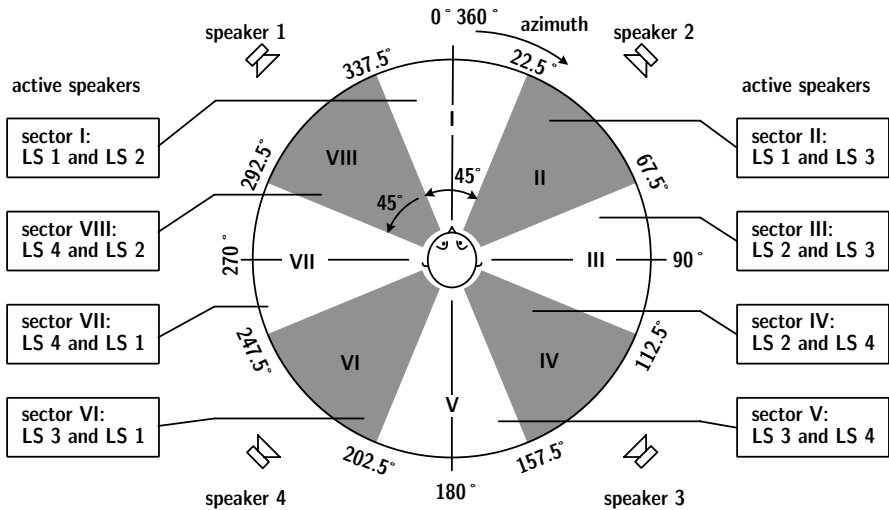
Fig. 16.23. Getting rid of the limitations in listener rotation: Sectors of activation of loudspeaker pairs in a four-speaker dynamic CTC system

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 3-D sound reproduction is illustrated in Fig. 16.25. It consists of auralization software with high precision in the early part (accurate on scales in microseconds) of the impulse response and smooth updates of the late decay (accurateo in scales of milliseconds) which take place only after the listener has moved by some metres.

16.3   VR technology and integrated VR systems

Virtual reality (VR) is a computer-generated environment for interaction in real time. One important specification of VR is multimodality of the human–computer interface. Most VR systems were developed initially for 3-D vision. To obtain presence and immersion of the user, VR is not complete without the modalities of acoustics and haptics (and more). The driving forces for establishing VR applications are task-specific interaction scenarios, their acceptance by the user and user feedback.

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
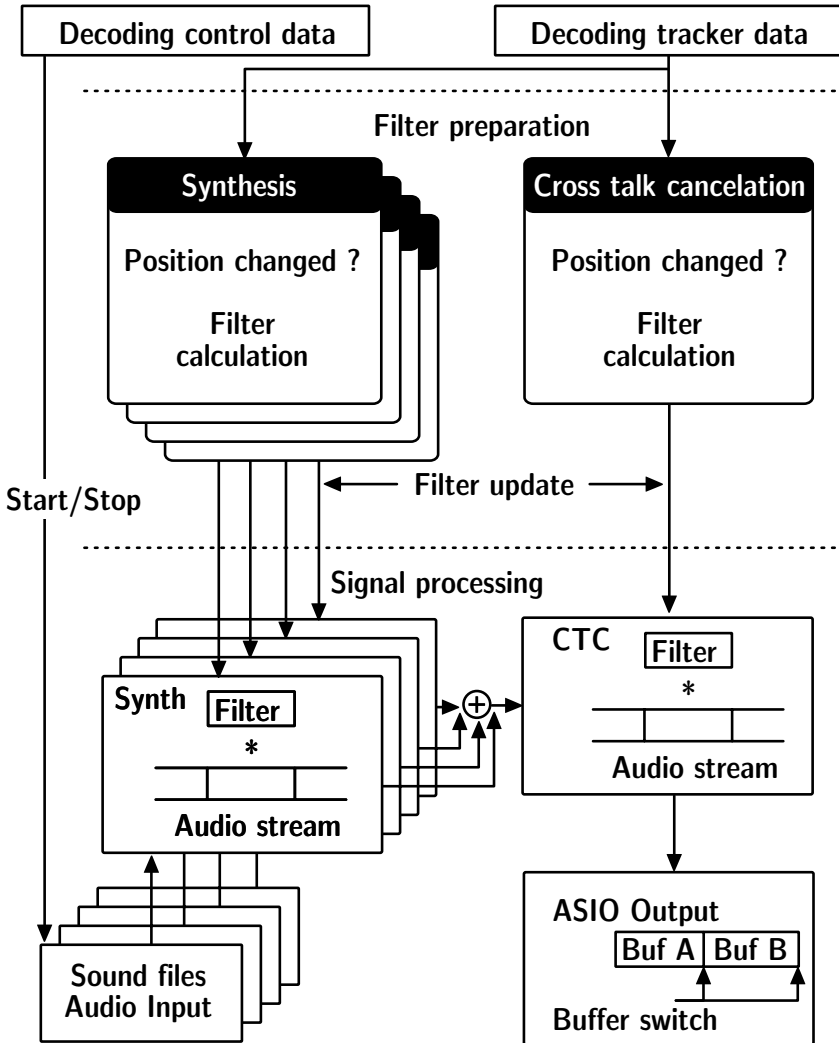
Fig. 16.24. Block diagram of dynamic CTC processing with head-tracker data

other technical object or system can be interpreted as a human–machine interface. Acoustic information may replace visual information displayed for system control or monitoring (Begault 1994).

A so-called "auditory display" such as an "icon" has the purpose of creating attention and information for the visual sense. Therefore acoustic icons are also called "earcons." When presented spatially, a large variety of situations and applications are 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, for example, (Bronkhorst
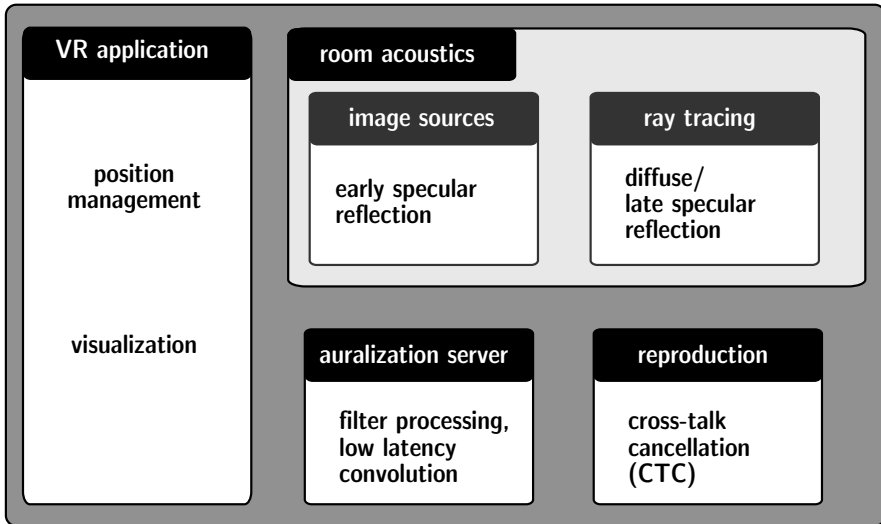
Fig. 16.25. VR System with integrated binaural room acoustics

et al. 1996)) or systems for orientation and navigation for blind persons. Earcons can hence represent information of function.

In the discipline of visual VR, the recent development of graphic processors enabled simulation and reproduction of quite complex scenarios with a high degree of realism. Furthermore, the state of technology for VR subsystems such as motion trackers and projections units has gained very high quality at reasonable cost. As soon as application software is available, it is not unexpected that VR technology will be transported from highly sophisticated laboratory setups to solutions 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, such as in fluid dynamics or molecular physics.

The rapid development of CPU and memory technology will also influence VR. PCs are being equipped with multicore technology and parallel concepts. Ray tracing, image source audibility check or other software components can well be implemented on parallel CPU architecture. Apart from constraints due to data traffic, the gain in simulation speed is almost given by the number of processors involved.

This computational power is required for physical interactive models of haptic, acoustic and visual feedback. In such an environment, the scenes are not simple effects,[91] but are calculated physical reactions in real time.

---

[91] such as in a movie, created once and for one specific intention of effect or feedback.

The main conceptual difference is that no database of a number of fixed reactions is used, such as 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 on memorized action and reaction. The user can operate more intuitively with the virtual scene. Interesting studies can also be performed with modified physical parameters such as objects falling and bouncing in gravitation on the Moon or Mars or alternative behaviour of the physical properties of objects.

To obtain the 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 have necessarily been included, and this is required more and more in large projection systems such as the L-bench or CAVE-like environment. Head-mounted displays were quite common in the 1990s, but today there seems to be a tendency toward releasing the user from wearing too much hardware.

The same holds for training in driving and flight simulators, where the projection unit is combined with a large mechanical platform that includes 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 much as possible.

From the 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. For an 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 (\). Spatial acoustic parameters such as in (OpenAL 2006) are useful for programming acoustic sources.[92] This approach is typically used for creating simple acoustic images for computer games or other fields of consumer electronics which don't have the demand for 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.

---

[92] but not for realistic spatial scenes in rooms

Fig. 16.26. The CAVE-like environment at RWTH Aachen University

Observing these activities, we can expect further progress in research and development of auralization and virtual reality systems, including physical acoustic models and 3-D sound field reproduction. From high-quality technology for acoustic laboratories, some hardware and software solutions will also inspire inventions of products for the consumer market.

Annex

Material data

The basic parameters of acoustic materials are the impedance and the surface shape. The equations given in Sect. 3.1 are applicable as first-order approximations of the impedance and reflection factor of acoustic materials. To discuss more details, other information such as angle-dependent impedance, porosity, tortuosity, etc., is required. These material data include all necessary information required for calculation of the reflected and the transmitted field. In many cases of sound prediction, however, the absorbed or transmitted energy is a sufficient quantity. Thus, for the scope of predicting the exterior sound field for auralization, the absorption coefficient and the scattering coefficient for random incidence are important input data used in room acoustics simulation software. To predict 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 in the project of the "round robin" on room acoustical computer simulations.[93] Other data were obtained from product data given by manufacturers.

The reference methods used for determining 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)[94] and the listed ISO standards.

These data are applicable for geometric or other energetic prediction models such as ray tracing or SEA. They don't have the precision and information required for numerical wave models.

---

[93] http://www.ptb.de/en/org/1/17/173/roundrobin.htm
[94] http://www.ptb.de/en/org/1/17/173/datenbank.htm

Tables of random-incidence absorption coefficients, α

Unless not explicitly specified otherwise, 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	1k	2k	4k	8k
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 a smooth surface	0.01	0.01	0.01	0.02	0.02	0.02	0.02
Limestone walls	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	1k	2k	4k	8k
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	1k	2k	4k	8k
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	1k	2k	4k	8k
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	1k	2k	4k	8k
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	–
6 mm pile carpet bonded to closed-cell foam underlay	0.03	0.09	0.25	0.31	0.33	0.44	0.44

Floor coverings (cont'd)

Material	Octave band frequency in Hz						
	125	250	500	1k	2k	4k	8k
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	1k	2k	4k	8k
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 (2 seats per m^2)

Material	Octave band frequency in Hz						
	125	250	500	1k	2k	4k	8k
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, 2 persons per m^2)

	Octave band frequency in Hz						
	125	250	500	1k	2k	4k	8k
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	1k	2k	4k	8k
Fabric-covered panel, 6 pcf rockwool core	0.46	0.93	1.0	1.0	1.0	1.0	1.0
Fabric-covered panel, 8 pcf rockwool 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 kg/m^2/cm	0.17	0.36	0.66	0.65	0.62	0.68	–
Rockwool 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	086
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	1k	2k	4k	8k
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	1k	2k	4k	8k
Microperforated foil "Microsorber" (Kaefer)	0.06	0.28	0.70	0.68	0.74	0.53	–
Microperforated glass sheets, 5 mm 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, of single objects in m^2

Material	Octave band frequency in Hz						
	125	250	500	1k	2k	4k	8k
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	–
Single person in group, sitting or standing, 1 per 6 m^2 area; typical minimum	0.05	0.10	0.20	0.35	0.50	0.65	–
Single 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
Single 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}

Material	Octave band frequency in Hz						
	125	250	500	1k	2k	4k	8k
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 Sect. 3.3.2) are related to the surface shape and size. Relevant surface parameters are the characteristic length and depth of surface corrugations. It is assumed that the total surface area is large compared with the corrugations and the wavelengths. Note that this concept is not applicable to single scattering objects.[95]

To consider the amount of scattering, it is essential that the shape is described by its characteristic dimensions, the average structural depth, h, and the average structural length, a. The random-incidence scattering coefficients, s, given, are dependent on 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; Jeon et al. 2004; Embrechts et al. 2004) and results from other measurements. More information on diffusers and more detailed data are in given in (Cox and D'Antonio 2004).
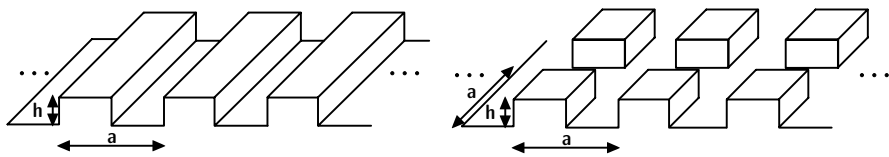
Fig. A.1. Definition of surface dimensions of 1-D and 2-D corrugations

---

[95] See Cox and D'Antonio (2004) for more information on the single-object diffusion coefficient.

2-D 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	–

2-D surfaces (cont'd)

Shape of corrugation	a/λ						
	0.125	0.25	0.5	1	2	4	8
Wooden boxes of various sizes, random pattern, average $h/a = 0.5$	0.05	0.05	0.15	0.4	0.7	0.9	–

	a/λ						
Trapezoidal grating $h/a \approx 0.5$	0.05	0.05	0.1	0.9	0.8	0.9	0.9

1-D 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	–

1-D surfaces (cont'd)

Shape of corrugation	a/λ						
	0.125	0.25	0.5	1	2	4	8
Periodically distributed hemicylinders $h/a = 0.25$	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	–
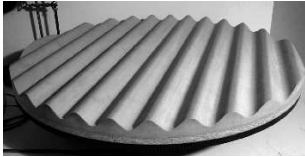
Diffusers

Type	Octave band frequency in Hz						
	125	250	500	1k	2k	4k	8k
RPG "Skyline"	0.01	0.08	0.45	0.82	1.0	–	–
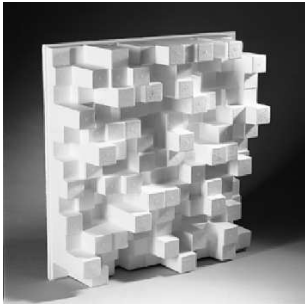
RPG "QRD"	0.06	0.15	0.45	0.95	0.88	0.91	–

(Courtesy of RPG Diffusor Systems, Inc.; www.rpginc.com)

Seating and audience

Shape of corrugation	Octave band frequency in Hz						
	125	250	500	1k	2k	4k	8k
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 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	1k	2k	4k	8k
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	–
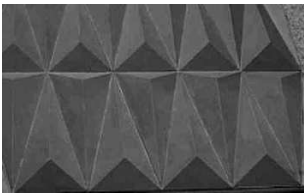
Tables of sound reduction indices, *R*

Masonry

Material	Octave band frequency in Hz						
	125	250	500	1k	2k	4k	8k
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	1k	2k	4k	8k
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	1k	2k	4k	8k
Chipboard. 13 mm, P 3 20 mm, chipboard. 13 mm	19.2	35.0	37.9	3.9	35.6	42.0	38
Veener 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	1k	2k	4k	8k
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

Ahnert W, Feistel R (1991) Binaural Auralization from a Sound System Simulation Programme. Proc. 91th AES Convention, New York, Preprint 3127

Alarcão D, Bento Coelho JL (2003) Lambertian enclosures – a first step towards fast room acoustics simulation. Building Acoustics 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

Atal BS, M.R. 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. Proc. 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 od 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. Proc.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. Proc. Fisita, World Automotive Congress, Yokohama, Japan

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. Acustica united wih Acta Acustica 86, 903

Berkhout AJ (1988) A holographic approach to acoustic control. J. Audio Eng. Soc. 36, 977

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

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

Börger G, Blauert J, Laws P (1977) Sterephone 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. Acustica united with Acta Acustica 84, 943

Bork I (2005a) Report on the 3rd round robin on room acoustical computer simulation, Part I: Measurements. Acta Acustica united with Acustica 91, 740

Bork I (2005b) Report on the 3rd round robin on room acoustical computer simulation, Part II: Calculations. Acta Acustica united with Acustica 91, 753

Braasch J (2005) Modelling of binaural hearing. In: Blauert J (ed) Communication acoustics, Springer Berlin Heidelberg New York.

Bronkhorst AW, Veltman JAH, van Breda L (1996) Application of a three-dimensional auditory display in a flight task. Hum. Fact.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. Proc. 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 Acustica united with Acustica 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, Proceedings of the Royal Society of London A, no. 323, 201

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. Proc. 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, London

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 Acustica united with Acustica 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

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, Proc. 15th ICA, Trondheim, 429

Dalenbäck B-I, Strömberg M (2006) Real time walkthrough auraliation – the first year. Proc. IoA spring conference Copenhagen

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 cross talk 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

EBU – Sound quality assessment material (1988). Recordings for subjective tests. SQAM Compact Disk, Tech. 3253-E

Embrechts JJ, De Geetere L, Vermeir G, Vorländer M, Sakuma T (2004) Calculation of the random-incidence scattering coefficients of a sine-shaped surface. Acta Acustica united with Acustica 92, 593

EN 12354 – Building acoustics: Estimation of acoustic performance of buildings from the performance of elements, 6 Parts.

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. Proc. EURONOISE, Lyon, 55

Farnetani A (2005) Investigation on the acoustics of ancient theatres by means of modern technologies. PhD thesis, University of Ferrara, Italy

Fastl H, Stoll G (1997) Scaling of pitch strength. Hearing Research 1, 293

Fastl H (1998) Pitch strength and frequency discrimination for noise bands or complex tones. In: Psychophysical and Physiological Advances in Hearing, A. R. Palmer et al. Eds, Whurr Publishers, London

Fels J, Buthmann, Vorländer M (2004) Head-related transfer functions of children. Acta Acustica united with Acustica 90, 918

Freiheit R (2005) Historic recording gives choir "alien" feeling: In: Anechoic space, no one can hear you sing – experiment benefits acoustical research and design. Proc. ASA/NOISE-CON 05, Minneapolis, MN

Fruhmann M (2004a) On the pitch strength of bandpass noises. Proc. ICA 2004, Kyoto, Japan

Fruhmann M (2004b) On the pitch strength of harmonic complex tones and comb-filter noises. Proc. 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. Computer Graphics, SIGGRAPH '98, 21

Funkhouser TA, Min P, Carlbom I (1999) Real-time acoustic modelling for distributed virtual environments. ACM Computer Graphics, Proc. SIGGRAPH '99, 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. Proc. 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. Proc. SAE Noise & Vibration Conference, Traverse City

Gerretsen E (1979) Calculation of the sound transmission between dwellings by partitions and flanking structures. Appl. Ac. 12, 413

Gerretsen E (1986) Calculation of airborne and impact sound insulation between dwellings. Appl. Ac. 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 Acustica united with Acustica 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. Proc. 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: Teleoperators & Virtual Environments 16, 157

Hahn JK, Geigel J, Lee JW, Gritz L, Takala T, Mishra S (1995) An Integrated Approach to Motion and Sound. J. of Visualization and Computer Animation 6, 109

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. Acustica united with Acta Acustica 88, 303

Hammershøi D, Møller H (2005) Binaural technique – basic methods for recordung, synthesis and reproduction. In: Blauert J (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

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. Ac. 38, 145

Hodgson M (1983) Theoretical and physical models as tools for the study of factory sound fields. Doctoral thesis, University of Southampton, U.K.

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. Proc. Int. Conf. Auditory Display (ICAD '96), Palo Alto, California, 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

Lake_Huron.htm (2005) http://www.lake.com.au. Version 2005

ISO 140 – Acoustics – Measurement of sound insulation of buildings and of building elements (various parts)

ISO 17497 – 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)

ISO 354 – Acoustics – Measurement of sound absorption in a reverberation room (ISO 354:2003)

ISO 9613 – 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. Proc. ICCVG '04, Warsaw, Poland

Jeon JY, Zhu L, Yoo K (2003) Miral Concert Hall: "Ceramic Palace" for sound scattering. Proc. Internoise Jeju, 545

Jeon JY, Lee SC, Vorländer M (2004) Development of scattering surfaces for concert halls. Appl. Ac. 65, 341

Jeon JY, Ryu JK, Jeong JH (2006) Review of the impact ball in evaluating floor impact sound. Acta Acustica united with Acustica 92, 777

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

Kirkeby O, Nelson PA (1999) Digital filter design for inversion problems in sound reproduction. J. Audio Eng. Soc. 47, 583

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

Kob M (2002) Physical modelling of the singing voice. Doctoral thesis, RWTH Aachen University, Germany

Kohlrausch AG, van de Par SLJDE, van Eijk RLJ, 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

Krokstad A, Strøm S, Sørsdal S (1968) Calculating the acoustical room response by the use of a ray-tracing technique. J. Sound Vib. 8, 118

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. Ac. 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 (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 (2000) Room acoustics, 4th edn. E&FN Spon, London

Kuttruff H (2007) Acoustics – an introduction, Taylor & Francis, London

Kuttruff H, Thiele R (1954) Über die Frequenzabhängigkeit des Schalldrucks in Räumen. Acustica 4, 614

LaViola JJ (2003) A testbed for studying and choosing predictive tracking algorithms in virtual environments. Prof. 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. Proc. IEEE workshop on application of signal processing to audio & acoustics, New Paltz, NY

Lehnert H, Blauert J (1992) Principles of binaural room simulation. Appl. Ac. 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. Proc. 117th Convention Audio Eng. Soc., San Francisco

Lentz T, Schröder D, Vorländer M, Assenmacher I (2007) Virtual reality system with integrated sound field simulation and reproduction. In: EURASIP Journal on Applied Signal Processing, Special Issue on Spatial Sound and Virtual Acoustics.

Lewers T (1993) A combined beam tracing and radiant exchange computer model of room acoustics. Appl. Ac. 38, 161.

Lievens M, J. Brunskog J (2007) Model of a person walking as a structure borne sound source. Proc. 19th ICA Madrid.

Lindau A, Hohn T, Weinzierl S (2007) Binaural resynthesis for comparative studies of acoustical environments. Proc. 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. 7. Auflage, Springer Berlin

Lyamshev LM (1959) A question in connection with the principle of reciprocity in acoustics. Soviet Physics Doklady 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. Ac. 1, 157

Magalhães MBS, Tenenbaum RA (2004) Sound sources reconstruction techniques: A review of their evolution and new trends. Acta Acustica united with Acustica 90, 199

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

McGrath D (1996) Real-time auralization with the Huron digital audio convolution workstation. J. Acoust. Soc. Am. 100, 2579

Mechel FP (ed) et al. (2002) Formulas of acoustics. Springer, Berlin Heidelberg New York

Mechel FP (2002) Improved mirror source method in roomacoustics. J. Sound Vib. 256, 873

Meesawat K, Hammershøi D (2003) The time when the reverberant tail in binaural room impulse response begins. 115th AES Convention, New York, Preprint 5859

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. 117th AES Convention, San Francisco, Preprint 6275

Möser M (2004) Engineering acoustics – an introduction to noise control. Springer, Berlin Heidelberg New York

Møller H (1992) Fundamentals of binaural technology. Appl. Ac.. 36, 171

Mommertz E (1996) Untersuchung akustischer Wandeigenschaften und Modellierung der Schallrückwürfe in der binauralen Raumsimulation, Doctoral thesis RWTH Aachen University, Germany

Morse PM, Ingard KU (1968) Theoretical acoustics. McGraw-Hill, New York

Mourjopoulos J (1994) Digital equalization of room acoustics. J. Audio Eng. Soc. 42, 884

Naef M, Staadt O, Gross M (2002) Spatialized. In: Proceedings of the ACM symposium on virtual reality software and technology. Hong Kong, China, 65

Naylor GM (1993) ODEON – another hybrid room acoustical model. Appl. Ac. 38, 131.

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

Otondo F, Rindel JH (2005) A new method for the radiation representation of musical instruments in auralizations. Acta Acustica united with Acustica 91, 902

Papoulis A (1981) Signal analysis. McGraw-Hill, New York

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

Petersson BAT, Gibbs BM (2000) Towards a structure-borne sound source characterization. Appl. Ac. 61, 325

Pösselt C, Schroeter J, Opitz H, Divenyi P, Blauert J (1986) Generation of binaural signals and home entertainment. Proc. 12th ICA, Toronto

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. Ac. 7, 243

Rindel JH (1993) Modelling the angle-dependent pressure reflection factor. Appl. Ac. 38, 223

Rindel JH, Christensen CL (2003) Room acoustic simulation and auralization – how close can we get to the real room? Proc. WESPAC8, Melbourne, Australia

Rindel JH, Otondo F, Christensen F (2004) Sound source representation for auralization. Proc. international symposium on room acoustics – design and science, Awaji, Japan

Rindel JH (2006) Auralization of airborne sound insulation including the influence of source room. Proc. EURONOISE '07, Tampere. Finland

Römer B (2004) Simulation der Schallabstrahlung von Kfz-Motoren (Simulation of sound radiation from combustion engines). Diploma (MSc) thesis RWTH Aachen University, Germany

Sakuma T, Yasuda Y (2002) Fast multipole boundary element method for large-scale steady-state sound field analysis. Part I: Setup and validation. Acta Acustica united with Acustica 88, 513

Savioja L (1999) Modelling techniques for virtual acoustics. Doctoral thesis, Helsinki University of Technology, Finland

Savioja L, Välimäki V (2000) Reducing the dispersion error in the digital waveguide mesh using interpolation and frequency-warping techniques. IEEE Trans. Speech and Audio Processing 8, 184

Savioja L, Huopaniemi J, Lokki T, Väänänen R (1999) Creating interactive virtual acoustic environments. J. Audio Eng. Soc. 49, 9, 675

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. Proc. DAGA '01, Hamburg-Harburg, 24

Schmitz A (1994) Naturgetreue Wiedergabe kopfbezogener Schallaufnahmen über zwei Lautsprecher mit Hilfe eines Übersprechkompensators. Doctoral thesis, RWTH Aachen University, Germany

Scholl W (2001) Impact sound insulation: The standard tapping machine shall learn to walk! Building Acoustics 8, 245

Schröder D (2004) Integration of real-time room acoustical simulations in VR environments. Diploma 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. Proc. 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. Proc. 4th ICA, Copenhagen, M21

Sellerbeck P (2003) Untersuchungen zur reziproken Transferpfadanalyse im Fahrzeug mit Hilfe eines binauralen Schallsenders. Diploma (MSc thesis) RWTH Aachen University and HEAD acoustics, Aachen, Germany

Shaw EAG (1982) External ear response and sound localization. In: W. Gatehouse (ed) Localization of sound: Theory and applications. Amphora, Groton CT, 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. Proc. EURONOISE Naples

Sottek R (2004) Design of vehicle interior sound using a noise synthesis technology. Proc. 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. Proc. SAE noise & vibration conference, Detroit, paper 2007-01-2226

Sottek R, Riemann D, Sellerbeck P (2004) Virtual binaural auralisation of vehicle interior sounds. Proc. 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. FVV 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. Proc. 5th ICA Vol II, Liége, 313

Spors S, Teutsch H, Kuntz A, Rabenstein R (2004) Sound field synthesis. In: Y.Huang, and J.Benesty (eds) Audio signal processing for next-generation multimedia communication systems. Springer US

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. Acustica united with Acta Acustica 82, 517

Stephenson UM (2004a) Derivation of the reduction of computation time by the voxel crossing technique. Proc. 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. Ac. 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. Proc. DAGA '07, Stuttgart

Storms R (1995) Npsnet-3d sound server: An effective use of the auditory channel, Naval Postgraduate School, Monterey CA, MSc thesis

Suh JS, Nelson PA (1999) Measurement of transient response of rooms and comparison with geometrical acoustic models. J. Acoust. Soc. Am 105, 2304

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. Proc. 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. Acustica united with Acta Acustica 86, 919

Torres JCB, Petragli MR, Tenenbaum RA (2004) An efficient wavelet-based HRTF model for auralization. Acta Acustica united with Acustica 90, 108

Tsingos N, Funkhouser T, Ngan A, Carlbom I (2001) Modelling acoustics in virtual environments using the uniform theory of diffraction. Proc. SIGGRAPH, Computer Graphics

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. Proc. IEEE int. symp. on circuits and systems, Chicago, Illinois, 355

van Maercke D (1986) Simulation of sound fields in time and frequency domain using a geometrical model. Proc. 12th ICA Toronto, E11-7

VDI 3760 – VDI directive (1996) Berechnung und Messung der Schallausbreitung in Arbeitsräumen (Computation and measurement of sound propagation in workrooms) Association of German engineers

Vian J-P, van Maercke D (1986) Calculation of the room impulse response using a ray-tracing method. Proc. ICA symposium on acoustics and theatre planning for the performing arts, Vancouver, Canada, 74

Vian J-P, Martin J (1992) Binaural room acoustics simulation: Practical uses and applications. Appl. Ac. 36, 293

Vigeant MC, Wang LM, Rindel JH (2007) Investigations of multi-channel auralization technique for solo instruments and orchestra. Proc. 19th ICA, Madrid 2007.

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. Proc ICA Trondheim, 689

Vorländer M (2000) A fast room acoustical simulation algorithm based on the free path distribution. J. Sound Vib. 232, 129

Vorländer M, Mommertz E (2000) Definition and measurement of random-incidence scattering coefficients. Appl. Ac. 60, 187

Vorländer M, Thaden R (2000) Auralisation of airborne sound insulation in buildings. Acustica united with Acta acustica 86, 70

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

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. Academic Press, London

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 Acustica united with Acustica 89, 28

Zienkiewicz OC (1977) The finite element method, 3rd edn. McGraw-Hill, London

Zwicker E, Fastl H (1999) Psychoacoustics, fact and models, 2nd ed. Springer, Berlin-Heidelberg-New York

Index