Graz University of Technology

Auralization in Room Acoustics

Bachelor's Thesis
by

Marius Förster

Graz University of Technology
Institute of Broadband Communications

Head: Univ.-Prof. Dipl.-Ing. Dr.techn. Gernot Kubin

Advisor: Dipl.-Ing. Dr.techn. Franz Graf

Graz, July 30, 2008

Contents

Chapter 1

Introduction

The aim of auralization in room acoustics is to make the acoustics of a room audible. That is to say the result of an auralization is an aural impression of a source playing inside an existing or non-existing room. Kleiner et al. defined the term auralization as follows [Kleiner, 1993]: ”Auralization is the process of rendering audible, by physical or mathematical modeling, the sound field of a source in a space, in such a way as to simulate the binaural listening experience at a given position in the modeled space.” The motivation is obvious: instead of describing the acoustic properties of a room by abstract numerical quantities it is possible to directly listen to the sound, which makes the results of room acoustics design more concrete. Another advantage is that people who are no experts in room acoustics can get an impression of different steps of a concert hall. A more general definition of auralization is given by [Vorländer, 2008]: ”Auralization is the technique of creating audible sound files from numerical (simulated, measured, or synthesized) data.”

In the past, scale models (1:5, 1:10, 1:20) were used to predict and evaluate room acoustic characteristics. The frequencies played back in the model were scaled in the same ratio. The first attempts of auralization go back to the 1930s when Spandöck played back frequency-scaled signals in a scale model, picked them up again and replayed them in a binaural fashion [Kleiner, 1993]. As a result of the development of digital computers and the dramatic increase of computational power, scale models were more and more replaced. The computer-based simulation and prediction is much faster and cheaper. Nevertheless, scale models are still being used in parallel due to certain limitations of computer models [Ahnert, 2003]. Computer modeling of room acoustics was first used by [Krokstad, 1968] and [Schroeder, 1973]. Today, different room acoustics simulation software, for example CATT-Acoustic, ODEON and EASE, present a pow-

erful tool for acoustic engineers.

Three components have to be defined and modeled for auralization:

- source
- medium
- receiver.

During the simulation, the source radiates rays according to its directivity. The medium, which in room acoustics is the room, defined by its geometry and surface properties, transmits the rays. Finally, the receiver, modeled by applying so-called head-related transfer functions (HRTFs), keeps track of the hits. Hence, a binaural room impulse response (BRIR) can be calculated. The last step is to convolve the BRIR with an anechoic signal. The result can be listened to via headphones or a cross-talk canceled stereo loudspeaker set. The three components will be discussed in Chapters 2, 3 and 4. In Chapter 5 some aspects of real-time auralization will be reviewed. The content and the structure of Chapters 2, 3, 4 and 5 follow closely that of an excellent book by [Vorländer, 2008]. Finally, an auralization of the Florenintersaal, a concert hall of the University of Music and Dramatic Arts Graz Austria, was carried out by the author (see Chapter 6). This was accomplished by using CATT-Acoustic.

Chapter 2

Representation of sound sources

Characterization and modeling of sound sources is the first component in the auralization process. There are different options to represent the radiation patterns of sources and it has to be distinguished between different types of sources, for example musical instruments, the human voice or even groups of sources. Therefore, the reader gets a general idea of the options and problems concerning the modeling of sound sources.

The modeling and auralization of sources is typically based on anechoic recordings made in the main radiation axis. The directivity is taken into account by fixed radiation patterns defined in octave bands. Such databases are available for the human voice, loudspeakers and musical instruments. However, there are many musical instruments, groups of instruments or noise sources, which require different approaches [Vorländer, 2008]. The technique of making directivity measurements and anechoic recordings is discussed in Section 2.2.

2.1 Examples of sound sources

As already mentioned, not all sources can be treated in the same way. To illustrate this, different types of sources are described in the following.

Loudspeakers. Loudspeakers are time invariant and, due to their small size, well specified in terms of frequency response and radiation patterns. Thus, applicable database material to include their directivity is available [Kleiner, 1993].

Musical instruments. The directivity of musical instruments can be very complex. The radiation of string instruments, for example, can be ascribed to

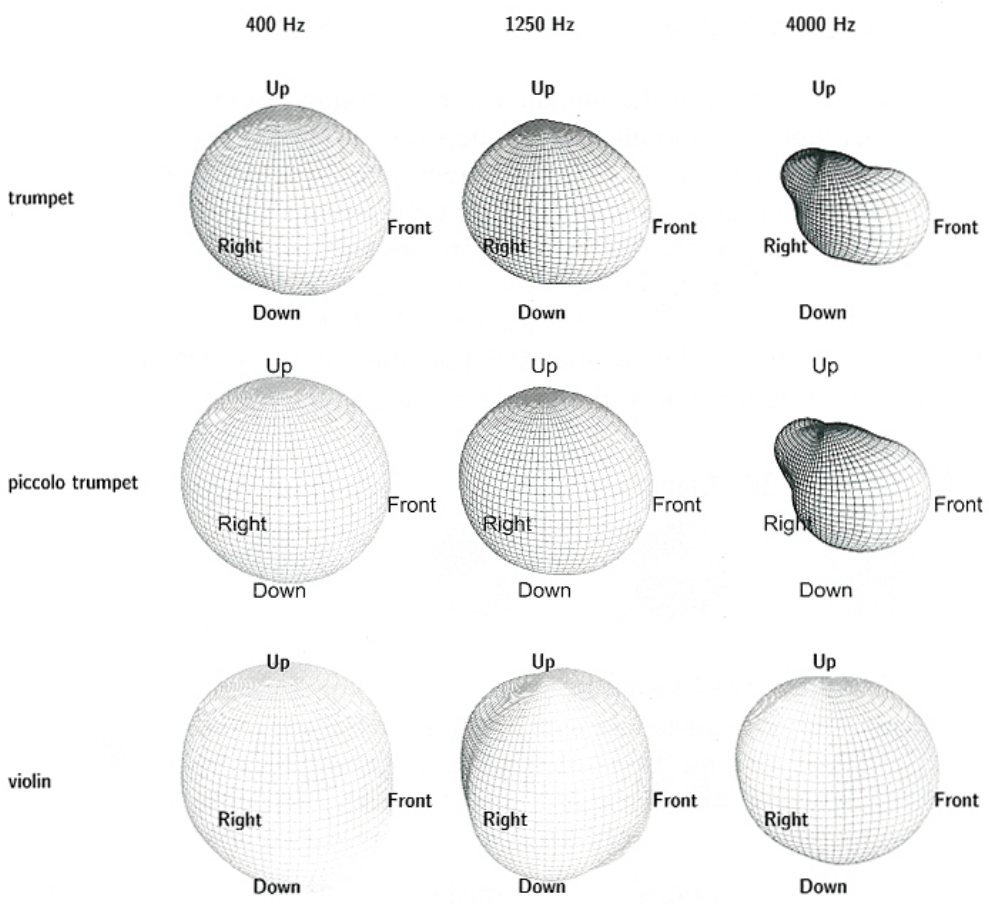

Figure 2.1: Directional characteristics of a trumpet, a piccolo trumpet, and a violin at different frequencies [Vorländer, 2008].

the resonant mode frequencies of the instrument. Each mode frequency has a different radiation pattern, whereas the sound emitted from the strings can be neglected. Another example are woodwind instruments. The sound can be radiated from different parts of the instrument depending on the note played. Furthermore, it should be considered that the player of the instrument influences the radiation because of directional reflections and masking. Figure 2.1 illustrates radiation characteristics of three different instruments [Savioja, 1999].

The human voice. The head and torso of the human body cause directional filtering. However, the directivity of the human voice, for speaking as well as for singing, is not constant over frequency. The reason is that the mouth opening depends on the text which is spoken or sung. Furthermore, it has to

be distinguished between the singing and speaking voice due to the different mouth opening [Vorländer, 2008].

2.2 Source directivity measurements and anechoic recordings

Single channel method. The directivity of a musical instrument can be measured in an anechoic room. A set of microphones is arranged around the musician in the horizontal and vertical plane according to Figure 2.2. During the measurements single notes are played and each note is filtered in octave bands from 125 Hz to 8000 Hz. Afterwards, for each octave band an averaged directivity is calculated including all notes.

In the simulation process, the room impulse response is calculated taking the obtained directivity data into account. The auralization then uses a monophonic recording of the music which is recorded in the main radiation axis (usually the frontal direction) [Rindel, 2004].

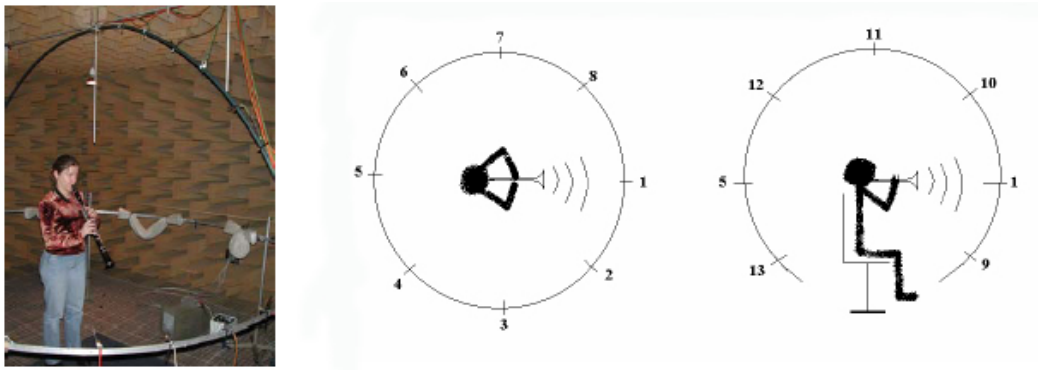

Figure 2.2: *Left:* Directivity measurements in an anechoic chamber. *Right:* Microphone positions in the horizontal and vertical plane [Rindel, 2004].

Multi channel method. Many musical instruments have dynamically changing radiation patterns. I.e. the same frequency may be radiated in different ways depending on the note played. Woodwind instruments, for example, have valves which are opened and closed. One possibility to cover asymmetries of the instrument, movements of the musician and changes in the radiation of different notes, is to perform simultaneous anechoic recordings with a set of microphones arranged around and above the musician.

In the reproduction situation during the simulation, each recorded channel is represented by a particular source, according to the position in the recording. These sources can be defined as neutral and omni-directional within a certain angle. Together, these sources form a compound source following every change of the recording situation (see Figure 2.3) [Rindel, 2004].

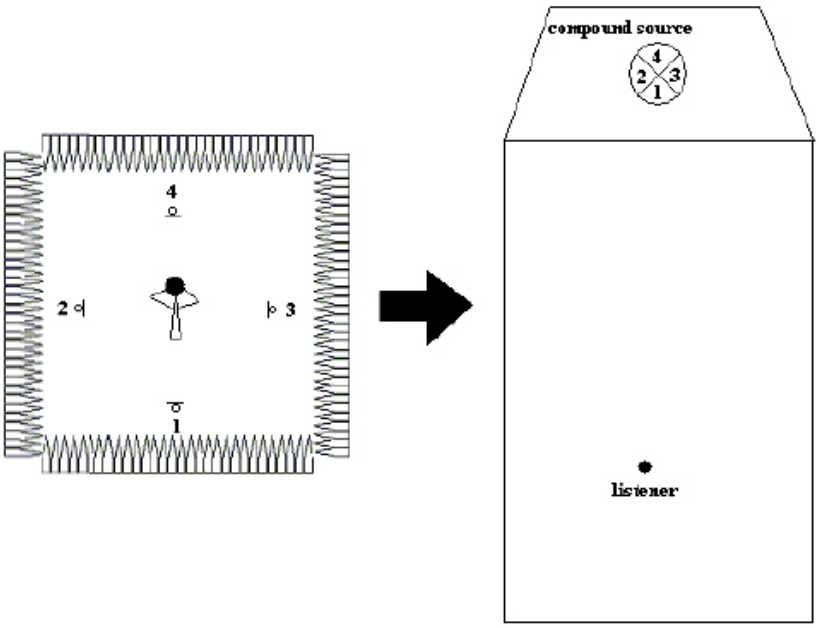

Figure 2.3: *Left:* The multichannel recording method. *Right:* The reproduction in the simulation [Rindel, 2004].

Orchestra recordings. Sometimes stereophonic anechoic recordings of orchestras are available. In the simulation, the two channels can be represented by two source positions, one at the left and one at the right side of the orchestra. The enhancement, in comparison with a monophonic representation, is clearly audible [Rindel, 2004].

Chapter 3

Simulation methods of room acoustics

After having characterized sources, the next step towards auralization is the simulation of the room under study. In most cases, a room can be modeled by means of an LTI-system. It is fully characterized by its impulse response which can be obtained by different methods. The main distinction can be made between computational simulation and scale measurements (for example 1:10, 1:20). The advantage of scale modeling is the natural including of complicated effects, for example diffraction or scattering, in a correct way instead of relying on mathematical approximations. The method can also be used for auralization at predefined measurement positions. This approach can be further divided into direct and indirect acoustic scale-model auralization. In the indirect method, the measurement of the binaural room impulse response and the convolution process are separated, whereas in the direct method the anechoic audio material is directly replayed in the scale model [Kleiner, 1993].

However, scale measurements are increasingly replaced by computer simulations. Commercially available software products have been further developed and consequently became more accurate, more user-friendly and cheaper. After having defined a computer model of the room taking wall materials, source and receiver positions into account, the simulation can be started. Another great benefit of computer simulation is that modifications can be effected easily [Vorländer, 2008].

Computational modeling of sound propagation can be further divided into wave based algorithms, geometrical acoustics and artificial methods (Figure 3.1). Due to computational constraints, not all of these techniques can be used for auralization purposes, especially not in real-time applications. Consequently, this

chapter mainly focuses on the state-of-the-art techniques of calculating the room impulse response (RIR), i.e. geometrical acoustics.

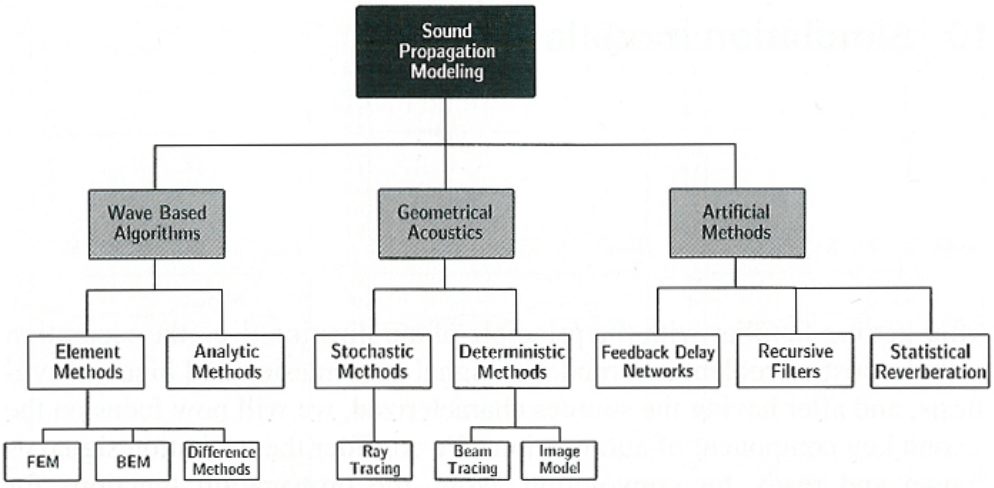

Figure 3.1: Simulation models for sound propagation [Vorländer, 2008].

3.1 Geometrical acoustics

The basic concept of geometrical acoustics is the assumption of sound propagation along straight lines comparable to ray optics. A sound ray is an energy bundle perpendicular to the wavefront. It represents a spherical wave with an infinitely small opening angle $d\Omega$ (see Figure 3.2). The intensity of the ray decreases with $1/r^2$ (r is the distance from the origin). The prerequisite for the validity of this method, however, is that the wavelengths is small compared to the area of the surfaces and large compared to their roughness. These approximations are valid above the Schroeder frequency

$$f_s \approx 2000\sqrt{\frac{T}{V}}\,[Hz] \tag{3.1}$$

due to heavily overlapping resonances. T denotes the reverberation time and V the room volume. In addition, it should be kept in mind that no wave effects, e.g. interference or near-field effects, are taken into consideration. Generally, two different algorithms of implementing geometrical acoustics, ray tracing and the image source method, can be distinguished. Ray tracing is a stochastic method (energy spreading by ray density, energy detection by volumes), whereas the image

source algorithm belongs to the deterministic methods (energy spreading according to distance law, energy detection by points). Both of them have their merits and shortcomings. Hence, a reasonable combination of both algorithms, so-called hybrid models, leads to RIRs very close to measurement results [Ahnert, 2003], [Vorländer, 2008].

Figure 3.2: A sound ray [Vorländer, 2008].

3.1.1 Stochastic ray tracing

In the basic algorithm, the source radiates sound rays at $t = 0$ in various directions. Every time a surface is hit, their initial amount of energy is reduced and their direction is changed. If a detector (mostly spheres) is hit by a ray, the energy, the direction of incidence and the time difference since radiation are detected. The result can be presented as a histogram (see Figure 3.3).

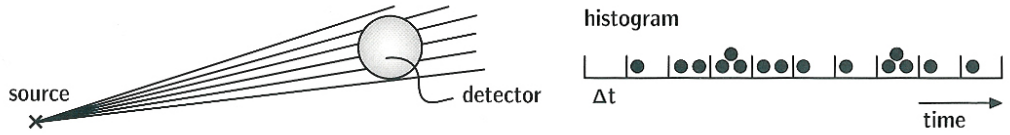

Figure 3.3: *Left:* In stochastic ray tracing energy spreading is implemented by counting. *Right:* The impulse response is created by counting events. An array called a histogram stores the time, angle and energy of each detected ray [Vorländer, 2008].

The directivity of sound sources can be implemented either by defining a frequency and direction-dependent start energy and a uniform radiation pattern or by direction-dependent density, but particles having the same energy. Spherical weighting functions of the ray density or the ray energy can be obtained by measurement (see Chapter 2) or by calculation. Of course it is also possible to create a spherically uniform distribution [Ahnert, 2003], [Vorländer, 2008].

If a ray hits a boundary it looses energy due to absorption. The modeling of absorption can be done in two ways, either by multiplying the incident ray energy with $(1 - \alpha)$ or by stochastic annihilation (α is the absorption coefficient).

The absorption by multiplication is implemented by giving each particle a start energy e_0 and tracing it until a minimum energy e_{min} or until a maximum travel time t_{max} is reached. Absorption by annihilation can be modeled by comparing a random number $z \in (0, 1)$ to the absorption coefficient α. If $z < \alpha$ the particle is annihilated and the next particle is started (see Figure 3.4) [Vorländer, 2008].

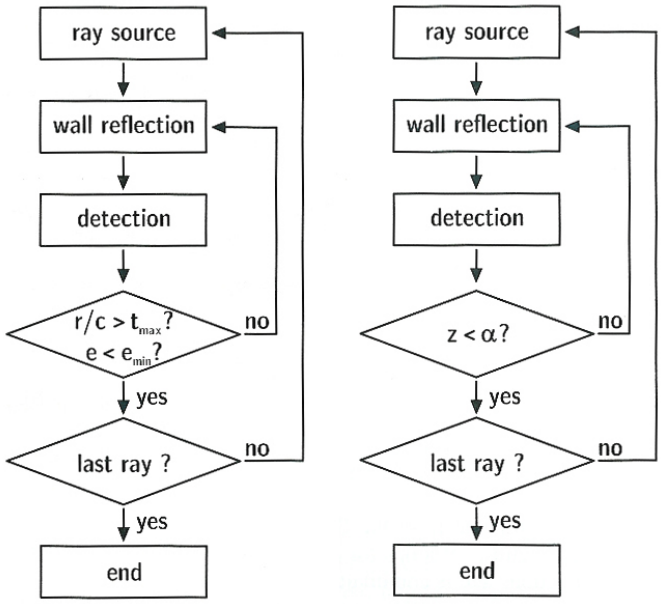

Figure 3.4: *Left:* Flow diagram of ray tracing with absorption by $(1 - \alpha)$-multiplication. *Right:* Flow diagram of ray tracing with absorption by random annihilation [Vorländer, 2008].

Both methods differ in their calculation time and their uncertainties. For absorption by multiplication the typical error of integral parameters such as clarity or strength is given by

$$\sigma_L = 4.34\sqrt{\frac{A}{8\pi N r_d^2}}\,[dB]. \tag{3.2}$$

Herein N is the number of rays launched, A is the equivalent absorption area and r_d is the radius of the detector sphere. For absorption by annihilation the error is given by

$$\sigma_L = 4.34\sqrt{\frac{A}{4\pi N r_d^2}}\,[dB]. \tag{3.3}$$

The calculation time is derived by

$$t_{calc} \approx N\overline{n}t_{max}\tau\,[s] \tag{3.4}$$

for $(1-\alpha)$-multiplication and by

$$t_{calc} = \frac{N\tau}{\overline{\alpha}}\,[s] \tag{3.5}$$

for absorption by annihilation. $\overline{n}$ denotes the mean reflection rate of a ray, $\overline{\alpha}$ the mean absorption coefficient and τ the computation time for one reflection (elementary time) [Vorländer, 2008].

In the ray tracing algorithm, it is also possible to take scattering into account. It occurs when a random number $z \in (0,1)$ exceeds the scattering coefficient δ. Otherwise the incident ray is reflected specularly. This model of switching between the two types of reflections is considered sufficient since the average characteristic is much more important than each individual behavior. The energies of reflections (normalized with respect to the incident plane wave) are given by

$$E_{spec} = (1-\alpha)(1-\delta) \equiv (1-a) \tag{3.6}$$

for the specularly reflected energy,

$$E_{total} = (1-\alpha) \tag{3.7}$$

for the total reflected energy (see Figure 3.5). α denotes the absorption coefficient and a is called the "specular absorption coefficient". From these equations the random-incidence scattering coefficient can be calculated by

$$\delta = \frac{a-\alpha}{1-\alpha} = 1 - \frac{E_{spec}}{E_{total}}\,. \tag{3.8}$$

Scattering is mostly implemented statistically according to Lambert's law. The direction of the scattered sound is independent of the angle of the incident sound ray. The directions of scattering are distributed in a way that they result in a constant emission of energy into all spatial angles. This is obtained by choosing two more random numbers $z_1, z_2 \in (0,1)$. The azimuthal angle Φ is given by

$$\Phi = 2\pi z_2 \tag{3.9}$$

and the polar angle θ is derived by

$$\theta = \arccos\sqrt{z_1} \tag{3.10}$$

[Mechel, 2002], [Schröder, 2007], [Vorländer, 2000], [Vorländer, 2008].

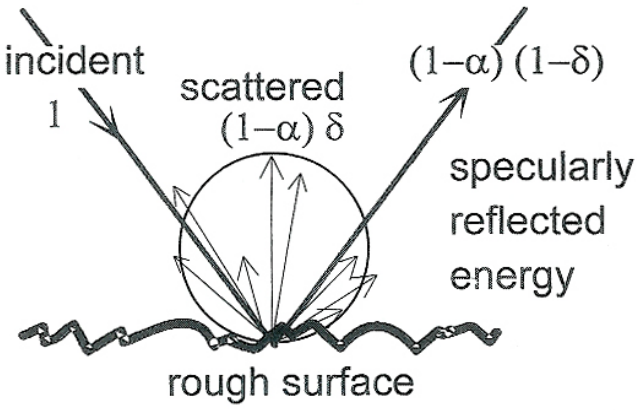

Figure 3.5: On rough surfaces the incident sound is partly specularly reflected and partly scattered [Vorländer, 2000].

3.1.2 Image source method

In the basic image source principle the sound source is mirrored at each plane. The obtained image sources are again mirrored which leads to image sources of second, third, etc. order. As a result the original room can be represented by an infinite pattern. Now, the idea is to replace the tracing of individual sound rays by adding the contributions of all image sources. In the process the energy spreading by distance is included by the $1/r^2$ law. Since this model is strictly deterministic, receivers are points.

Although the reflection factor depends on the angle of incidence, it is assumed to be angle-independent, which is equivalent to a quasi-plane wave. For ideally reflecting surfaces the image source has the same power as the original source, if the surface has an absorption factor $\alpha > 1$ the power is reduced.

The construction of image sources (see Figure 3.6) is accomplished as follows. Given is a plane with the foot point $\vec{A}$ of the wall normal. $\vec{S}$ denotes the source position, $\vec{S_n}$ the position of the mirror image, $\vec{r}$ the vector between the foot point and the source, $\vec{n}$ the wall normal and α the angle between $\vec{n}$ and $\vec{r}$, $\vec{S_n}$ can be calculated by

$$\vec{S_n} = \vec{S} - 2\vec{d} = \vec{S} - 2d\vec{n} \tag{3.11}$$

in which

$$d = \left|\vec{d}\right| = |\vec{r}| \cos\alpha. \tag{3.12}$$

In a uniform manner, image sources of second order are constructed handling image sources of first order as mother sources, and so forth. The procedure is

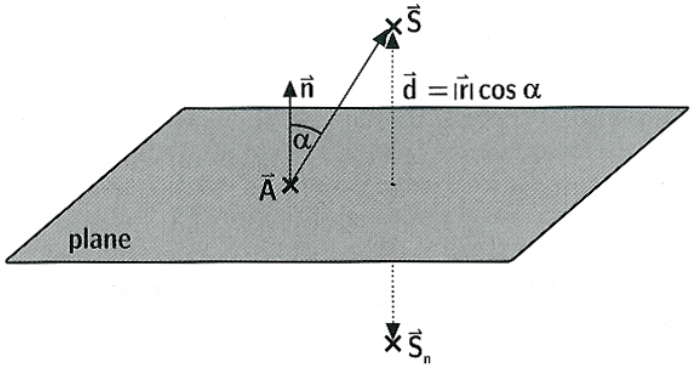

Figure 3.6: Image source construction [Vorländer, 2008].

carried out until a defined maximum time t_{max} is reached.

A large number of image sources can be calculated by this process. However, not all of them are audible at the receiver position. Therefore, a so-called audibility check has to be performed for each image source .

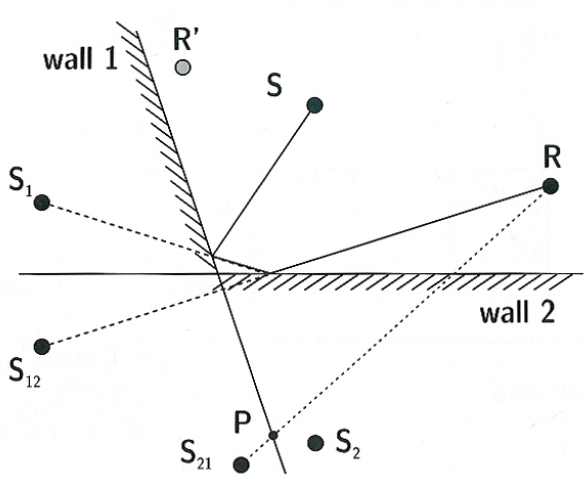

Figure 3.7: Audibility test of image sources [Vorländer, 2008].

All image sources can be seen as the last element of a chain

$$S \rightarrow S_{n_1} \rightarrow S_{n_1 n_2} \rightarrow \ldots \rightarrow S_{n_1 n_2 \ldots n_{i-1}} \rightarrow S_{n_1 n_2 \ldots n_i}. \tag{3.13}$$

The order of the indexes $n_k \neq n_{k\pm1}$ with $n_k \in (1, n_w)$ denote the order of the walls hit, for example S_{12} means that first wall 1 and then wall 2 is hit. The audibility check is performed by tracing back each chain from the receiver to the

source. If the test is positive for each element of the chain the last image source $S_{n_1 n_2 \ldots n_i}$ actually is audible. Figure 3.7 shows an example with a source S, two walls and a receiver R. For example, the image source S_{12} is audible which can be explained as follows. The intersection point of $\overline{RS_{12}}$ with the wall polygon 2 is located inside the polygon. Thus, the test of S_{12} is positive and S_1 has to be inspected. This is done by drawing a straight line between S_1 and the previous intersection point. The result is also positive. In contrast, the image source S_{21} is inaudible at the receiver position R. The reason is that the intersection point P of $\overline{RS_{21}}$ with the wall polygon 1 is located outside the polygon. Remark S_{21} would be audible from the receiver position R'. To find out whether a point is inside or outside a polygon a so-called point-in-polygon test is performed. By the way, the point-in-polygon test is also used during the ray tracing algorithm.

Finally, the amplitudes of all audible image sources are stored in the sound pressure impulse response. Each contribution is delayed by $t_{IS} = r_{IS}/c$ with r_{IS} being the distance between the image source and the receiver.

The classical image source method, however, is limited to low order reflections and consequently only provides good results for simple geometries and short impulse responses. The reason is the rapid increase of possible image sources which is exponential if no strategies for excluding sources are applied. Therefore, a maximum order i has to be chosen which determines the average maximum truncation time

$$t_{max} = \frac{i}{\overline{n}}\,[s]. \tag{3.14}$$

Herein, the mean reflection rate is given by $\overline{n} = \frac{cS}{4V}$ with c denoting the speed of sound, S the surface area of the room and V the volume.

Furthermore, the image source method is limited because of the plane wave assumption, which means that reflection takes place with a constant angle of incidence. In actual fact, sound is reflected at various angles due to spherical waves (see Figure 3.8). Hence, the distances from the source and the receiver to the boundaries should be large in the model in order to obtain reliable results. For short distances ($d \leq \lambda$) and grazing incidence ($\vartheta_0 > 60°$) the systematic errors are considerably audible which sets limits for the simulation of small rooms at low frequencies [Vorländer, 2008].

The computational load of the image source method can be reduced by using preprocessing techniques. One possibility is to subdivide the 3D space into smaller subspaces which has been adopted by [Funkhouser, 1998]. Precomputed and stored spatial data structures encode all possible transmission and reflection paths from all sources in order to determine reverberation paths to a moving listener during the auralization process. Spatial data structures can be implemented

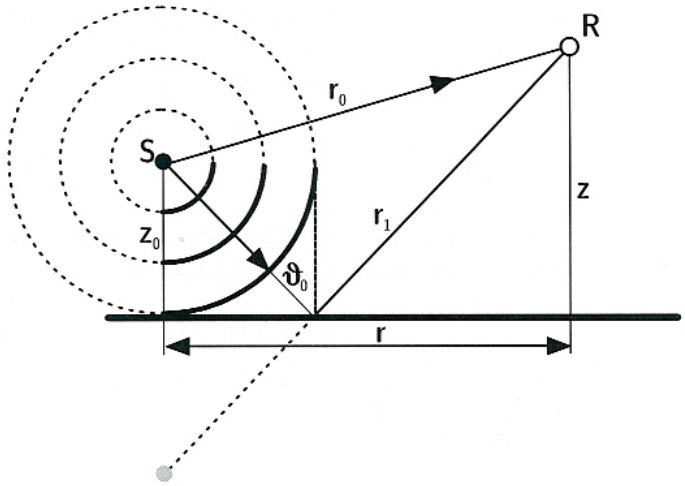

Figure 3.8: A sperical wave reflected at a plane [Vorländer, 2008].

by binary space partitioning (BSP) which is explained by a short example in the following [Schröder, 2006]. The aim is to subdivide the geometry in sets of convex polygons (a set of polygons is convex if the vertices of all polygons are located behind all other polygons) which are then stored in the leafs of a BSP tree (see Figure 3.9). The nodes of the BSP tree contain the particular partitioner dividing the space into subspaces. The minus at the branches means that the partition lies behind the partitioner while the plus stands for in front.

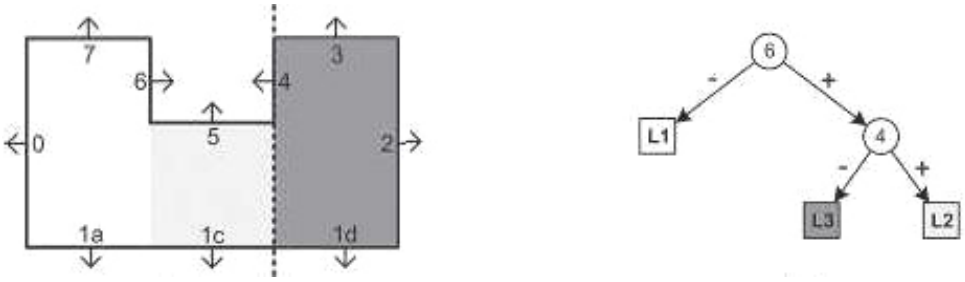

Figure 3.9: *Left:* A room geometry is divided into three partitions. *Right:* The nodes of the BSP tree contains the partitioners, the leafs contain the sets of convex polygons [Schröder, 2006].

After the whole geometry is encoded in the BSP tree, the location of an arbitrarily located point can be determined very fast by testing its position only against a small subset of planes instead of all planes. The test starts at the tree's root node and ends at a leaf representing a subspace which cannot be further subdivided. On

this basis intersections between polygons and line segments, for example between receiver and sender, can be computed efficiently [Schröder, 2006].

3.1.3 Hybrid models

As already mentioned in the introduction of this chapter, stochastic ray tracing as well as the deterministic image source method both have their advantages and disadvantages. The image source method performs much better in the temporal resolution which is crucial for the auralization process. On the other hand, it is possible to include scattering in the ray tracing algorithm. Furthermore, it can be used to speed up the audibility test of image sources. Hence, combining both methods in a reasonable way yields to plausible approximations of the room impulse response of a real room. Such models are summarized by the term hybrid models [Vorländer, 2008].

Hybrid image source models (deterministic ray tracing)

These algorithms perform the audibility test in the forward direction by specular ray tracing. The underlying principle is that a ray hitting a receiver must refer to an audible image source. Still, the contributions to the impulse response are represented by image sources which also means that the detection of sound energy is different from a stochastic ray tracing process (see Figure 3.10). Different dialects of this approach are cone tracing, beam tracing or pyramid tracing.

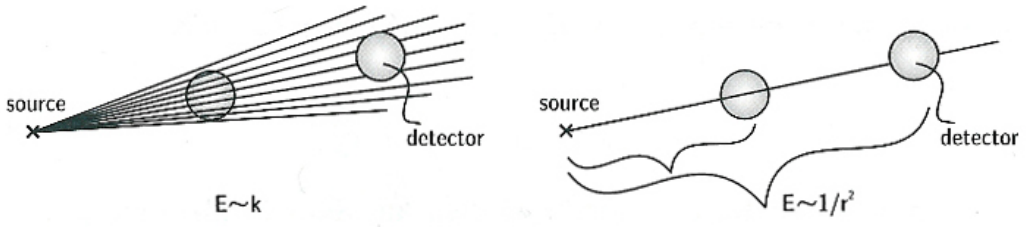

Figure 3.10: Energy spreading and detection in stochastic ray tracing (*left*) and in the deterministic image source method (*right*) [Vorländer, 2008].

The computation time of such algorithms is shorter than in the generic image source method. It can be calculated by

$$t_{calc} = \frac{4c^2 \overline{n} t_{max}^3}{r_d^2} \tau \; [s] \tag{3.15}$$

with c denoting the speed of sound, $\overline{n}$ the mean reflection rate, t_{max} the maximum travel time of a ray, r_d the radius of the receiver and τ the computation time for one reflection [Vorländer, 2008]. In Figure 3.11 the computation times of the classical ray tracing and image source method and the hybrid image source method are compared.

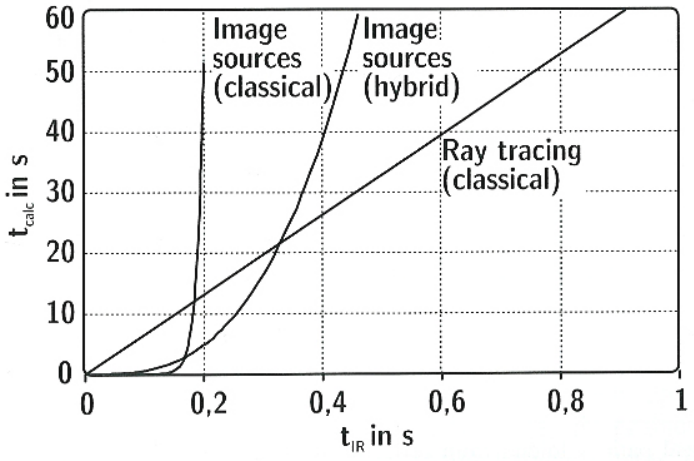

Figure 3.11: A comparison of the computation times of the classical ray tracing and image source methods and the hybrid image source method [Vorländer, 2008].

Hybrid deterministic-stochastic models

Generally, the impulse response of a room can be split up into a first part, containing the direct sound and the early reflections, and a second part, containing the late reverberation. The first part is locally different and its directional and temporal structure is essential. The late reverberation is also very important and gives a general impression of the room but practically no directional information is contained and its fine structure is less important (see Figure 3.12) [Kuttruff, 1993].

In contemporary room acoustical simulation programs, the first part of the room impulse response is usually calculated by some sort of an image source model containing delayed Dirac pulses with according amplitudes. These are sampled with a certain temporal resolution. In the late portion ($> 100\,\text{ms}$) of the impulse response such a high resolution as in the first part is not necessary. Furthermore, a reduction of the computational load is always favorable. This can be achieved by the application of stochastic methods, for example ray tracing [Vorländer, 2008].

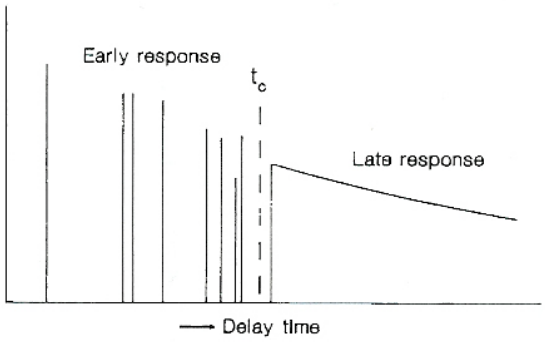

Figure 3.12: Early and late part of the impusle response [Kuttruff, 1993].

Stochastic methods come also into play for the modeling of surface scattering (see Section 3.1) which cannot be done by deterministic methods. The importance of scattering is illustrated by Figure 3.13 which shows the conversion of specularly into diffusely reflected sound energy. In this example 25% of the reflected energy is scattered and 75% is reflected specularly (the absorption coefficent is uniformely 0.2). Additionally, listening tests were carried out by [Torres, 2000] which proved that, depending on the input signal, changes in the diffusion coefficient are clearly audible within a wide frequency range.

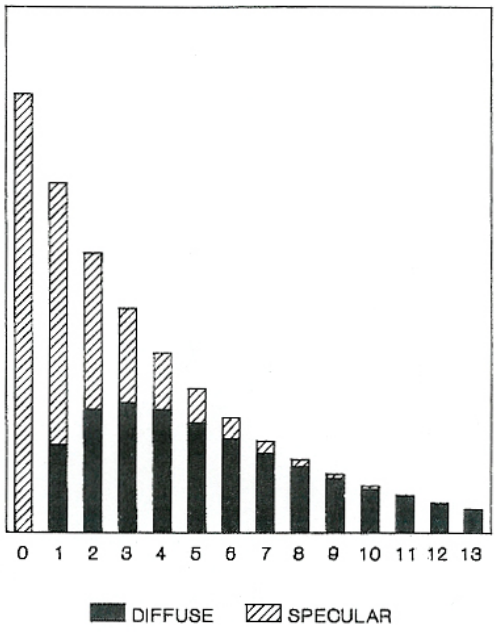

Figure 3.13: Conversion of specularly into diffusely reflected sound energy, illustrated by an example. The numbers denote the order of reflection [Kuttruff, 1995].

The specific implemented algorithms differ and cannot be discussed in detail here. Some well known software solutions are listed in Section 6.4.

3.1.4 The binaural room impulse response (BRIR)

The next step towards auralization in room acoustics is the calculation of the binaural room impulse response (BRIR). It is the basis for binaural reproduction and has to be considered in combination with Section 4.2. The spectrum $\underline{H}_j$ of the jth reflection, respectively the jth image source, is calculated by

$$\underline{H}_j\Big|_{left,right} = \frac{\mathrm{e}^{-j\omega t_j}}{ct_j} \cdot \underline{H}_{source}(\theta,\phi) \cdot \underline{H}_{air} \cdot HRTF(\vartheta,\varphi)\Big|_{left,right} \cdot \prod_{i=1}^{n_j} \underline{R}_i. \quad (3.16)$$

Herein, the delay is denoted by t_j, the phase by $j\omega t_j$, the distance law of a spherical wave by $1/(ct_j)$, the directivity dependent transfer function of the source by $\underline{H}_{source}$, the air attenuation (low pass behaviour) by $\underline{H}_{air}$, the complex geometrical reflection factors of the involved walls by $\underline{R}_i$, and the head-related transfer function of the left/ right ear at a certain orientation of the listener by $HRTF(\vartheta,\varphi)\Big|_{left,right}$. Subsequently, the component $\underline{H}_j$ is transformed into the time domain by an inverse Fourier transformation and then added to the BRIR [Vorländer, 2008].

A HRTF is direction dependent and describes the linear distortions caused by head, pinnae and torso. It is defined as the relation between the sound pressure at the eardrum of a test person and the sound pressure in the middle of the head with the test person absent (more details in Section 4.2.1). HRTFs are available from measurements in a certain spatial resolution. More generally, the HRTF may be replaced by a directivity dependent transfer function for the receiver $\underline{H}_{rec}$ [Ahnert, 2003]. The spectra of the reflection factors in equation (3.16), which are mostly available in frequency bands, can be derived by interpolation.

Equation (3.16) is based on image sources. Since scattered sound dominates the late response and is also essential in the first part it has to be implemented to (3.16). This is achieved by constructing equivalent reflections $\underline{H}_j$. The transition time between the early and the late part is crucial since echoes may occur [Vorländer, 2008].

3.2 Further simulation models

In the following, some other approaches to room acoustical simulation are briefly introduced. One option is given by the wave-based models, namely the finite-element method (FEM) and the boundary-element method (BEM). The principle is to discretize the given geometry by the use of a mesh into small elements which can be of different form, for example rectangular. The size of these elements is determined by the wavelength. The largest possible size can be calculated according to the sampling theorem [Vorländer, 2008]. FEM requires a mesh covering the whole volume, whereas for BEM only the boundary is covered (see Figure 3.14). The results of both methods are complex transfer functions in the frequency domain which can be transformed into impulse response data. However, due to the large number of elements needed, these models are only applicable for small enclosures and low frequencies [Kleiner, 1993].

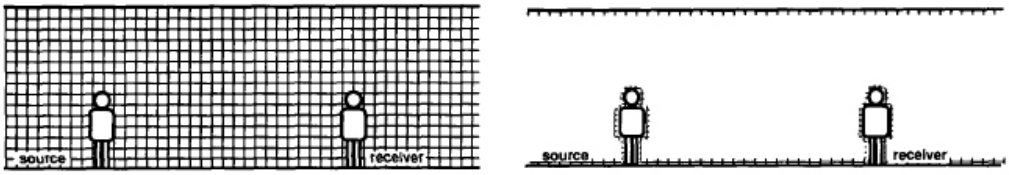

Figure 3.14: *Left:* For finite-element modeling (FEM) the whole room is coverd with a mesh. *Right:* For boundary-element modeling (BEM) only the boundaries (surfaces) of the room are covered [Kleiner, 1993].

Another possibility of room acoustical simulation is given by the finite-difference time-domain (FDTD) methods. A variant is the digital waveguide mesh model. In a three-dimensional room a rectangular grid is constructed in which each node is connected to six neighboring nodes. The result is a digital waveguide mesh containing one-dimensional waveguides connected at each intersection point. A waveguide is a bidirectional delay line. Figure 3.15 illustrates an example showing a two-dimensional slice of the whole grid. With c denoting the speed of sound, N the dimension and Δx the distance between two nodes the update frequency of the mesh can be calculated by

$$f_s = \frac{c\sqrt{N}}{\Delta x}\,[Hz]. \tag{3.17}$$

At the same time f_s is the sampling frequency of the resulting impulse response. For a three-dimensional room and $c = 340\ m/s$ this equation can be approximated by $f_s = \frac{588.9}{\Delta x}$. The accuracy depends mainly on the density of the mesh [Savioja, 1999].

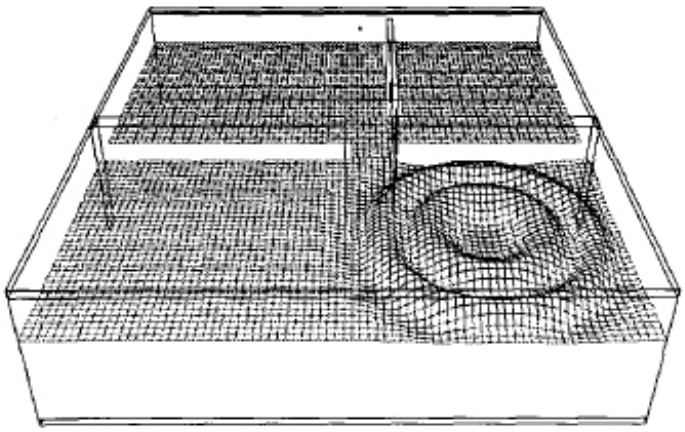

Figure 3.15: An example of FDTD is shown with a horizontal slice of a digital waveguide mesh [Savioja, 1999].

Another time-domain approach to simulation is provided by the radiosity model which is based on the prerequisite of a diffusely scattered sound field. Energy is irradiated and reradiated from surface elements (so-called patches). It is possible to calculate the temporal process of energy transmission to the receiver (room impulse response) if the time is also discretized [Vorländer, 2008].

Chapter 4

Reproduction methods

The third component in auralization is reproduction. In this chapter an overview about different methods suitable for auralization purposes is given, whereas the binaural and transaural technique is described in detail. However, ambisonics and wave field synthesis (WFS) are introduced in a more general manner. In either case, the goal is a high degree of realism and a completely neutral reproduction without any distortions in order to obtain an authentic result corresponding to 3-D perception.

But before that, an intermediate step is briefly reviewed: the ”connection” of a source signal with the simulated room which is done by convolution.

4.1 Convolution

Any source signal which is recorded properly according to Section 2 can be listened to providing the impression as if the sound source would be playing in the real room. For that, it has to be convolved with the RIR derived from simulation (see Figure 4.1). This can be calculated either in time or in frequency domain.

In time domain the discrete convolution is given by

$$g[n] = s[n] * h[n] = \sum_{k=0}^{N-1} s[k]h[n-k] \; . \tag{4.1}$$

Herein, the source signal is denoted by $s[n]$, the impulse response by $h[n]$ and the output signal by $g[n]$. Thus, for a length of N for the input signal and a length of L for the impulse response, $N \cdot L$ floating point multiplications which require

Figure 4.1: A source signal can be convolved with the room impulse response. The output can be listened to [Vorländer, 2008].

the most computational cost have to be performed. The computational load for adding and storing can be neglected.

In the frequency domain FFT convolution is calculated as follows. The temporal input signal and impulse response are stored in vectors of the same length. Then, they are transformed into frequency domain by FFT and multiplicated element-wise. Afterwards, an inverse FFT is processed in order to obtain the output signal in time domain. However, in real-time applications it is not possible to accomplish the whole calculation in one block, especially not for continuous input signals. A solution is provided by so-called segmented convolution. That is, cutting the signal into temporal segments (frames) and completing the calculation frame-wise. The resulting frame is transfered to the output [Vorländer, 2008].

4.2 Binaural reproduction

Binaural reproduction can be achieved directly by headphones or by two loudspeakers with crosstalk cancellation (transaural) (see Section 4.2.3). In both methods, the aim is to reproduce the same signals at the listeners ears as if the listener were in the real room. This is obtained by listener modeling with the aid of head-related transfer functions (HRTF) (corresponding to the according head-related impulse responses (HRIR)). The basic idea is outlined in Figure 4.2 and it is applied to calculate the BRIR (see Section 3.1.4). A mono signal $s(t)$ can be shifted to any direction by convolving it with a pair of HRIRs [Vorländer, 2008]:

$$\begin{aligned} p_{\text{right ear}}(t) &= s(t) * HRIR_{\text{right ear}} \\ p_{\text{left ear}}(t) &= s(t) * HRIR_{\text{left ear}} \, . \end{aligned} \tag{4.2}$$

The concept of HRTFs which is related to binaural hearing is described in Section 4.2.1 in more detail.

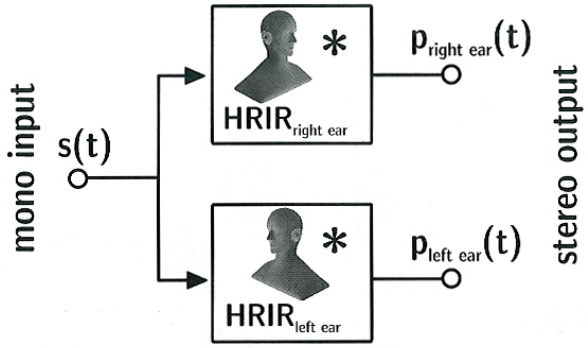

Figure 4.2: Binaural synthesis: a signal can shifted to any direction by convolving it with a set of HRIRs [Vorländer, 2008].

4.2.1 Binaural hearing and HRTFs

At first, a head-related coordinate system is introduced (see Figure 4.3). The azimuthal angle φ between 0° (frontal direction) and 360° degrees (counterclockwise) describes the horizontal plane. The median plane is described by the polar angle θ between 0° (frontal direction) and +90°/ -90° (upper hemisphere/ lower hemisphere).

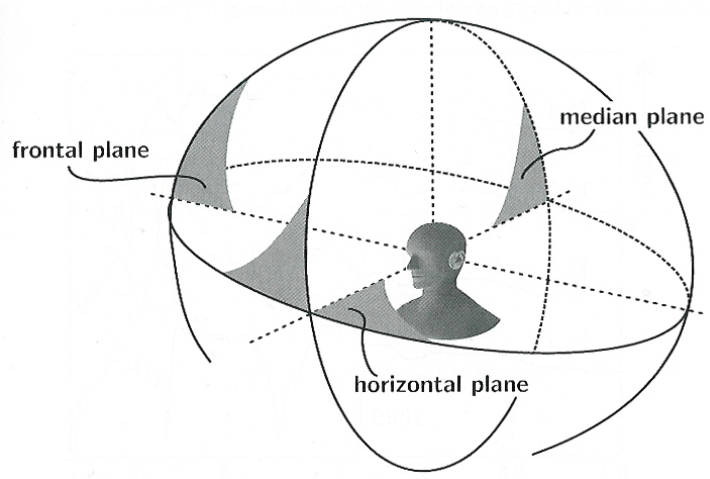

Figure 4.3: A head-related coordinate system [Vorländer, 2008].

Sound can be localized due to spatial hearing. For example a sound wave incident from the right-hand side of a listener has a shorter traveling time to the right ear than to the left one. These direction-dependent differences are called interaural time differences (ITD). Furthermore, the sound wave is frequency- and direction-

dependent reflected, absorbed and diffracted by the head and torso. These differences in amplitude between the two ears are called interaural level differences (ILD). In the horizontal plane the localization is much better than in the median plane. The reason is that in the median plane only monaural cues (spectral cues) can be evaluated, whereas in the horizontal plane also time differences help to localize the sound. Nevertheless, it is possible to hear whether a sound arrives from the frontal or from the back direction [Vorländer, 2008]. In addition, the accuracy in localization depends on the type of signal [Blauert, 1996].

The total linear distortions caused by head, pinna and torso are described formally by the HRTFs which are, by the way, individual since the size and shape of heads differ. HRTFs are defined by the proportion of the sound pressure at the eardrum or the entrance of the ear canal to the sound pressure in the center of the head with the head absent. This relation is of course dependent on the direction of sound incidence. Figure 4.4 shows a HRIR and the corresponding HRTF of sound incident from the left-hand side. In the time-domain it can be seen that the sound arrives at the right ear with a certain delay and a lower amplitude. In the frequency-domain the frequency-dependent damping can be seen.

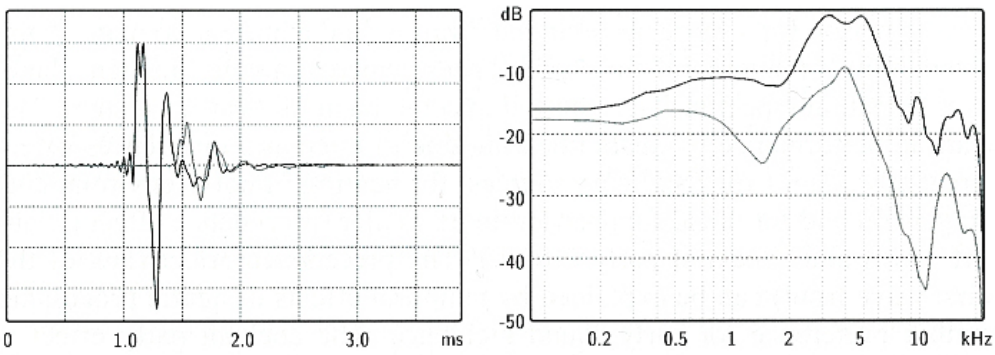

Figure 4.4: *Left:* A set of HRIRs. *Right:* The corresponding HRTFs [Vorländer, 2008].

The measurement of HRTFs can be accomplished with dummy heads (databases can be found on the Internet[1]) or with individual human heads. In the design of a dummy head the challenging task is to create a head which matches best with the average human head. Admittedly, in a listening test with an arbitrary listener, errors in localization might occur when applying data from standard dummy head measurements. Typical problems are the confusion of the front and

[1] http://sound.media.mit.edu/KEMAR.html
http://recherche.ircam.fr/equipes/salles/listen/download.html
http://interface.cipic.ucdavis.edu/CIL_html/CIL_HRTF_database.htm

the back direction and the localization within the median plane. The reason is that the individual anatomic proportions of a human being may differ a lot from those of the dummy head. The best results, however, are achieved by individual measured HRTFs due to individual features in transfer functions, in particular above 6 kHz [Vorländer, 2008]. Localization can be improved by head movements [Mackensen, 2004]. In interactive auralization this can be implemented by head-tracking (see Section 5.1).

4.2.2 Reproduction via headphones

Headphones are widely in use for auralization purposes. Although they are well-qualified, there are also some disadvantages. One problem is the so-called in-head localization which decreases the immersion dramatically. It can be solved with proper headphone equalization, individual measured HRTFs and head-tracking since head movements are essential in sound localization and externalization. Additional effect may be caused by unnatural ear occlusion which affects the transfer impedance of the ear canal and the wearing comfort.

One purpose of the prementioned headphone equalization (see Figure 4.5) concerns the radiation into the listeners ear canal. Since the path through the ear canal is already included in the recording situation it is included twice in the replay situation via headphones. Hence, it must excluded once. Figure 4.6 shows the involved impedances when a headphone is mounted on the ear. $\underline{Z}_{HP}$ denotes the headphone source impedance, $\underline{Z}_{ec}$ the ear canal impedance and $\underline{Z}_{tym}$ the termination impedance of the eardrum. The pressure at the eardrum represents the whole excitation signal (if only the path of sound transmission through the air is considered). If the input signal at the ear canal is well-defined and a model for the ear canal impedance $\underline{Z}_{ec}$ is available, the pressure at the eardrum can be calculated immediately. It should also be noted that the ear canal is individually different.

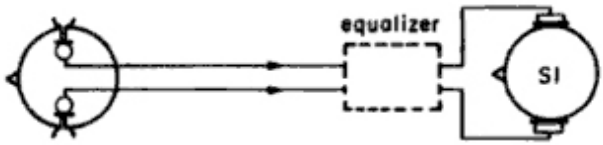

Figure 4.5: Headphone equalization [Blauert, 1996].

Other equalization filters are utilized for compatibility between loudspeaker and headphone reproduction of binaural signals [Vorländer, 2008].

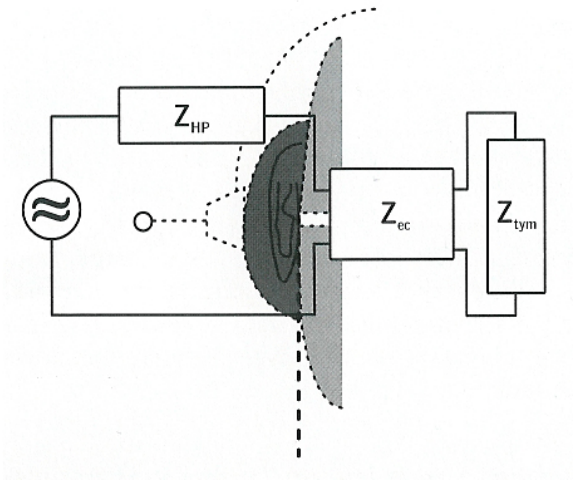

Figure 4.6: The figure illustrates the involved impedance when a headphone is mounted at the ear [Vorländer, 2008].

4.2.3 Crosstalk canceled reproduction via loudspeaker

It is possible to reproduce binaural signals by means of a stereo loudspeaker setup. The binaural loudspeaker setup should act like a virtual headphone but in contrast to a real headphone it lacks sufficient channel separation due to crosstalk. Only the ipsilateral ear[2] should be treated with the signal emitted from the according loudspeaker. Instead, the signal also arrives at the contralateral ear[3]. The effect

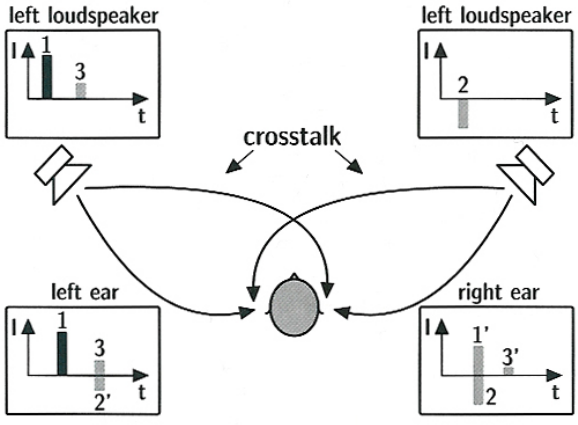

Figure 4.7: The principle of iterative crosstalk compensation [Vorländer, 2008].

[2]The ear oriented towards the direction of sound incidence.
[3]The ear located in the shadow region.

is a drastic decrease of localization accuracy. This problem can be overcome by so-called crosstalk cancellation (CTC). Probably the first method of a crosstalk cancellation filter was presented by [Atal, 1966] as an iterative subtraction filter. For example, assume a signal (1) emitted by the left loudspeaker. It arrives at the left ear with a certain delay (1) but it is also audible to the right ear with a lower amplitude (1'). Hence, for compensation, an accordant signal (2) has to be emitted from the right loudspeaker. However, signal 2 also arrives at the left ear (2') with a lower amplitude due to crosstalk. Consequently, a compensation signal (3) has to be radiated by the left loudspeaker and so forth (see Figure 4.7). Usually, a channel separation of 20 to 25 dB is achieved with five iterations [Vorländer, 2008].

In the following, a closed solution for the iterative CTC process is explained. The crosstalk paths H_{LR} and H_{RL} have to be canceled by the system. According to Figure 4.8 the ear signals Z_L and Z_R are given by

$$\begin{aligned} Z_L &= Y_L \cdot H_{LL} + Y_R \cdot H_{RL} = X_L \\ Z_R &= Y_R \cdot H_{RR} + Y_L \cdot H_{LR} = X_R \,. \end{aligned} \tag{4.3}$$

Solving the system of equations (4.3) for Y_L and Y_R the required CTC-filters are

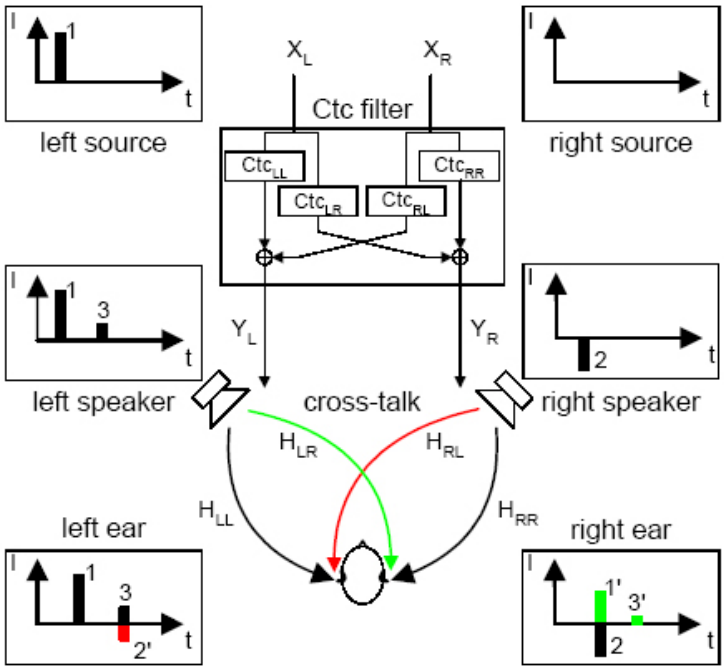

Figure 4.8: A closed solution for crosstalk cancellation [Lentz, 2004].

obtained (labeled with brackets) :

$$\begin{aligned} Y_L &= \underbrace{\frac{H_{RR}}{H_{LL}\cdot H_{RR} - H_{LR}\cdot H_{RL}}}_{CTC_{LL}} \cdot X_L - \underbrace{\frac{H_{RL}}{H_{LL}\cdot H_{RR} - H_{LR}\cdot H_{RL}}}_{CTC_{RL}} \cdot X_R \\ Y_R &= \underbrace{\frac{H_{LL}}{H_{LL}\cdot H_{RR} - H_{LR}\cdot H_{RL}}}_{CTC_{RR}} \cdot X_R - \underbrace{\frac{H_{LR}}{H_{LL}\cdot H_{RR} - H_{LR}\cdot H_{RL}}}_{CTC_{LR}} \cdot X_L \,. \end{aligned} \tag{4.4}$$

This static method, however, is only valid for a specific position (sweet spot). The four transfer functions H_{LL}, H_{LR}, H_{RL} and H_{RR} needed to produce the filters are measured with an artificial head at the specific position [Lentz, 2002], [Lentz, 2004].

Using a head-tracking system (see Section 5.1) and a HRTF database it is possible to turn the head and move within an area of about 1 m^2. Hence, a valid filter set can be calculated for the current position. However, the stability is only given within the angle spanned by the two loudspeakers. If for any frequency the denominator

$$D = H_{LL}\cdot H_{RR} - H_{LR}\cdot H_{RL} \tag{4.5}$$

becomes small, the filter produces ringing effects. A complete rotation of the listener can be provided by a four speaker system (see Figure 4.9) [Lentz, 2004].

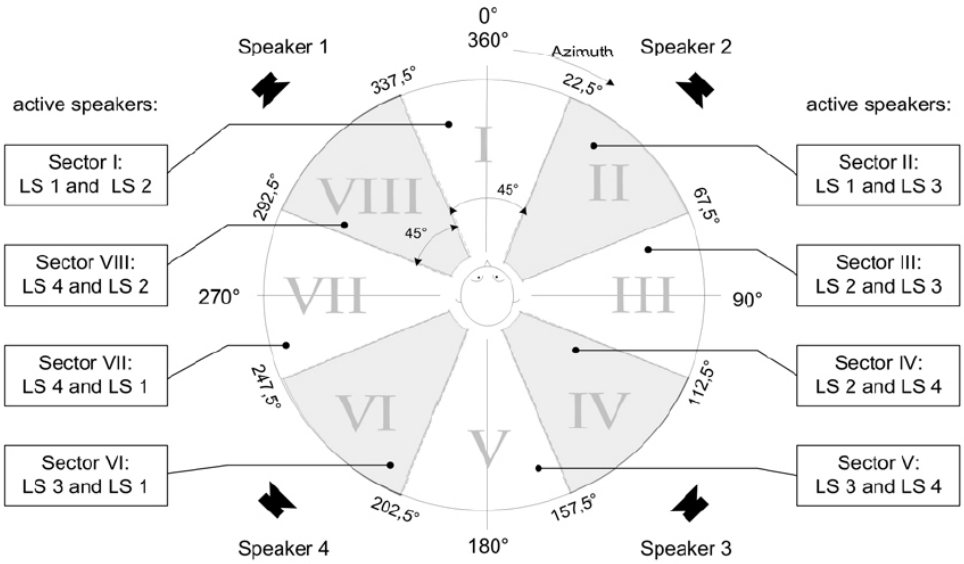

Figure 4.9: Free rotation is provided by a four speaker setup. The system always chooses the appropriate two speakers which run at the same time. Between the areas two CTC filters are superimposed [Lentz, 2004].

4.3 Multichannel reproduction

The binaural reproduction methods via headphones or crosstalk-canceled loudspeakers described in the previous section are one-listener solutions. Head-related signals are used in order to reproduce the correct input signals at the listener's eardrums. In this chapter, multichannel reproduction methods generally suitable for auralization purposes are reviewed. Yet, this is done in a more general sense. The aim of these methods is to produce a sound field in the listening area enabling more listeners to be involved (see Figure 4.10).

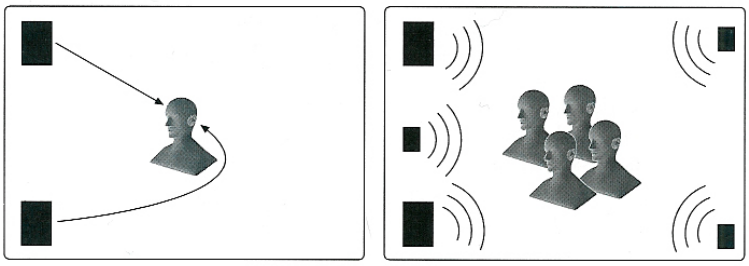

Figure 4.10: *Left:* One-listener solution. *Right:* Solutions for more listeners [Vorländer, 2008].

4.3.1 Ambisonics

Ambisonics[4] has been invented by Michael Gerzon [Gerzon, 1976]. It is a technique for recording and for the replay of spatial sound fields with the aim of creating a true 3D sound image. The information needed for 3D sound field encoding is transmitted by four channels: W, X, Y and Z. This realization is called B-format. The channel X denotes the left-right signal, Y the front-back signal and Z the up-down signal. They are recorded by figure-of-eight microphones and correspond to the particle velocity in the specific room direction. The sound pressure at the recording point is represented by the component W which is recorded by an omnidirectional microphone. In case of horizontal sound field encoding, only three channels are needed: W, X and Y. Recordings can be made with special sound field microphones. In the reproduction situation the signals played by the loudspeakers are constructed by linear combinations of the four channels.

An advantage of Ambisonics is that there is no longer a sweet spot and the

[4] http://www.ambisonic.net/
http://www.york.ac.uk/inst/mustech/3d_audio/ambis2.htm

image can even be perceived when walking outside the main range of the speakers. In addition, there are no fixed loudspeaker positions.

4.3.2 Wave field synthesis (WFS)

Wave field synthesis (WFS) is a method to reproduce a sound field which is completely identical to the real one. An incoming sound field is measured by an array of microphones and reconstructed by an appropriate loudspeaker array. This is based on Huygens' principle: each wave front incident on the microphones can be synthesized by the addition of secondary sources (loudspeakers) located at the same positions. Generally, the positions of the microphone and the loudspeaker array are different. Therefore, numerical extrapolation of the positions has to be applied (see Figure 4.11). The anechoic source signal is recorded separately from the room impulse response. In the replay situation, it is convolved with the RIR which can also be derived from room acoustics simulation software. In any case, the RIR of each source position has to be measured or simulated for each microphone position within the microphone array. The simulation can be done according to Chapter 3. The binaural synthesis has to be excluded since the elementary ear signals emerge naturally [Berkhout, 1988], [Theile, 2003].

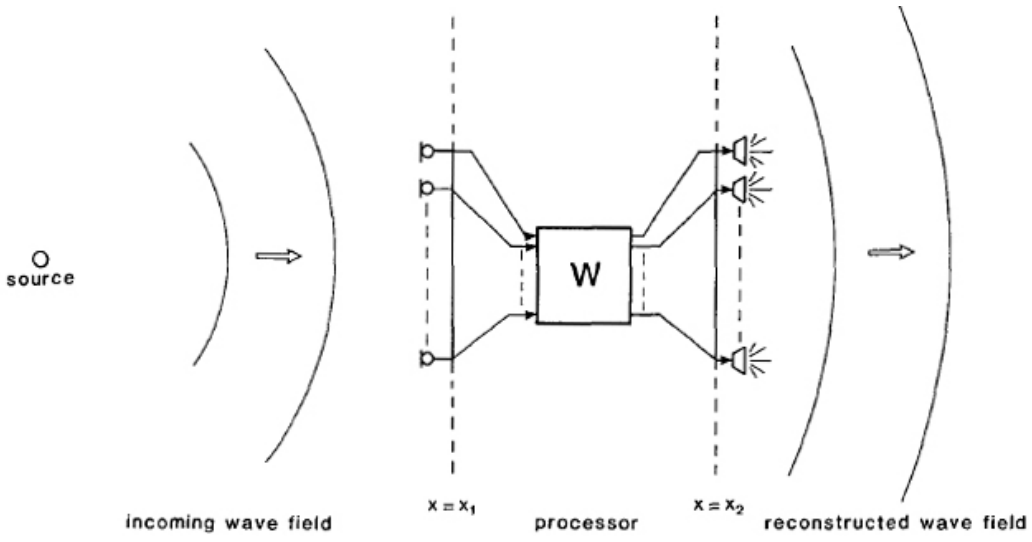

Figure 4.11: Wave field synthesis [Berkhout, 1988].

Chapter 5

Aspects of real-time auralization

Real-time auralization plays an important role, for example, in virtual reality (VR) applications. The interacting user should be immersed in a multimodal way, including visual, haptic and acoustic cues. The visual component is highly developed and of major interest, whereas the acoustic component is not integrated to the same extent. Accordingly, in this chapter, aspects of real-time auralization are overviewed.

A crucial factor is the overall latency which is introduced by hardware elements such as head trackers or interfaces and the time needed for acoustic simulation and binaural synthesis. A total delay of 50 ms is considered acceptable. The time for simulation and synthesis, however, conflicts with the update rate which is the time in which the input is changed, thus, the simulation recalculated and transfered to the output. If the update rate is high which is necessary for smooth transitions of head rotations or between positions, few computational time for the acoustic simulation is available. Update rates of 60 Hz (corresponding to a processing time of 17 ms) are considered adequate. This indicates that, in the end, the level of authenticity is dependent on computer performance [Vorländer, 2008].

5.1 Precalculated dynamic auralization and head tracking

An intermediate step on the way towards complete real-time auralization is presented by [Dalenbäck, 2006]. A certain grid of receiver positions is introduced to the room under study. The local density of which has to be chosen depending on how fast the RIR changes in different parts of the room. Then, the echograms are

preprocessed for each of these positions. In a second step, one B-format (W, X, Y, Z) impulse response set per position is created which is sufficient for free rotation. During the so-called walkthrough (dynamic replay situation), the convolution with anechoic sound is calculated and down-mixed for binaural reproduction. Between the positions, interpolation is necessary (see Figure 5.1).

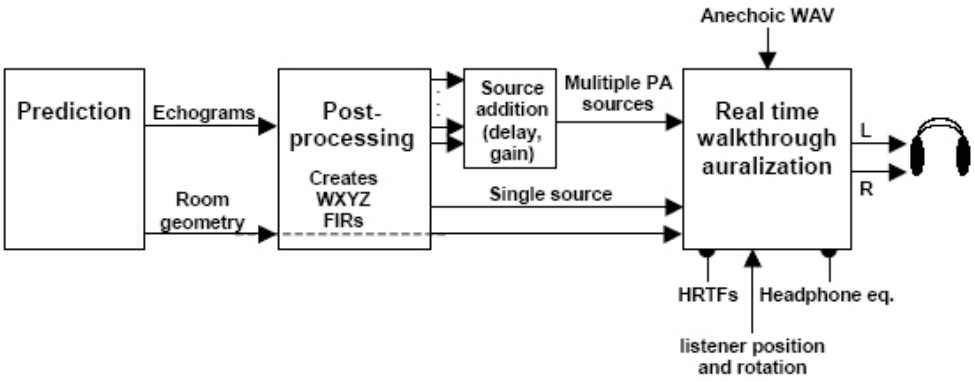

Figure 5.1: Processing for walkthrough [Dalenbäck, 2006].

Head tracking

An important feature of real-time auralization is the so-called head tracking which enables real-time binaural synthesis. A head tracker usually consists of a receiver somehow mounted at the listener and a transmitter located in front or above the head. This enables the determination of the current observer position and orientation relative to the environment. Hence, the BRIR can be synthesized according to equation (3.16), always containing the correct set of HRTFs. Devices for tracking head movements may be based on different technologies: ultrasound, electromagnetic, optical or mechanical. The trend is to get rid of antennas or sensors mounted on the head. One possibility is, for example, eye tracking [Vorländer, 2008].

5.2 Real-time processing of image sources and reverberation

The first part of the BRIR consists of specular reflected and of scattered sound. The specular part is calculated according to Section 3.1.2. However, in real-time processed dynamic auralization the scene and accordingly the impulse response

changes due to interaction. At first, the cloud of image sources is calculated for one fixed position of the source. Hence, all relevant parameters are known including the position, the total time delay, the order and the wall reflections. If the scene changes dynamically it is not necessary to recalculate the positions of the image source from the beginning. Instead, a faster way is to translate or rotate the existing ones which is explained in the following [Vorländer, 2008].

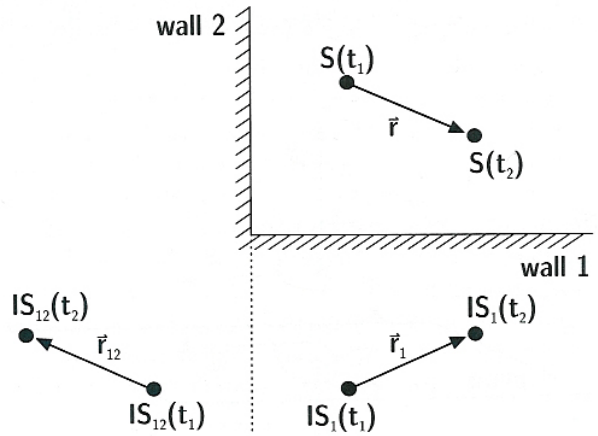

Figure 5.2: Translation of the original source and the corresponding image sources [Vorländer, 2008].

From time t_1 to t_2 the source is translated by the vector $\vec{r}$ (see Figure 5.2). This translation vector has to be mirrored at the corresponding mirroring wall plane in order to obtain the translation vector of the first-order image source. Thus, a matrix has to be found which modifies the direction of $\vec{r}$ in such a way that the first-order image source is moved corresponding to the translation of the original source. The matrix

$$T = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix} \tag{5.1}$$

is obtained from the translation of the original source in the orthonormal system of the mirroring wall plane (see Figure 5.3).

The translation of a second-order image source is derived by mirroring the first-order translation vector. The total shift vector is calculated by

$$T_{total} = \prod T_{walls} \tag{5.2}$$

and the corresponding vector to the new image source by

$$\vec{r}_{i,new} = \vec{r}_i + T_{total} \cdot \Delta\vec{r}_{source} \, . \tag{5.3}$$

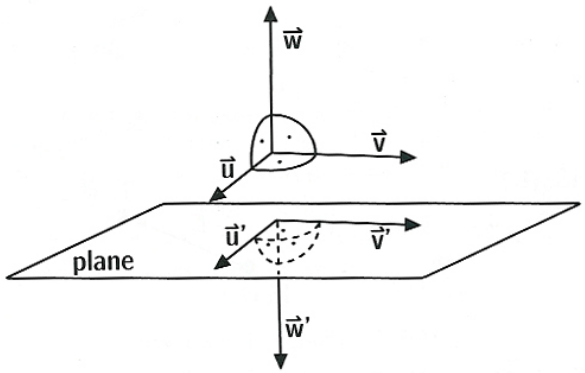

Figure 5.3: Mirrored orthonormal system [Vorländer, 2008].

Still, image source processing has high computational demands and thus the calculation of the impulse response has to be truncated. Late reverberation is added separately. One possibility to do so is by reverberation processors. Anyhow, this is only a rough approximation. Especially for coupled rooms or long and flat rooms a more detailed and physically-based solution is needed.

One possibility is presented by [Schröder, 2007] based on a fast ray tracing algorithm. The result of the ray tracing are histograms for each detection sphere (receiver) and each frequency band containing the angle of incidence, particle's energy, and running time. Figure 5.4 illustrates a detection sphere subdivided into directivity groups which are spatial angle intervals. The energies of the detected particles are assigned to these directional groups in accordance with their angle of incidence. Since the temporal resolution of the histograms is usually lower than the resolution provided by the sampling rate an appropriate fine structure has to be generated by post-processing.

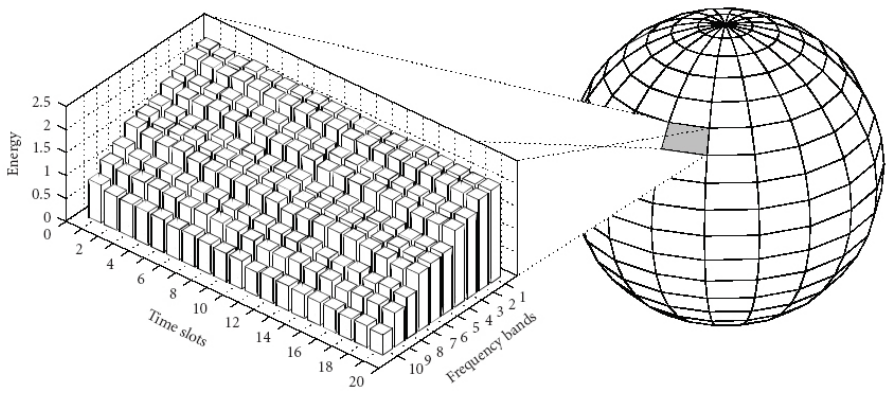

Figure 5.4: A histogram of one directivity group is illustrated [Lentz, 2007].

Chapter 6

Auralization of the Florentinersaal

An auralization of the Florentinersaal in Graz (see Figure 6.1 and 6.2) was realized in different variants (see Table 6.1 and Figure 6.4). The Florentinersaal is a concert hall of the University of Music and Dramatic Arts Graz. The dimensions are: lenght = 15.5 m, width = 7.5 m and hight = 8 m. Hence, the volume amounts to 930 m^3. The auralization was carried out in the binaural post-processing module of CATT-Acoustic. After the simulation and auralization were accomplished, a measurement has been carried out in the Florentinersaal using the binaural recording system "Source" [Graf, 1999] (see Section 6.3). This was done in order to compare the sound of the auralization to the one of the measurement.

Figure 6.1: Florentinersaal (view of the stage)

Figure 6.2: Florentinersaal (view of the audience area)

6.1 Different simulation variants

In the following the different simulation variants are introduced. The post-processing and auditory impressions are reviewed in 6.2. The following settings have been similar to all variants:

- number of rays/ octave = 15000 (more rays than the default of 10000 were chosen since more rays means a longer early part of the IR and thus more specific spatial properties for a specific receiver position)
- 9 receiver positions with the head direction towards the stage; 3 in the front and 6 in the rear audience area (except for walkthrough)
- height of receivers = 1.2 m (except for walkthrough and variant 8)
- the source was electroacoustic with the directivity Catt.SD0; the directivity at 1 kHz is shown in Figure 6.3
- the source was located in the center of the stage directing horizontally into the hall; height of source = 2.30 m = 1.8 m above the stage (except for variant 8)

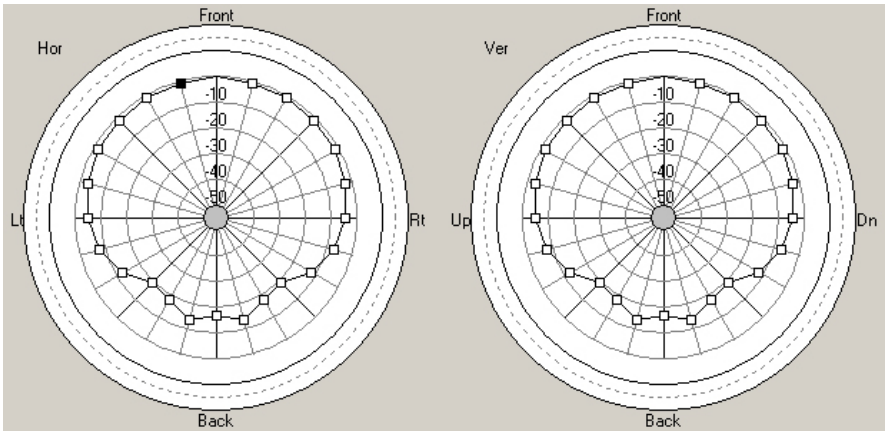

Figure 6.3: Directivity of the electroacoustic source used in the simulation at 1 kHz

Variant 1	with seating, no audience, default scattering with 80 % over all frequencies, T-30 as in [Raumakustikskript]
Variant 2	with seating, no audience, absorption coefficients were tuned until T-30 of [Raumakustikskript] was reached, scattering coefficients were entered manually
Variant 3	like variant 2 but with audience
Variant 4	like variant 2 but without seating and audience
Variant 5	like variant 4 but without curtains
Variant 6	like variant 2 but one side wall is completely reflecting
Variant 7	like variant 2 but walkthrough (different receiver positions)
Variant 8	like variant 2 but hight of loudspeaker = 1.3 m above stage and height of receivers = 1.3 m

Table 6.1: Summary of the different simulation variants

Variant 1

The first aim was to approximate the reverberation time (T-30) of the simulation to the one of the reverberation measurement of 1994 of the Florentinersaal [Raumakustikskript]. Thus, seating was in the hall but no audience (see Figure 6.6). The materials were chosen as realistic as possible. The longest expected reverberation time was 1240 ms at 125 Hz and, hence, the ray truncation time was set to 1300 ms. However, since simulation results for T-30 were still too high, default scattering was enabled with 80 % scattering over all frequency bands (octave bands from 125 Hz to 16 kHz). The average reverberation time of all 9 receivers is shown in Figure 6.5. The maximum difference to the measurement is 150 ms at 125 Hz. It should also be noted that the results of the reverberation measurement

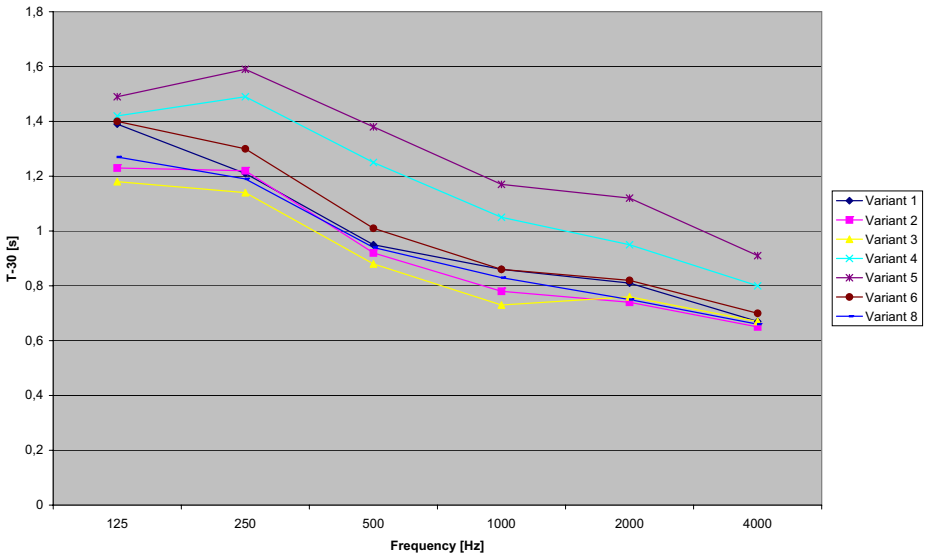

Figure 6.4: T-30 of the different simulation variants

were in third octave bands, whereas the simulation was accomplished in octave bands.

	125	250	500	1k	2k	4k	
EyrT	1,17	1,05	0,82	0,72	0,69	0,61	s
EyrTg	1,16	1,05	0,82	0,72	0,68	0,61	s
SabT	1,20	1,09	0,87	0,77	0,74	0,66	s
T-15	1,38	1,21	0,94	0,83	0,81	0,66	s
T-30	1,39	1,21	0,95	0,86	0,81	0,67	s
Tref	1,20	1,20	1,00	0,80	0,70	0,70	s
AbsC	15,59	17,12	21,37	23,86	24,41	25,32	%
AbsCg	15,71	17,22	21,38	23,93	24,53	25,44	%
MFP	4,95	4,96	4,95	4,96	4,96	4,95	m
Diffs	80,47	80,77	80,85	80,87	80,88	80,80	%

Figure 6.5: *Variant 1:* with seating, no audience; T-30 as in [Raumakustikskript]

Variant 2

The aim was also to approximate the reverberation time of the simulation to the one of the measurement which means that the geometry is the same as in variant 1 (see Figure 6.6). The only difference to variant 1 is that default scattering was disabled and the absorption coefficients of the four walls were tuned until the

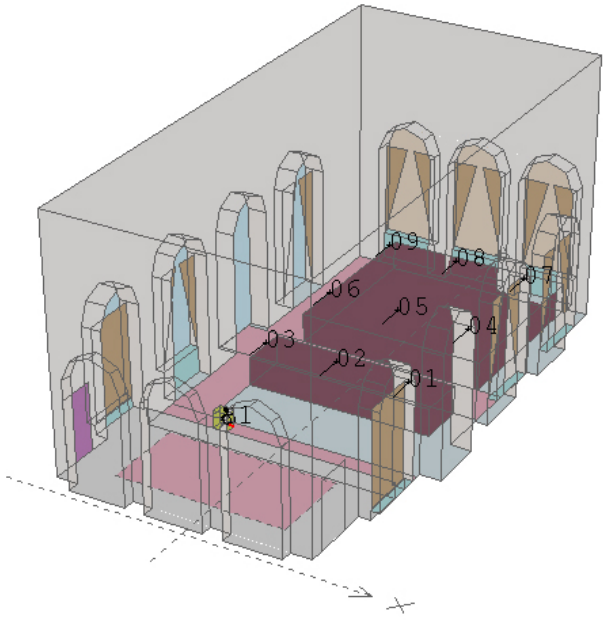

Figure 6.6: *Variants 1 and 2:* with seating, no audience

reverberation time matched. The maximum difference of T-30 between the simulation and the measurement was smaller than 100 ms. Scattering coefficients were

	125	250	500	1k	2k	4k	
EyrT	0,94	0,93	0,75	0,64	0,61	0,58	s
EyrTq	0,93	0,92	0,75	0,64	0,61	0,57	s
SabT	0,98	0,97	0,81	0,70	0,67	0,63	s
T-15	1,14	1,13	0,91	0,78	0,76	0,67	s
T-30	1,23	1,22	0,92	0,78	0,74	0,65	s
Tref	1,20	1,20	1,00	0,80	0,70	0,70	s
AbsC	19,05	19,19	23,06	26,27	26,80	26,76	%
AbsCq	19,13	19,27	23,09	26,32	26,93	26,81	%
MFP	4,94	4,94	4,95	4,95	4,94	4,94	m
Diffs	15,30	20,17	26,40	31,98	37,49	43,38	%

Figure 6.7: *Variant 2:* with seating, no audience; T-30 as in [Raumakustikskript]

entered manually for every material by estimating the roughness of the surfaces. Generally, they rise towards higher frequencies. The average reverberation time of all 9 receivers is shown in Figure 6.7.

Variant 3

Variants 3, 4 and 5 are related to variant 2, only the occupancy of the seating or audience changes. In variant 3 the audience area is completely occupied by audience (see Figure 6.8). Thus, the average reverberation time of all 9 receivers decreases as shown in Figure 6.9.

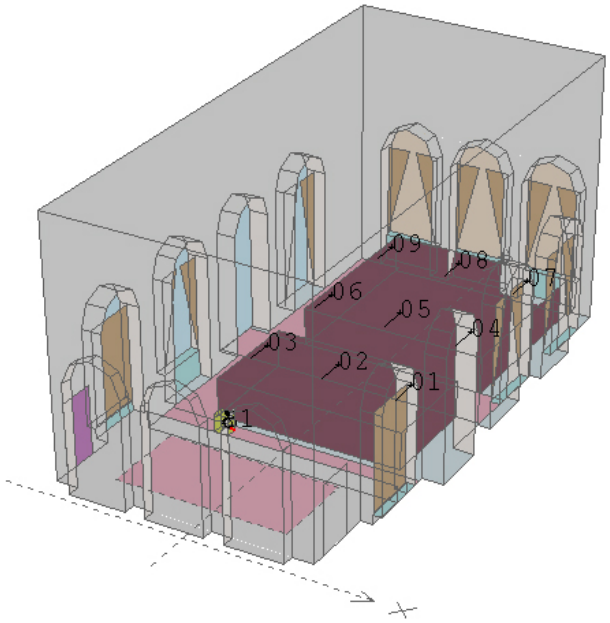

Figure 6.8: *Variant 3:* with audience

	125	250	500	1k	2k	4k	
EyrT	0,86	0,83	0,66	0,57	0,55	0,53	s
EyrTg	0,86	0,82	0,66	0,57	0,55	0,53	s
SabT	0,92	0,88	0,73	0,64	0,62	0,59	s
T-15	1,09	1,04	0,82	0,72	0,71	0,63	s
T-30	1,18	1,14	0,88	0,73	0,76	0,67	s
Tref	1,20	1,20	1,00	0,80	0,70	0,70	s
AbsC	19,87	20,60	24,88	28,28	28,64	28,16	%
AbsCg	19,98	20,67	24,88	28,29	28,77	28,26	%
MFP	4,76	4,76	4,76	4,77	4,77	4,76	m
Diffs	15,68	20,78	27,10	32,81	38,48	44,42	%

Figure 6.9: *Variant 3:* with audience

Variant 4

No audience and no seating is located in the hall (see Figure 6.10). Accordingly, the reverberation time rises as can be seen in Figure 6.11.

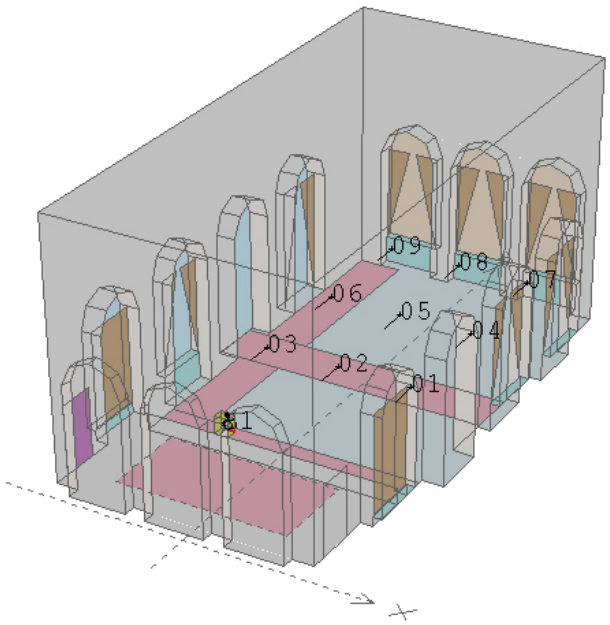

Figure 6.10: *Varian 4:* without audience and seating

	125	250	500	1k	2k	4k	
EyrT	1,19	1,33	1,11	0,94	0,88	0,75	s
EyrTg	1,19	1,33	1,12	0,95	0,88	0,75	s
SabT	1,24	1,38	1,17	1,01	0,94	0,81	s
T-15	1,35	1,45	1,22	1,02	0,93	0,79	s
T-30	1,42	1,49	1,25	1,05	0,95	0,80	s
Tref	1,20	1,20	1,00	0,80	0,70	0,70	s
AbsC	16,37	14,77	17,28	19,87	20,72	21,99	%
AbsCg	16,41	14,72	17,14	19,77	20,69	21,94	%
MFP	5,31	5,31	5,32	5,31	5,32	5,31	m
Diffs	13,88	18,29	24,16	29,27	34,37	39,82	%

Figure 6.11: *Varian 4:* without audience and seating

Variant 5

This variant is exactly like variant 4 but in addition the curtains are removed (see Figure 6.12). The average reverberation time can be seen in Figure 6.13.

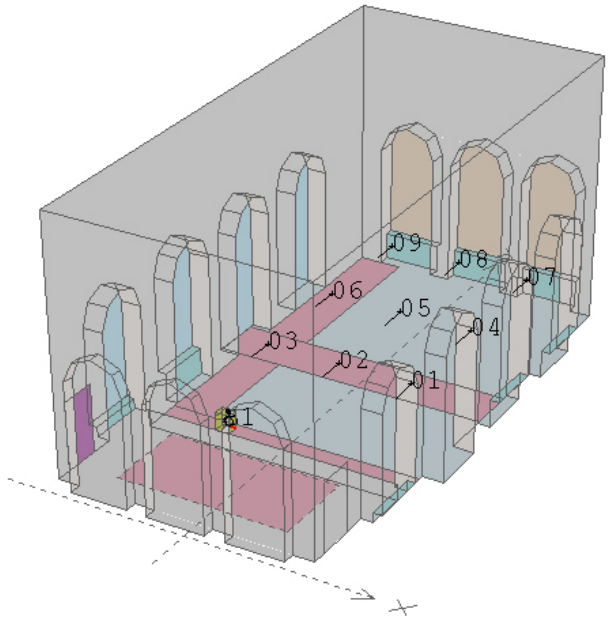

Figure 6.12: *Variant 5:* without audience, seating and curtains

	125	250	500	1k	2k	4k	
EyrT	1,32	1,49	1,31	1,08	1,08	0,90	s
EyrTg	1,32	1,49	1,32	1,09	1,08	0,90	s
SabT	1,38	1,54	1,37	1,15	1,14	0,95	s
T-15	1,42	1,56	1,37	1,15	1,10	0,89	s
T-30	1,49	1,59	1,38	1,17	1,12	0,91	s
Tref	1,20	1,20	1,00	0,80	0,70	0,70	s
AbsC	15,80	14,07	15,71	18,51	18,04	19,34	%
AbsCg	15,82	14,02	15,58	18,38	18,01	19,34	%
MFP	5,67	5,67	5,68	5,68	5,68	5,68	m
Diffs	14,10	18,49	24,37	29,58	34,67	40,00	%

Figure 6.13: *Variant 5:* without audience, seating and curtains

Variant 6

This variant is almost identical to variant 2. The only difference is that the right wall (as seen from the receiver view) is completely reflecting (see Figure 6.14). The average reverberation time can be seen in Figure 6.15.

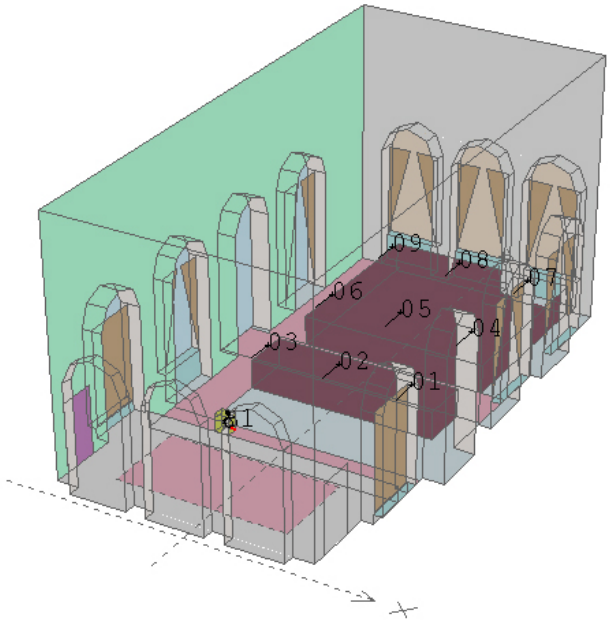

Figure 6.14: *Variant 6:* like variant 2 but one wall is completely reflecting

	125	250	500	1k	2k	4k	
EyrT	1,01	0,96	0,78	0,67	0,64	0,59	s
EyrTg	1,00	0,96	0,78	0,67	0,64	0,59	s
SabT	1,05	1,01	0,83	0,73	0,70	0,64	s
T-15	1,26	1,21	0,95	0,82	0,76	0,67	s
T-30	1,40	1,30	1,01	0,86	0,82	0,70	s
Tref	1,20	1,20	1,00	0,80	0,70	0,70	s
AbsC	17,83	18,49	22,39	25,33	25,84	26,11	%
AbsCg	17,92	18,55	22,36	25,36	25,96	26,21	%
MFP	4,94	4,93	4,95	4,94	4,95	4,94	m
Diffs	14,66	18,92	25,24	30,08	35,63	40,86	%

Figure 6.15: *Variant 6:* like variant 2 but one wall is completely reflecting

Variant 7

This variant is a so-called walkthrough. The hall is exactly the same as in variant 2. 30 receiver positions were programmed along the path shown in Figure 6.16. The height of the receivers was 1.7 m. A walkthrough can be generated in the binaural post-processing module. Instead of convolving each BRIR separately with an anechoic signal the ”Walkthrough convolver” has to be used. Thereby it is necessary to create a ”Walkthrough convolver script (*.WCS)” which contains the order and time of receiver positions. In between the positions, it is interpolated. In the end, a single WAV-file is generated. More information about how a walkthrough is created can be obtained from [CATT User's Manual].

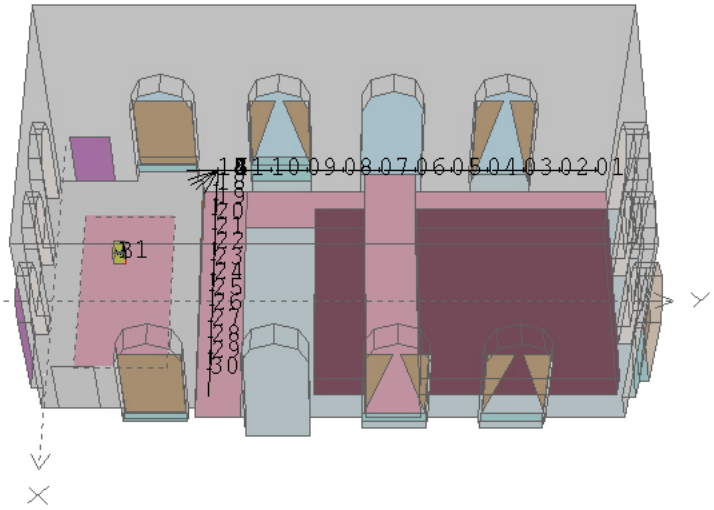

Figure 6.16: *Variant 7:* walkthrough; geometry like variant 2

Variant 8

This variant was simulated after the measurement with the binaural recording system (see Section 6.3) has been carried out. The reason was that during the measurement the height of the loudspeaker was only 1.3 m instead of 1.8 m above the stage. Furthermore, the height of the binaural recording system was 1.3 m instead of 1.2 m. Hence, the source and receiver dimensions of this simulation variant were adjusted to the ones in the measurement. The geometry is the same as in variant 2. The average reverberation time can be seen in Figure 6.17.

	125	250	500	1k	2k	4k	
EyrT	0,94	0,93	0,75	0,64	0,62	0,58	s
EyrTg	0,93	0,92	0,75	0,64	0,61	0,58	s
SabT	0,98	0,97	0,81	0,70	0,67	0,63	s
T-15	1,14	1,12	0,90	0,77	0,73	0,65	s
T-30	1,27	1,19	0,94	0,83	0,75	0,66	s
Tref	1,20	1,20	1,00	0,80	0,70	0,70	s
AbsC	19,04	19,22	23,08	26,28	26,79	26,74	%
AbsCg	19,13	19,27	23,09	26,32	26,93	26,81	%
MFP	4,94	4,94	4,95	4,95	4,94	4,94	m
Diffs	15,24	20,20	26,40	31,96	37,53	43,29	%

Figure 6.17: *Variant 8:* different height of source and receivers; geometry like variant 2

6.2 Post-processing of the simulations

In CATT-Acoustic the first step is to enter the geometry and run a so-called ”Full detailed calculation” using the prediction module. It is important to select ”Save data for Post processing (ECH)”. Thereby it is made sure that echograms (*.ECH-files) are created which contain all necessary information needed to synthesize a BRIR as well as other formats of impulse responses. The calculation of the BRIR and the convolution with an anechoic source signal is completed in the binaural post-processing module. The last step is to calibrate the obtained WAV-files. The reason for calibration is that the convolver auto-scales each WAV-file for maximum dynamic range. Thus, for comparability of different receiver positions with the same source signal, relative calibration has to be performed [CATT User’s Manual]. An overview of the signal flow is shown in Figure 6.18.

The applied HRTF was CATT1_plain_44.DAT and the headphone equalization filter was BEYER_DT990PRO_plain_44.DAT. Different anechoic source signals can be chosen in CATT-Acoustic. In this auralization each BRIR was convolved with two anechoic signals, one guitar signal (ag13_44_an.wav) and one speech signal of a female talker (ft_44_an.wav). The binaural output signals were named ssss_ss_rr.wav, for example for receiver 01, source B1 and source signal ag13_44_an.wav the output signal was named ag13_B1_01.wav. They are located in the OUT-folder of the specific simulation variant.

Auditory impressions

The author’s subjective auditory impressions were:

- Generally, it was possible to localize the source.

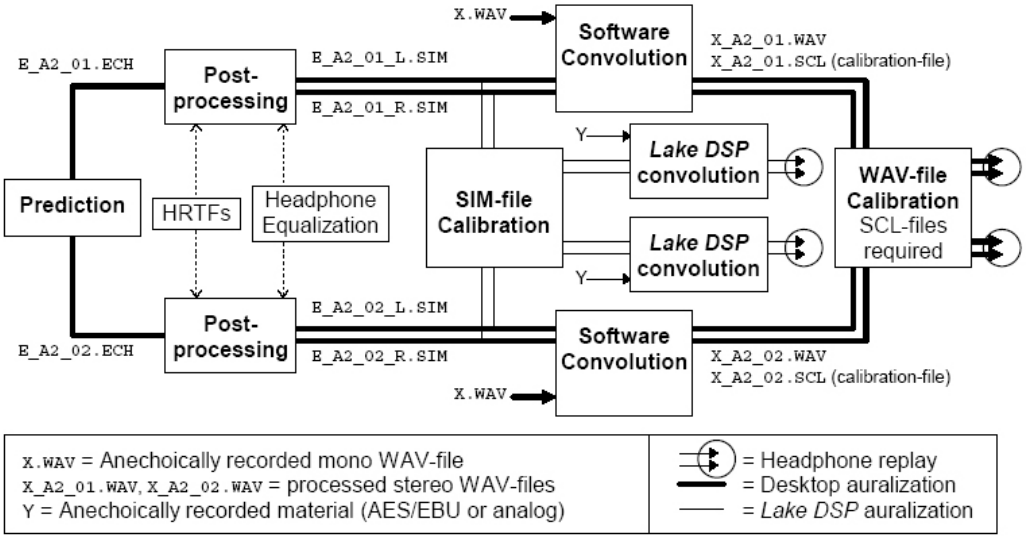

Figure 6.18: Signal flow from prediction to binaural post-processing and auralization [CATT User's Manual]. Desktop auralization was used.

- For the front receiver positions localization was easier than for the rear positions.
- For the rear positions the sound was perceived as more diffuse.
- Between variant 1 and 2 no considerable differences could be perceived.
- In positions 2, 5 and 8 the source was localized on the right of the middle although it should be in the center.
- Speech was easier to localize than music.
- Variant 3 sounds a little more dull than variant 1 and 2. Only little differences could be perceived since the seating in variant 1 and 2 is already heavily upholstered.
- The variants 4 and 5 have considerably more reverberation than the variants 1, 2 and 3.
- For variant 6 no differences could be perceived compared to variant 2.
- Variant 8 sounds as if the source was located further away compared to variant 1 and 2.
- Walking towards the source and passing it could be perceived in the walkthrough (variant 7).

6.3 Measurement in the Florentinersaal

The positions of source and receivers were the same as in the simulation. The height of the source was 1.3 m above the stage. The height of the receivers was 1.3 m. Thus, simulation variant 8 is the best one to compare with the measurement. The following equipment was utilized:

- device to play the WAV-files back: a Toshiba laptop with iTunes
- audio interface: USB audio interface Link.USB from Tapco
- loudspeaker: Genelec 1030A Bi-Amplified Monitoring System
- 2 microphones: each with AKG CK 92 omni capsule and SE 300 B powering unit
- the binaural recording system ”Source” (see Figure 6.19)
- recording device: 2-channel mobile digital recorder M-Audio MicroTrack

Figure 6.19: The binaural recording system ”Source” [Graf, 1999] with 2 AKG microphones.

The same anechoic signals as used in the auralization (ag13_44_an.wav and ft_44_an.wav) were also played back in the Florentinersaal and recorded at the

Figure 6.20: The loudspeaker is located on the stage and the binaural recording system is located on position 1.

Figure 6.21: The binaural recording system is located on position 5.

9 positions. The recorded signals were named ssss_rr_Messung.wav, for example AG13_01_Messung.wav is the signal ag13_44_an.wav recorded at position 1.

Comparing the recorded signals to those of the auralization they sound very different. The recorded signals sound brighter and can be localized easier. Furthermore, the room is better audible.

6.4 Auralization software

As mentioned above, the simulation and auralization of the Florentinersaal has been completed with **CATT-Acoustic**[1] (Computer Aided Theater Technique). It has been developed by Dr. Dalenbäck in cooperation with the Chalmers University of Technology, Sweden. A free demo-version including auralization can be downloaded on the web-site.

However, room acoustic modeling and auralization programs are also provided by other developers:

ODEON [2] developed by the Technical University of Denmark in cooperation with a group of consulting companies. There is also a free demo-version available on the web-site.

EASE [3] (Electro-Acoustic Simulator for Engineers) developed by Acoustic Design Ahnert, Germany.

RAMSETE [4] developed by the University of Parma, Italy. A free demo-version can be downloaded on the web-site.

AUVIS [5] (AUralization of VIrtual Studios) developed for the simulation of recording studios by the Institut für Rundfunktechnik, Munich, Germany.

Other LMS RayNoise [6] and Bose Auditioner [7].

[1] http://www.catt.se/
[2] http://www.odeon.dk/
[3] http://www.ada-acousticdesign.de/set_en/setsoft.html and http://www.auralisation.de/
[4] http://www.ramsete.com/Ramsete_Ultimo/home.htm
[5] http://www.irt.de/en/products/acoustics/auralisation-with-simulation-software-auvis.html
[6] http://www.lmsintl.com/RAYNOISE
[7] http://pro.bose.com/ProController?url=/pro/technologies/auditioner/index.jsp

Chapter 7

Conclusions

For auralization three components affect the result and, thus, have to be defined and modeled:

- source
- medium
- receiver.

Usually, the directivity of the source is accounted for by fixed radiation patterns defined in octave bands which are available in directivity databases. For measurements of radiation patterns, a set of microphones is arranged around the source to cover the directional characteristics. The signal played back by the source during the simulation has to be anechoic since the signal should only be filtered by one room, the room under study. Mostly, this anechoic recording is made in the main radiation axis of the source.

In room acoustics the medium is the room which is defined by its geometry and surface properties, that is absorption and scattering coefficients. The state-of-the-art simulation models are based on geometrical acoustics. So-called hybrid models which are a combination of stochastic (ray tracing) and deterministic methods (image source modeling) are used. At the receiver position the direction, energy, and time of the incident sound rays is registered.

The modeling of the receiver (listener) is done with the aid of HRTFs. Hence, the auralization should sound as if being in the real room and listening to the source. However, the auralization of the Florentinersaal in Graz sounded different in comparison to the recording. This indicates that a measurement in the real room should always be preferred, if possible. Anyhow, differences between the

simulation variants were audible. Thus, auralization can be a help for decisions during the construction phase of a room. In addition, it is a striking tool for presentations to clients.

It should be noted that simulation models based on geometrical acoustics are only an approximation of reality. In the future more simulation methods, for example wave-based methods, will be applicable due to the rising of computational power.

Bibliography

[Ahnert, 2003] Ahnert W., Feistel S., Schmitz O.: Modern tools in acoustic design of concert halls and theatres - use and limitations of computer simulation and auralisation, XIII Session of the Russian Acoustical Society, p. 863-874, Moscow, Russia, August 2003

[Atal, 1966] Atal B. S., Schroeder M. R.: Apparent sound source translator, US Patent No. 3,236,949, February 1966

[Berkhout, 1988] Berkhout A. J.: A holographic approach to acoustic control, J. Audio Eng. Soc., Vol. 36, No. 12, p. 977-995, December 1988

[Blauert, 1996] Blauert J.: Spatial hearing: the psychophysics of human sound localization, MIT Press, 1996

[CATT User's Manual] CATT-Acoustic v8.0 Room Acoustics Prediction and Auralization, User's Manual, 2002

[Dalenbäck, 2006] Dalenbäck B.-I., Strömberg M.: Real time walkthrough auralization - the first year, Proc. IoA Copenhagen, Denmark, May 2006

[Funkhouser, 1998] Funkhouser T., Carlbom I., Elko G., Pingali G., Sondhi M., West J.: A beam tracing approach to acoustic modeling for interactive virtual environments, Proc. 25th annual conference on computer graphics and interactive techniques, SIGGRAPH 1998

[Gerzon, 1976] Gerzon M. A.: Multidirectional sound reproduction systems, US Patent No. 3,997,725, December 1976

[Graf, 1999] Graf F.: Entwicklung eines Aufnahmesystems für psychoakustische Analysen, Master's Thesis (Diplomarbeit), TU Graz, March 1999

[Kleiner, 1993] Kleiner M., Dalenbäck B.-I., Svensson P.: Auralization - an overview, J. Audio Eng. Soc., Vol. 41, No. 11, p. 861-875, November 1993

[Krokstad, 1968] Krokstad A., Strøm S., Sørsdal S.: Calculating the acoustical room response by the use of a ray tracing technique, J. Sound Vib., Vol. 8, p. 118-125, 1968

[Kuttruff, 1993] Kuttruff H.: Auralization of impulse responses modeled on the basis of ray-tracing results, J. Audio Eng. Soc., Vol. 41, No. 11, November 1993

[Kuttruff, 1995] Kuttruff H.: A simple iteration scheme for the computation of decay constants in enclosures with diffusely reflecting boundaries, J. Acoust. Soc. Am., Vol. 98, No. 1, p. 288-293, July 1995

[Lentz, 2002] Lentz T., Schmitz O.: Realisation of an adaptive cross-talk cancellation system for a moving listener, Proc. AES 21st Conference, St. Petersburg, Russia, June 2002

[Lentz, 2004] Lentz T., Behler G.: Dynamic cross-talk cancellation for binaural synthesis in virtual reality environments, Proc. AES 117th Convention, San Francisco, CA, USA, October 2004

[Lentz, 2007] Lentz T., Schröder D., Vorländer M., Assenmacher I.: Virtual reality system with integrated sound field simulation and reproduction, EURASIP Journal on Advances in Signal Processing, January 2007

[Mackensen, 2004] Mackensen P.: Auditive localization. Head movements, an additional cue in localization, Dissertation, Technical University Berlin, 2004

[Mechel, 2002] Mechel F. P. (ed.) et al.: Formulas of acoustics, Springer-Verlag Berlin Heidelberg New York, 2002

[Raumakustikskript] Skript zur Vorlesung Raumakustik, Version 4.0, TU Graz, Sommersemester 2006

[Rindel, 2004] Rindel J. H., Otondo F., Christensen C. L.: Sound source representation for auralization, Proc. international symposium on room acoustics: design and science, Hyogo, Japan, April 2004

[Savioja, 1999] Savioja L., Huopaniemi J., Lokki T., Väänänen R.: Creating interactive virtual environments, J. Audio Eng. Soc., Vol. 47, No. 9, p. 675-705, September 1999

[Schröder, 2006] Schröder D., Lentz T.: Real-time processing of image sources using binary space partitioning, J. Audio Eng. Soc., Vol. 54, No. 7/8, p. 604-619, July/August 2006

[Schröder, 2007] Schröder D., Dross P., Vorländer M.: A fast reverberation estimator for virtual environments, Proc. AES 30th International Conference, Saariselkä, Finland, March 2007

[Schroeder, 1973] Schroeder M. R.: Computer models for concert hall acoustics, Am. J. Phys., Vol. 41, p. 461-471, April 1973

[Theile, 2003] Theile G., Wittek H., Reisinger M.: Wellenfeldsynthese - Neue Möglichkeiten der räumlichen Tonaufnahme und -wiedergabe, Fernseh- und Kino-Technik - Sonderausdruck Nr. 5 und 6/2003

[Torres, 2000] Torres R. R., Kleiner M., Dalenbäck B.-I.: Audibility of "diffusion" in room acoustics auralization: an initial investigation, Acustica united with Acta Acustica, Vol. 86, p. 919-927, March 2000

[Vorländer, 2000] Vorländer M., Mommertz E.: Definition and measurement of random-incidence scattering coefficients, Appl. Ac., Vol. 60, p. 187-199, 2000

[Vorländer, 2008] Vorländer M.: Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality, Springer-Verlag Berlin Heidelberg, 2008