Creating Interactive Virtual Acoustic Environments*

LAURI SAVIOJA[1,3,4], *AES Member*, **JYRI HUOPANIEMI**[2], *AES Member*, **TAPIO LOKKI**[1,3], **AND RIITTA VÄÄNÄNEN**[1,4]

[1]*Helsinki University of Technology, FIN-02015 HUT, Finland*
[2]*Nokia Research Center, Speech and Audio Systems Laboratory, FIN-00045 Nokia Group, Finland*

The theory and techniques for virtual acoustic modeling and rendering are discussed. The creation of natural sounding audiovisual environments can be divided into three main tasks: sound source, room acoustics, and listener modeling. These topics are discussed in the context of both non-real-time and real-time virtual acoustic environments. Implementation strategies are considered, and a modular and expandable simulation software is described.

0 INTRODUCTION

The audiovisual technology (audio, still picture, video, animation, and so on) is rapidly converging to integrated multimedia. Research on audiovisual media modeling has increased dramatically in the last decade. Standardization of advanced multimedia and virtual reality rendering and definition has been going on for several years. Examples are the Moving Picture Experts Group (MPEG) [1] and the Virtual Reality Modeling Language Consortium (VRML) [2] standardization bodies. Similar work (audiovisual scene description) has also been carried out in the Java three-dimensional application programming interface specification [3], [4] and in the Interactive Audio Special Interest Group (IASIG) [5]. The progress of multimedia has introduced new fields of research and application into audio and acoustics, one of which is virtual acoustic environments, also called virtual acoustic displays (VADs), and their relation to graphical presentations.

0.1 Virtual Acoustics

In this paper we use the term "virtual acoustics" to cover the modeling of three major subsystems in an acoustical communication task:

1) Source modeling
2) Transmission medium (room acoustics) modeling
3) Receiver (listener) modeling.

This simple source–medium–receiver model is common to all communication systems, but in this paper it is discussed from the viewpoint of acoustics [6], [7]. Interactive virtual acoustic environments and binaural room simulation systems have been studied in recent years (for example, [6]–[13]), but until recently the physical relevance and the relation of acoustical and visual content have not been of major interest.

In this paper underlying DSP concepts for efficient and auditorily relevant multimedia sound processing are studied, and a modeling schematic is proposed for virtual acoustics. The goal is to deliver an acoustical message in a virtual reality system from the source to the receiver as it would happen in a real-world situation. The process of implementing virtual acoustic displays can be divided into three different stages, as depicted in Fig. 1 [14]: a) definition, b) modeling, and c) reproduction. The definition of a virtual acoustic environment includes a priori knowledge and data of the system to be implemented, that is, information about the sound sources, the room geometry, and the listeners. The term "auralization" [15] is understood as a subset of the virtual acoustic concept referring to the modeling and reproduction of sound fields (as shown in Fig. 1). In the modeling stage we can divide the rendering to the three afore-mentioned tasks (source, room, listener). These tasks are described shortly in the following.

Source modeling (Fig. 1) consists of methods that produce (and possibly add physical character to) sound in an audiovisual scene. The most straightforward method is to use prerecorded digital audio. In most aural-

* Original version presented at the 100th Convention of the Audio Engineering Society, Copenhagen, Denmark, 1996 May 11–14 [46]; revised 1998 June 29 and 1999 July 1.
[3] Telecommunications Software and Multimedia Laboratory.
[4] Laboratory of Acoustics and Audio Signal Processing.

ization systems sound sources are treated as omnidirectional point sources. The approximation is valid for many cases. More accurate methods are, however, often needed. For example, most musical instruments have radiation patterns that are frequency dependent. Typically sound sources radiate more energy to the frontal hemisphere whereas sound radiation is attenuated and low-pass filtered when the angular distance from the on-axis direction increases. In this paper we present methods for efficient modeling of sound source directivity.

The task of modeling sound propagation behavior in acoustical spaces (Fig. 1) is discussed. The goal in most room acoustic simulations has been to compute an energy–time curve (ETC) of a room (squared room impulse response), and based on that, to derive room acoustical attributes such as reverberation time (RT_{60}). The ray-based methods—ray tracing [16], [17] and the image-source method [18], [19]—are the most often used modeling techniques. Recently computationally more demanding wave-based techniques such as finite-element method (FEM), boundary-element method (BEM), and finite-difference time-domain (FDTD) method have also gained interest [20]–[22]. These techniques are suitable for the simulation of low frequencies only [15]. In real-time auralization the limited calculation capacity calls for simplications, modeling only the direct sound and early reflections individually, for example, and the late reverberation by recursive digital filter structures [10], [23]–[25].

In listener modeling (Fig. 1) the properties of human spatial hearing are considered. Simple means for giving a directional sensation of sound are the interaural level and time differences (ILD and ITD), but they cannot resolve the front–back confusion (see [7], [15], [26]–[28] for fundamentals on spatial hearing and auralization). The head-related transfer function (HRTF), which models reflections and filtering by the head, shoulders, and pinnae of the listener, has been studied extensively during the past decade. With the development of HRTF measurement [29], [30] and efficient filter design methods [31]–[37] real-time HRTF-based three-dimensional sound implementations have become applicable in virtual environments.

Reproduction schemes of virtual acoustic environments can be divided into the following categories [15]: 1) binaural (headphone), 2) crosstalk canceled binaural (loudspeaker), and 3) multichannel reproduction. Binaural processing refers to three-dimensional sound image production for headphone or loudspeaker listening. For loudspeaker reproduction of binaural signals, crosstalk canceling technology was first implemented by Schroeder and Atal [38]. This theory has been further developed subsequently [36], [39]–[42]. The most common multichannel reproduction techniques are Ambisonics [43], [44] and vector base amplitude panning (VBAP) [45].

0.2 System Implementation

The integrated system described in this paper is called DIVA (Digital Interactive Virtual Acoustics) [46]. The DIVA system consists of several subparts which have been studied at the Helsinki University of Technology (HUT). In this publication the main emphasis is on the audio components of the system.

Our purpose in the present research has been to make a real-time environment for full audiovisual experience.

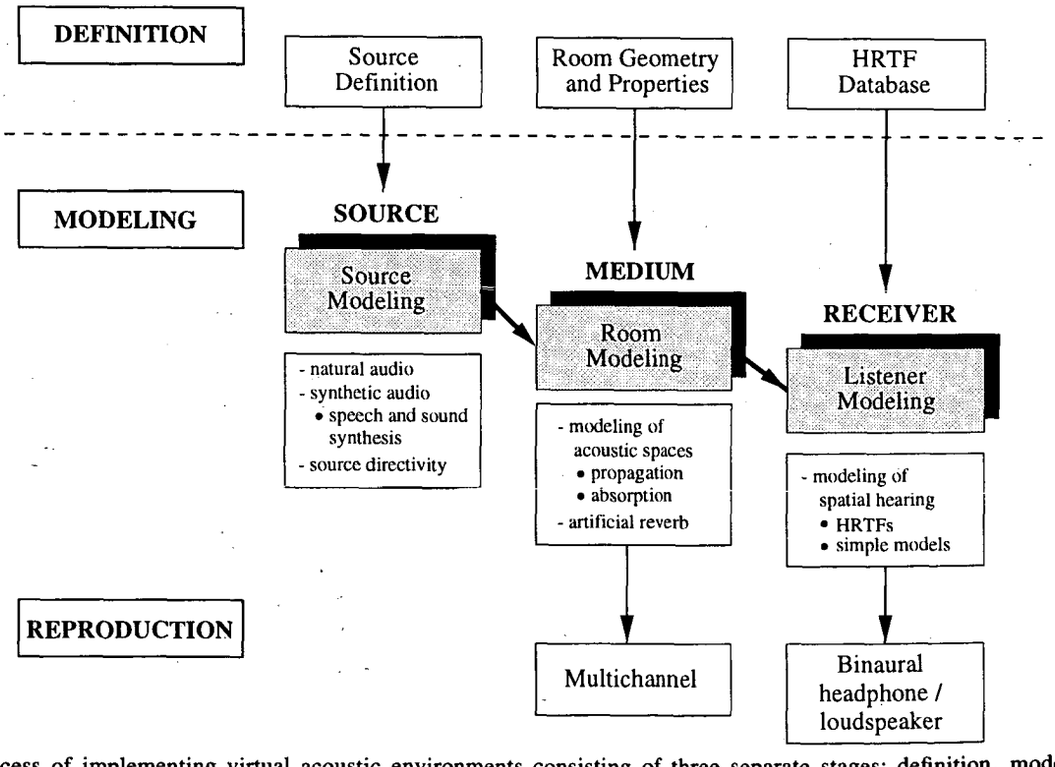


Fig. 1. Process of implementing virtual acoustic environments consisting of three separate stages: definition, modeling, and reproduction. Components to be modeled: sound source, medium, and receiver.

The system integrates the whole audio signal processing chain from sound synthesis through room acoustics simulation to spatialized reproduction. This is combined with synchronized animated motion. A practical application of this project has been a virtual concert performance [47], [48]. However, synchronized sound effects are an important extension and navigation aid in any virtual environment [49], [50]. The processing of audio signals with an imagined physical environment, even without visual display, may give composers and sound designers new powerful means of artistic expression. Synchronization of real-time audio and video is also an important topic in networked multimedia applications such as teleconferencing. Anticipating this, the DIVA software is implemented as a distributed system in which each part of the system may reside in a separate computing node of a high-speed network.

Fig. 2 presents the architecture of the DIVA virtual concert performance system. There may be two simultaneous users in the system, a conductor and a listener, who both can interact with the system. The conductor wears a tailcoat with magnetic sensors for tracking. With his or her movements the conductor controls the orchestra, which may contain both real and virtual musicians. The aim of the conductor gesture analysis is to synchronize electronic music with human movements. The main concern has been with capturing the rhythm. In addition, nuances of the performance such as crescendo and diminuendo can also be recognized. Our current system is based on an artificial neural network (ANN). A thorough discussion of this topic can be found in [51].

In the graphical user interface (GUI) of DIVA, animated human models are placed on stage to play music from MIDI files. The virtual musicians play their instruments at the tempo and loudness shown by the conductor. Instrument fingerings are found from predefined grip tables. All the joint motions of the human hand are found by inverse kinematics calculations, and they are synchronized to exactly perform each note on an animated instrument model. A complete discussion of computer animation in the DIVA system is presented in [52].

At the same time a listener may freely fly around in the concert hall. The GUI sends the listener position data to the auralization unit, which renders the sound samples provided by physical models and a MIDI synthesizer. The auralized output is reproduced through either headphones or loudspeakers.

This paper is organized as follows. Section 1 discusses different methods used in modeling sound sources for virtual acoustics. Section 2 briefly introduces the basics of computational room acoustics. The methods suitable for real-time simulation and auralization are discussed in more detail. The main principles of human spatial hearing and its modeling are presented in Section 3. Section 4 discusses the requirements of real-time interactive auralization. Sections 5 and 6 present the implemented software system and a case study of the auralization of an old gothic cathedral. Future plans and conclusions are presented in Sections 7 and 8.

1 MODELING OF SOUND SOURCES

Sound source modeling in the virtual acoustic concept refers to attaching sound to an environment and giving it properties such as directivity. The simplest approach has traditionally been to use an anechoic audio recording or a synthetically produced sound as an input to the auralization system. In an immersive virtual reality system, however, a more general and thorough modeling approach is desired that takes into account the directional characteristics of sources in an effective way. The following are general qualitative requirements for audio source signals in a virtual acoustic system (used for physical modeling).

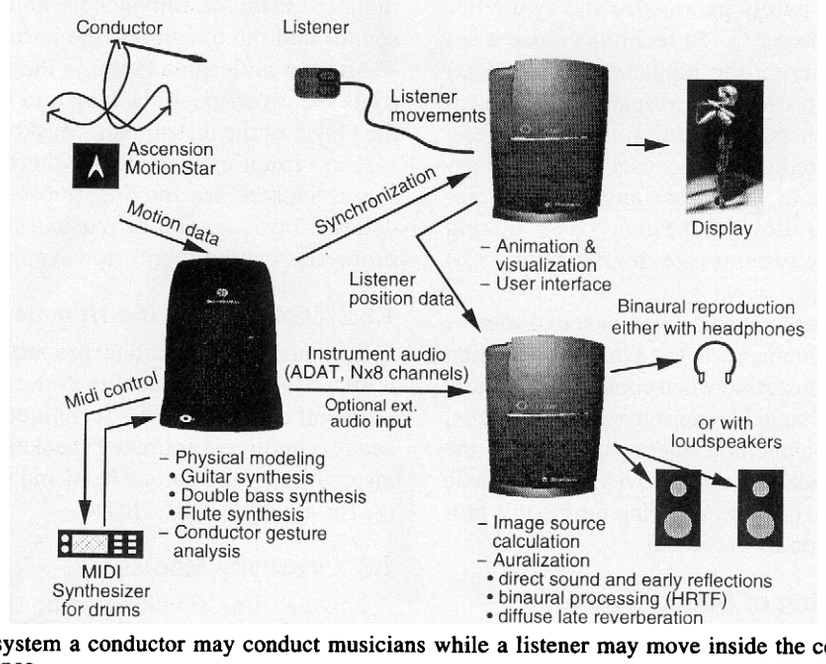


Fig. 2. In the DIVA system a conductor may conduct musicians while a listener may move inside the concert hall and listen to an auralized performance.

- Each sound source signal should be "dry," not containing any reverberant or directional properties, unless it is explicitly desired in the simulation (such as simulating stereophonic listening in a listening room).
- The audio source inputs are normally treated as point sources in the room acoustical calculation models (such as the image-source and ray-tracing methods discussed in Section 2). This requirement causes the source signals to be monophonic. A stereophonic signal emanating from loudspeakers can thus be modeled as two point sources.
- The quality (signal-to-noise ratio, quantization, sampling rate) of the sound source signal should be adequately high not to cause undesired effects in auralization.

From the coding and reproduction point of view, audio source signals can be divided into two categories (similarly as in the MPEG-4 standardization work [53]):

1) Natural audio
2) Synthetic audio.

The concept of natural audio refers to sound signals that have been coded from an existing waveform. This category includes all forms of digital audio, raw or bit-rate reduced, and thus the basic requirements for use in virtual reality sound are those stated in the previous section. By using various bit-rate reduction techniques relatively efficient data transfer rates for high-quality digital audio may be achieved [54].

Purely synthetic audio may be explained as sound signal definition and creation without the aid of an a priori coded sound waveform. Exceptions are sound synthesis techniques that involve wavetable data and sampling, which both use a coded waveform of some kind. These methods can be regarded as hybrid natural/synthetic sound generation methods. On the other hand, many sound synthesis techniques such as frequency modulation, granular synthesis, additive synthesis, and physical modeling are purely parametric and synthetic.

Text-to-speech synthesis (TTS) techniques use a set of parameters that describe the input (often pure text) as phonemes and as prosodic information. This also allows for very low-bit-rate transmission, because the speech output is rendered in the end-user terminal. Furthermore, it is possible to attach face animation parameters to TTS systems to allow for the creation of "talking heads" in virtual reality systems (see, for example, [53]) for more information).

The advantage of using synthetic audio as sound sources in virtual acoustic environments is the achieved reduction in the amount of data transferred when compared to natural audio (where the coded signal is transferred). This results, however, in added computation and complexity in the audio rendering. Parametric control provided by synthetic audio is a remarkable advantage, allowing for flexible processing and manipulation of the sound.

1.1 Physical Modeling of Sound Sources

An attractive sound analysis and synthesis method, physical modeling has become a popular technique among researchers and manufacturers of synthesizers in the past few years [55]. Physical-model-based sound synthesis methods are particularly suitable for virtual acoustics simulation due to many reasons. First one of the aims in virtual acoustics is to create models for the source, the room, and the listener that are based on their physical properties and can be controlled in a way that resembles their physical behavior. Second we may be able to incorporate source properties such as directivity in the physical modeling synthesizers [56], [57].

1.2 Properties of Sound Sources

Modeling the radiation properties and directivity of sound sources is an important (yet often forgotten) topic in virtual reality systems. We have investigated the directional properties of musical instruments and the human head, and implemented real-time algorithms for frequency-dependent directivity filtering for a point-source approximation. In general, the mathematical modeling of the directivity characteristics of sound sources (loudspeakers, musical instruments) can be a complex and time-consuming task. Therefore, in many cases measurements of directivity are used to obtain numerical data for simulation purposes [58].

1.2.1 On the Directivity of Musical Instruments

String instruments exhibit complex sound radiation patterns due to various reasons. The resonant mode frequencies of the instrument body account for most of the sound radiation (see, for example, [59]). Each mode frequency of the body has a directivity pattern such as monopole, dipole, quadrupole, or a combination thereof. The sound radiated from the vibrating strings, however, is weak and can be neglected in the simulation. In wind instruments, the radiation properties are dominated by outcoming sound from various parts of the instrument (the finger holes or the bell). For example, in the case of the flute, the directivity is caused by radiation from the embouchure hole (the breathier noisy sound) and the toneholes (the harmonic content) [57].

Another noticeable factor in the modeling of directivity is the directional masking and reflections caused by the player of the instrument. Masking plays an important role in virtual environments where the listener and the sound sources are moving freely in a space. A more detailed investigation of musical instrument directivity properties can be found, for example, in [58].

1.2.2 Directivity of the Human Head

Incorporating directional properties of the human head is attractive in virtual reality systems, for example, from the point of view of TTS techniques and related talking head technology (animated speakers). Directional filtering caused by the human head and torso are investigated in, for example, [60], [61].

1.3 Directivity Models

For real-time simulation purposes it is necessary to derive simplified source directivity models that are efficient from the signal processing point of view and as

good as possible from the perceptual point of view. Different strategies for directivity modeling have been considered in earlier studies [56], [57]. Of these methods, there are two strategies that are attractive to virtual reality audio source simulation in general: 1) directional filtering and 2) a set of elementary sources. In Fig. 3 these two methods are illustrated.

To obtain empirical data, we conducted measurements of source directivity in an anechoic chamber for two musical instruments: the acoustic guitar and the trumpet. In both cases two identical microphones were placed at 1-m distance from the source and the player, one being the reference direction α_0 (normally 0° azimuth and elevation) and the other being the measured direction α_m. An impulsive excitation was used, and the responses $H_{\alpha_0}(z)$ and $H_{\alpha_m}(z)$ were registered simultaneously. For modeling purposes, we normally consider directivity relative to main-axis radiation, that is, we compute the deconvolution of the reference and measured magnitude responses for each direction,

$$|D_{\alpha_m}(z)| = \frac{|H_{\alpha_m}(z)|}{|H_{\alpha_0}(z)|}. \quad (1)$$

1.3.1 Directional Filtering

The directivity properties of sound sources may be efficiently modeled in conjunction with geometrical acoustics methods such as the image-source method. In this method the monophonic sound source output $\hat{y}(n)$ is fed to M angle-dependent digital filters $D_{\alpha_m}(z)$, where $m = 1, \ldots, M$, representing each modeled output direction from the source [see Fig. 3(a)]. Generally the low-pass characteristic of the direction filter increases as a function of the angle, and very low-order filters have been found to suffice for real-time models.

As an example, a set of first-order IIR filters modeling the directivity characteristics of the trumpet (and the player) is represented in Fig. 4. The directivity functions were obtained using Eq. (1). A first-order IIR filter was fitted to each result. The filters were designed with a minimum-phase constraint. In our measurements we analyzed directivity only on the horizontal plane (0° elevation).

Another example is that of modeling the directivity of a spherical head (approximating a human head). Fig. 5 show results for fitting IIR filters (a two-pole and one–zero model) to spherical head directivity data [61] in different azimuth angles (0–150°). The transfer functions were obtained using an algorithm presented in [62]–[64] by applying the principle of reciprocity [61].

1.3.2 Set of Elementary Sources

The radiation and directivity patterns of sound sources may also be distributed, and a point-source approximation may not be adequate for representation. Such examples could, for example, be a line source. So far only a point-source type has been considered and more complex sources have been modeled as multiple point sources. The radiation pattern of a musical instrument may thus be approximated by a small number of elementary sources. These sources are incorporated in, for example, the physical model, and each of them produces an output signal (see [56]) [see Fig. 3(b)]. This approach is particularly well suited to flutes, where there are inherently two point sources of sound radiation, the embouchure hole and the first open tone hole. Another sound source example from physical reality could be a train, where the composite sound image consists of multiple elementary sounds emanating from different locations. A multiple-point-source approximation may also be valid in this case.

2 ROOM ACOUSTICS MODELING

In this section an overview of different room acoustic modeling methods is given. One wave-based method and two ray-based methods are introduced in more detail. Modeling techniques suitable for real-time auralization including the modeling of air and material absorption and late reverberation are presented.

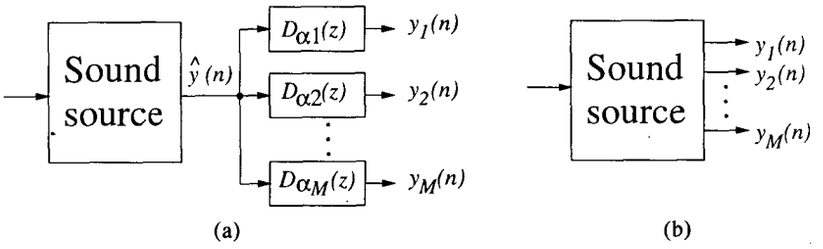
2.1 Overview of Room Acoustics Modeling Techniques

In modeling room acoustics, the basic division is made between computational simulation and scale modeling [15]. Fig. 6 represents different approaches of classification.

Scale models are used widely in the design of large concert halls. The technique is well established for measuring room acoustical attributes. The method is also suitable for auralization at predefined measurement locations using either the direct or the indirect scale-model auralization technique as described in Kleiner et al. [15].

Computers have been used for 30 years to model room acoustics [16], [65], [66], and nowadays computational modeling combined with the use of scale models is a relatively common practice in the acoustic design of concert halls. A good overview of modeling algorithms is presented in [20], [15].

There are three different approaches in the computa-



(a) (b)

Fig. 3. Two methods for incorporating directivity into sound source models. (a) Directional filtering. (b) Set of elementary sources.

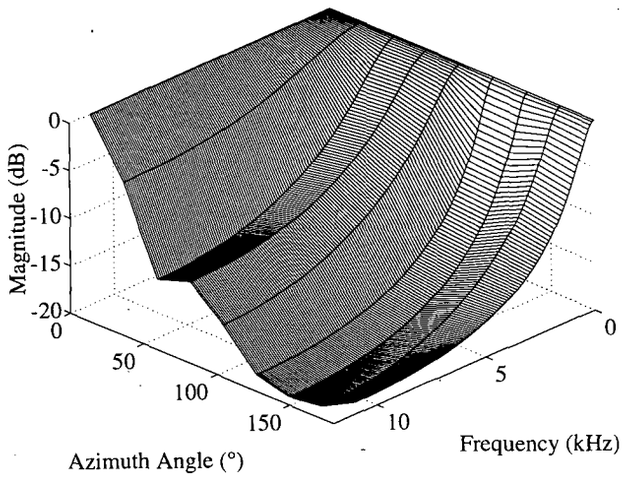


Fig. 4. Modeling directivity characteristics of trumpet with first-order IIR filters.

tional modeling of room acoustics, as illustrated in Fig. 6. The most accurate results can be achieved with the wave-based methods. An analytical solution for the wave equation can be found only in rare cases such as a rectangular room with rigid walls. Therefore some numerical wave-based methods must be applied. Element methods such as FEM and BEM are only suitable for small enclosures and low frequencies due to heavy computational requirements [15], [67]. FDTD methods provide another possible technique for room acoustics simulation [21], [22]. These time-domain methods which produce impulse responses are also better suited for auralization than FEM and BEM, which typically are calculated in the frequency domain [15].

The ray-based methods of Fig. 6 are based on the geometrical room acoustics, in which the sound is supposed to act like rays (see, for example, [68]). This assumption is valid when the wavelength of the sound

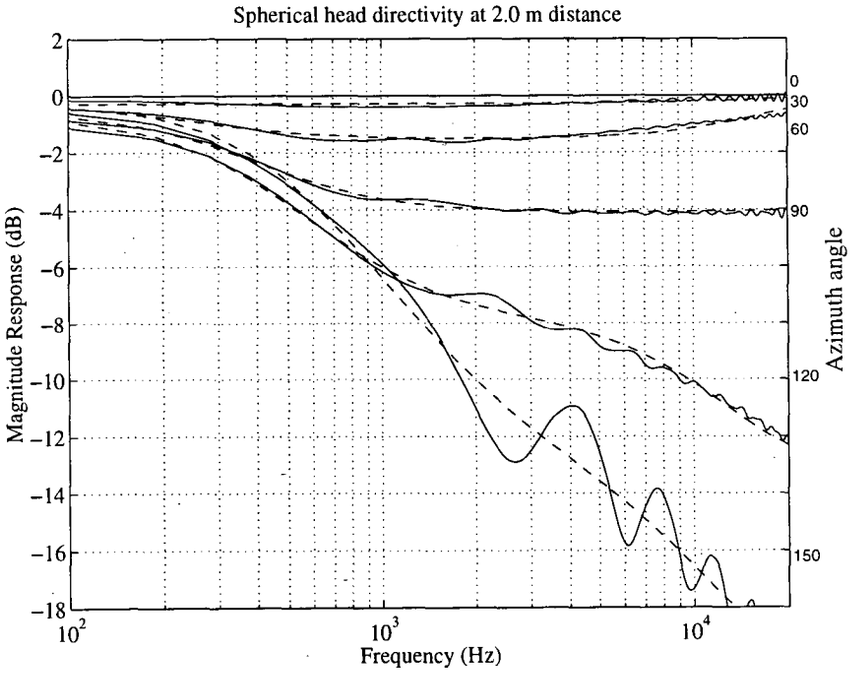


Fig. 5. Modeling directivity characteristics of human head with IIR filter (two-pole one–zero model).

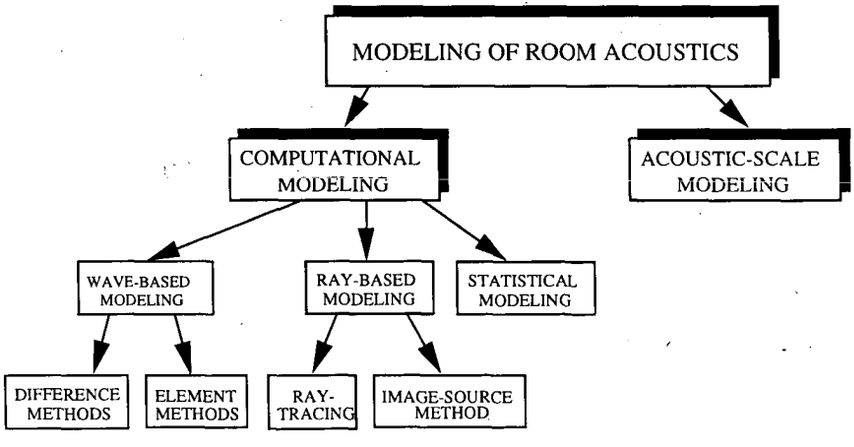


Fig. 6. Principal computational models of room acoustics are based either on sound rays (ray based) or on solving the wave equation (wave based). Different methods can be used together to form a valid hybrid model.

is small compared to the area of the surfaces in the room and large compared to the roughness of the surfaces. However, all phenomena due to the wave nature, such as diffraction and interference, are ignored. The ray-based methods are described in more detail in Section 2.3.

In addition there are also statistical modeling methods (see Fig. 6), such as the statistical energy analysis (SEA) [69]. Those methods are mainly used in the prediction of noise levels in coupled systems in which sound transmission by structures is an important factor, but they are not suitable for auralization purposes.

2.2 Digital Waveguide Mesh Method

The digital waveguide mesh method is a variant of the FDTD methods. A digital waveguide is a bidirectional digital delay line, and they are widely used in the physical modeling of musical instruments. In one-dimensional systems even real-time applications are easily possible [55], [70]–[73].

A digital waveguide mesh is a regular array of one-dimensional digital waveguides arranged along each perpendicular dimension, interconnected at their intersections. The resulting mesh of a three-dimensional space is a regular rectangular grid in which each node is connected to its six neighbors by unit delays. A detailed study on deriving the difference equations for the mesh and for the boundary conditions is presented in [74], [75].

An example of FDTD simulation is presented in Fig. 7, where one horizontal slice of a mesh is visualized. The space under study consists of three rooms, and the excitation signal has been in the largest one. The primary wavefront and the first reflection from the ceiling can be observed from Fig. 7.

The update frequency of an N-dimensional mesh is

$$f_s = \frac{c\sqrt{N}}{\Delta x} \approx \frac{588.9}{\Delta x} \quad \text{Hz} \tag{2}$$

where c represents the speed of sound in the medium and Δx is the spatial sampling interval corresponding to the distance between two neighboring nodes. The approximate value of Eq. (2) stands for a typical room simulation (c = 340 m/s, N = 3). That same frequency is also the sampling frequency of the resulting impulse response.

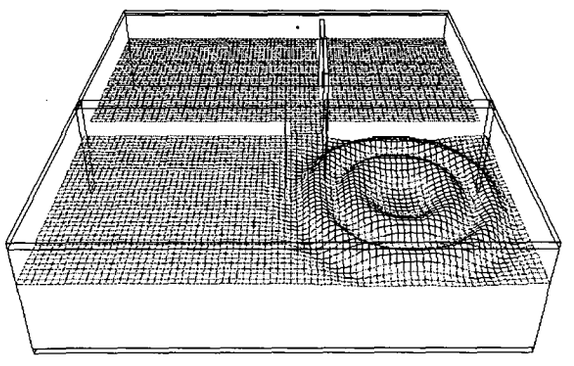


Fig. 7. Example of FDTD simulation. Model consists of three separate rooms. One horizontal slice of the digital waveguide mesh is visualized.

An inherent problem with the digital waveguide mesh method is the direction-dependent dispersion of wavefronts [74], [76]. This effect can be reduced by using structures other than the rectangular mesh, such as triangular or tetrahedral meshes [76], [77], or by using interpolation methods [78]. The direction-independent dispersion can be further reduced by using frequency warping [79], [80].

The accuracy of the digital waveguide mesh depends primarily on the density of the mesh. Due to the dispersion the model is practically only useful at frequencies well below the update frequency f_s of the mesh. In the estimation of room acoustical attributes phase errors are not severe, and results may be used approximately up to one-tenth of the update frequency. In that case there are at least six mesh nodes per wavelength. For auralization purposes the valid frequency band is more limited.

2.3 Ray-Based Room Acoustics Modeling

An impulse response of a concert hall can be separated into three parts: direct sound, early reflections, and late reverberation, as illustrated in Fig. 8(a). The response illustrated is a simplified one in which there are no diffuse or diffracted reflection paths. In real responses, as shown in Fig. 8(b), there is diffused sound energy also between early reflections, and the late reverberation is not strictly exponentially decaying. The results of ray-based models resemble the response in Fig. 8(a) since the sound is treated as sound rays with specular reflections. Note that in most simulation systems the result is the ETC, which is the square of the impulse response.

The most commonly used ray-based methods are ray tracing [16], [17] and the image-source method [18], [19]. The basic distinction between these methods is the way reflection paths are found. To model the ideal impulse response all the possible sound reflection paths should be discovered. The image-source method finds all the paths, but the computational requirements are such that in practice only a set of early reflections are found. The maximum achievable order of reflections depends on the room geometry and the available calculation capacity. Ray tracing applies the Monte Carlo simulation technique to sample these reflection paths, and thus it gives a statistical result. By this technique also higher order reflections can be searched, though there are no guarantees that all the paths will be found.

2.3.1 Ray Tracing

Ray tracing is a well-known algorithm in simulating the high-frequency behavior of an acoustic space [16], [17], [81], [82]. There are several variations of the algorithm, which are not all covered here. In the basic algorithm the sound source emits sound rays, which are then reflected at surfaces according to certain rules, and the listener keeps track of which rays have penetrated it as audible reflections. The most common reflection rule is the specular reflection. More advanced rules which

include, for example, some diffusion algorithm have also been studied (see [11], [83]). The listeners are typically modeled as volumetric objects, like spheres or cubes, but the listeners may also be planar. In theory a listener can be of any shape as long as there are enough rays to penetrate the listener to achieve statistically valid results. In practice a sphere is in most cases the best choice, since it provides an omnidirectional sensitivity pattern and is easy to implement.

2.3.2 Image-Source Method

From the computational point of view the image-source method is also a ray-based method. The method is thoroughly explained in many papers [18], [19], [84], [85]. In [86] Heinz presented how the image-source method can be used together with a statistical reverberation for binaural room simulation.

The basic principle of the image-source method is presented in Fig. 9. To find reflection paths from the sound source to the listener the source is reflected against all surfaces in the room. Fig. 9 contains a section of a simplified concert hall, consisting of floor, ceiling, back wall, and balcony. Image sources S_c and S_f represent reflections produced by the ceiling and the floor. There is also a second-order image source S_{fc}, which represents sounds reflected first from the floor and then from the ceiling. After finding the image sources, a visibility check must be performed. This indicates whether an image source is visible to the listener or not. This is done by forming the actual reflection path (P_c, P_{fc}, and P_f in Fig. 9) and checking that it does not intersect any surface in the room. In Fig. 9 the image sources S_c and S_{fc} are visible to the listener L instead of image source S_f, which is hidden by the balcony since the path P_f is intersecting it. It is important to notice that locations of the image sources are not dependent on the listener's position and only the visibility of each image source may change when the listener moves. The basic image-source method does not, however, take into account diffraction and diffusion. If these phenomena are to be included in the simulation, the visibility checking will be more complex.

There are also hybrid models in which ray tracing and the image-source method are used together. Typically first reflections are calculated with image sources due to the method's accuracy in finding reflection paths. To avoid the exponential growth of the amount of image sources, later reflections are handled with ray tracing [87], [88], [83].

2.4 Room Acoustics Modeling Methods in DIVA System

In the DIVA system we use various real-time and non-real-time techniques to model room acoustics, as illustrated in Fig. 10.

Performance issues play an important role in the making of a real-time application, and therefore there are few alternative modeling methods available. The real-time auralization algorithm of the DIVA system uses the traditional approach: the image-source method to calculate the early reflections and an artificial late reverberation algorithm to simulate the diffuse reverberant

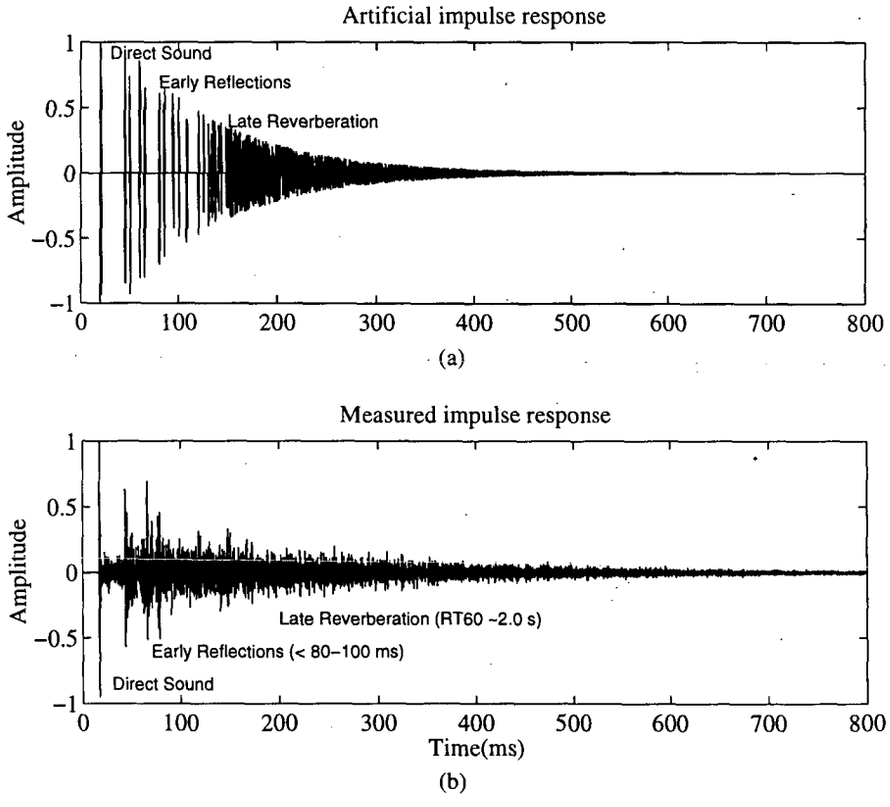


Fig. 8. (a) Imitation of an impulse response of a concert hall. In a room impulse response simulation, the response is typically considered to consist of three separate parts: direct sound, early reflections, and late reverberation. In the late reverberation part the sound field is considered diffuse. (b) Measured response of a concert hall.

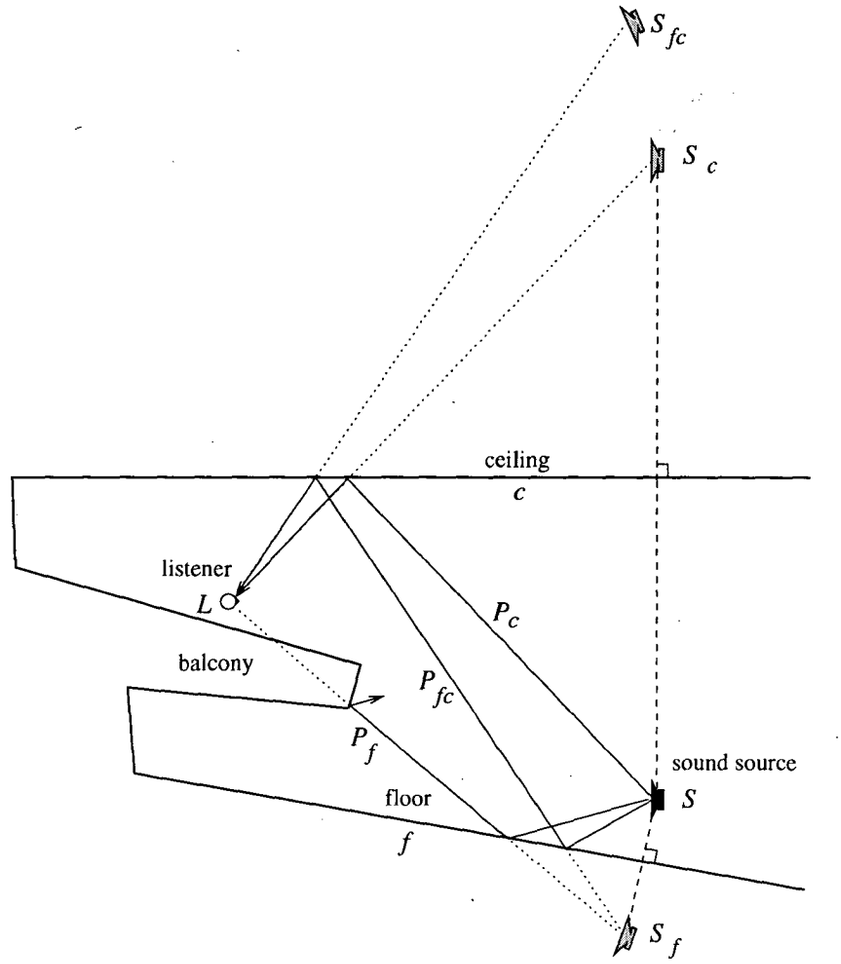


Fig. 9. Sound source is mirrored against all surfaces to produce image sources which represent the corresponding reflection paths. Image sources S_c and S_{fc} representing first- and second-order reflections from ceiling are visible to listener L while reflection from floor P_f is obscured by balcony.

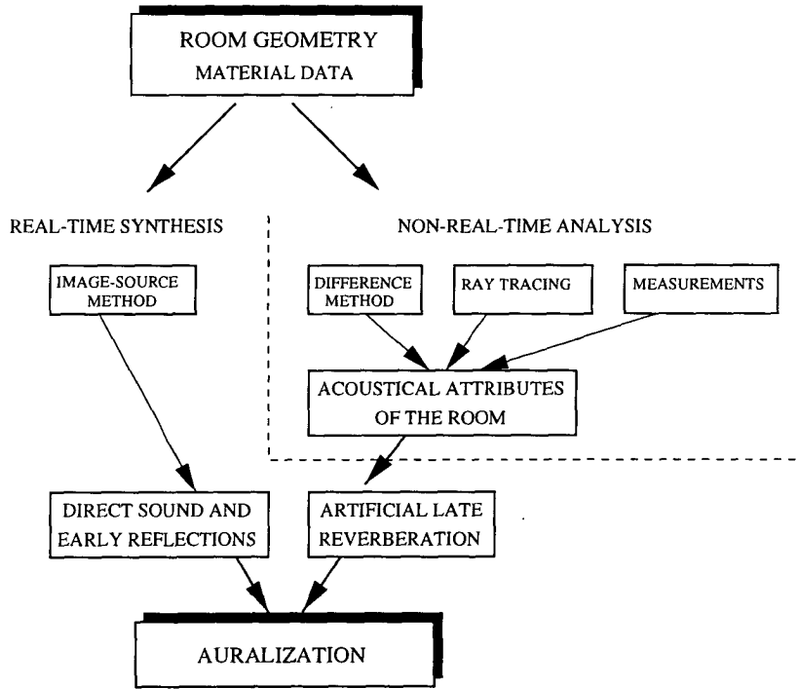


Fig. 10. Computational methods used in DIVA system. Model is a combination of real-time image-source method and artificial late reverberation, which is parametrized according to room acoustical parameters obtained by simulation or measurements.

field [89]. The image-source model was chosen since both ray tracing and the digital waveguide mesh method are too slow for real-time purposes.

The artificial late reverberation algorithm (described in Section 2.7) is parametrized based on room acoustical attributes obtained by non-real-time room acoustic simulations or measurements. The non-real-time calculation method is a hybrid one containing a difference method for low frequencies and ray tracing for higher frequencies. In the following sections the applied models are described in more detail. The difference method used in non-real-time simulations is the digital waveguide mesh method already discussed in Section 2.2

2.4.1 The Ray-Tracing Algorithm

In the ray-tracing algorithm of the DIVA system the sound source emits sound rays to predefined directions producing uniform distribution over a sphere. Surface materials have absorption coefficients for octave bands. The sound reflection paths are considered the same for all the frequencies, and thus all frequencies can be treated simultaneously. In addition there is one diffusion coefficient. Diffusion is implemented such that in each reflection the ray is sent either in a specular or in an arbitrary direction, based on the diffusion coefficient. Listeners are modeled as spheres. The algorithm yields an energy–time response in each octave band as the result.

2.4.2 Advanced Visibility Checking of Image Sources

The implemented image-source algorithm is quite traditional and follows the method presented in Section 2.3.2. However, there are some enhancements to achieve a better performance level.

In the image-source method the number of image sources grows exponentially with the order of reflections. However, only some of them actually are in effect because of occlusions. To reduce the number of potential image sources to be handled, only the surfaces that are at least partially visible to the sound source are examined. This same principle is also applied to image sources. The traditional way to examine the visibility is to analyze the direction of the normal vector of each surface. The source might be visible only to those surfaces that have normal pointing toward the source. After that check there are still unnecessary surfaces in a typical room geometry. To enhance the performance further a preprocessing run is performed with ray tracing to statistically check the visibilities of all surface pairs. The result is a Boolean matrix in which item $M(i, j)$ indicates whether surface i is at least partially visible to surface j or not. Using this matrix the number of possible image sources can be reduced remarkably.

One of the most time-consuming procedures in an interactive image-source method is the visibility check of image sources, which must be performed each time the listener moves. This requires a large number of intersection calculations of surfaces and reflection paths. To reduce this we use the advanced geometrical directory EXCELL [90], [91]. The method is based on regular decomposition of space, and it uses a grid directory. The directory is refined according to the distribution of data. The addressing of the directory is performed with an extendible hashing function.

2.5 Air Absorption

The absorption of sound in the transmitting medium (normally air) depends mainly on the distance, temperature, and humidity. There are various factors that participate in the absorption of sound in air [68]. In a typical environment the most important is the thermal relaxation. The phenomenon is observed as an increasing low-pass filtering as a function of distance from the sound source. Analytical expressions for the attenuation of sound in air as a function of temperature, humidity, and distance have been published in, for example, [92], [93].

Based on the standardized equations for calculating air absorption [93], transfer functions for various temperature, humidity, and distance values were calculated, and second-order IIR filters were fitted to the resulting magnitude responses [94]. The results of modeling for six distances from the source to the receiver are illustrated in Fig. 11(a). In Fig. 11(b) the effect of distance attenuation (according to the $1/r$ law) has been added to the air absorption filter transfer functions.

2.6 Material Parameters

The problem of modeling the sound wave reflection from acoustic boundary materials is a complex one. The temporal or spectral behavior of reflected sound as a function of the incident angle, the scattering and diffraction phenomena, and so on, makes it impossible to develop numerical models that are accurate in all aspects. This topic is more thoroughly discussed in [94], for example. For the DIVA system, computationally simple low-order filters were designed. Furthermore the modeling was restricted to only the angle-independent absorption characteristics.

In the DIVA system a precalculated set of all the boundary reflection combinations has been stored to enable efficient calculation. Those filters contain all the possible material combinations occurring in first-, second-, and third-order reflections. The number of required material filters N_f can be calculated from

$$N_f = \frac{\prod_{i=1}^{K} [n + (i - 1)]}{K!} \tag{3}$$

where K is the order of the reflections and n is the number of materials.

The most common characterization of acoustic surface materials is the absorption coefficients, given for octave bands 125, 250, 500, 1000, 2000, and 4000 Hz. The absorption coefficient is the energy ratio of the absorbed and the incident energies, and the relation between the absorption coefficient $\alpha(\omega)$ and the reflectance $R(j\omega)$ is given by

$$\alpha(\omega) = 1 - |R(j\omega)|^2 \tag{4}$$

where $|R(j\omega)| = \sqrt{1 - \alpha(\omega)}$ can be used to obtain the absolute value of the reflectance when absorption coefficients are given. (The negative value of the square root is possible in theory but almost never happens in practice [68].)

The algorithm for realizing cascaded absorption coefficient data with a low-order IIR filter was as follows. First all possible boundary absorption combinations were calculated and transformed into reflectance data. In the second phase the resulting amplitudes were transformed into a complex frequency response by adding a minimum-phase component. A frequency-domain weighted least-squares fitting algorithm was then applied to the complex reflectance data. As a result a vector containing reflection filter coefficients for all surface combinations was stored for use in the system.

In Fig. 12 the magnitude responses of first-order and third-order IIR filters designed to fit the corresponding target values are shown. Each set of data is a combination of two materials (second-order reflection): a) plas-

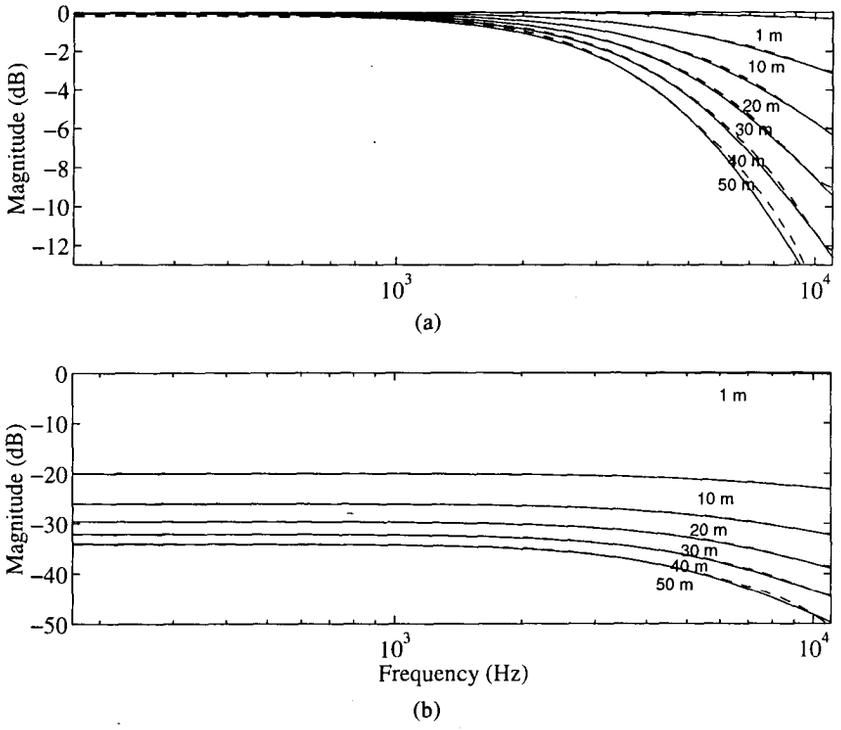


Fig. 11. (a) Magnitude of air absorption filters as a function of distance (1–50m) and frequency. —— ideal response; – – – filter response. For these filters the air humidity is chosen to be 20% and temperature 20°C. (b) Magnitude of combined air absorption and distance attenuation filters as a function of distance (1–50m) and frequency.

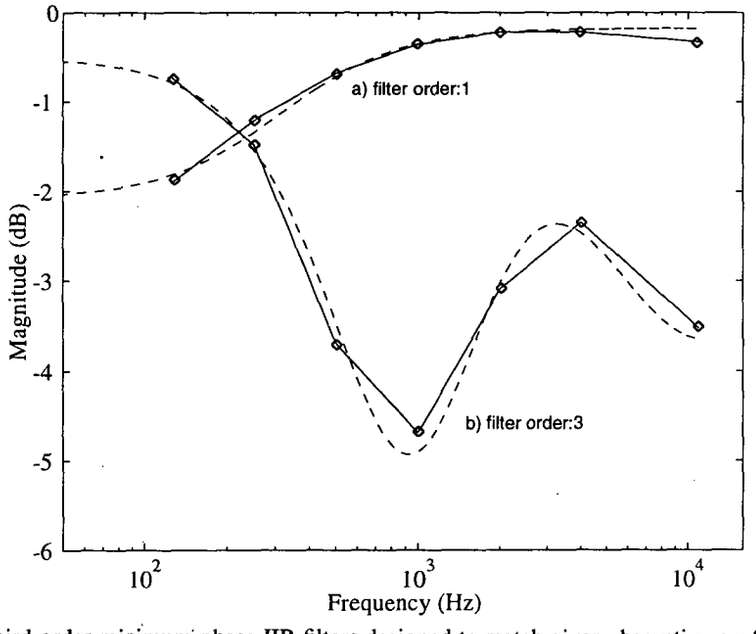


Fig. 12. First-order and third-order minimum-phase IIR filters designed to match given absorption coefficient data. —— target responses; – – – corresponding filter responses.

terboard on frame with 13-mm boards and 100-mm air cavity [95] and glass panel (6 + 2 + 10 mm, toughened, acoustolaminated) [96]; b) plasterboard (same as previously) and 3.5–4-mm fiberboard with holes, 25-mm cavity with 25-mm mineral wool [95].

2.7 Modeling of Late Reverberation

The late reverberant field of a room is often considered nearly diffuse and the corresponding impulse response exponentially decaying random noise [23]. Under these assumptions the late reverberation does not have to be modeled as individual reflections with certain directions. Therefore to save computation in late reverberation modeling, recursive digital filter structures are designed, whose response model the characteristics of real room responses such as the frequency-dependent reverberation time.

The following aims are essential in late reverberation modeling. One is to produce an exponentially decaying impulse response with a dense pattern of reflections to avoid fluttering in the reverberation. A second goal is to simulate the frequency domain characteristics of a concert hall which ideally has a high modal density, especially at low frequencies. Also, in simulating a concert hall, no mode should be emphasized considerably more than the others to avoid coloration of the reverberated sound, or ringing tones in the response. Third, in order to simulate the air absorption and low-pass filtering effect of surface material absorption, the reverberation time has to decrease as a function of frequency. Finally to attain a good spatial impression of the sound field, the late reverberation should produce partly incoherent signals at the listener's ears. Especially lateral reflections, which cause interaural time and level differences between the ears, are significant in producing low interaural coherence. This leads to simulating the diffuseness of the sound field, that is, a situation where the sound reflections arrive from different directions with equal probability. Producing incoherent reverberation with recursive filter structures has been studied, for example, in [23], [97]–[100]. A good summary of reverberation algorithms is presented in [25].

Two basic IIR filter elements are often used in late reverberation modeling, namely, a comb filter and a comb-allpass filter. Both filter types produce decaying pulse trains as their impulse response. The advantage of the use of comb filters in reverberation modeling is that their responses are decaying exponentially and that their reverberation times can be derived [23] from the feedback gain b and the length N of the delay line in samples,

$$T_{60} = \frac{3}{\log|1/b|} \cdot d \tag{5}$$

where $d = N/f_s$, that is, the delay line length in seconds. Strong coloration of individual comb filter responses is a disadvantage caused by regular resonances at the frequencies

$$f = \frac{n}{d} = \frac{n \cdot f_s}{N}, \qquad n = 0, 1, 2, \ldots . \tag{6}$$

Another disadvantage of individual comb filters is that they cause fluttering in the reverberation. To reduce the effect of these drawbacks, several comb filters are connected in parallel, and their delay lengths are chosen so that no reflections from different comb filters coincide at the same instant, and that the resonances do not occur at the same frequencies, causing stronger coloration [23].

A comb-allpass filter, on the other hand, produces a flat steady-state response, but the decay time cannot be controlled like in case of the comb filter response. It can be used to increase the reflection density by connecting it in series with parallel comb filters, for example, like in Schroeder's reverberator [23]. In [24] different combinations of comb and comb-allpass filters in room reverberation simulation are discussed. Other reverberators based on the comb-allpass filters have been studied in Gardner [101], who has developed reverberators based on nested comb-allpass filters.

Other advanced reverberator structures have been proposed by Jot [97] and Rocchesso and Smith [102] among others. These are feedback delay networks (FDN), where several parallel delay lines are feedback connected through a feedback matrix so that each delay line output is connected back to all delay line inputs, as illustrated in Fig. 13. Parallel comb filters are a special case of this structure, resulting from a diagonal feedback matrix. The advantage of feeding each output signal from the delay lines through a matrix is that the resulting impulse response has an increasing reflection density as a function of time, which is the case of the responses of real rooms. Another advantage is that the outputs of the delay lines are mutually incoherent, yet contain the energy of the modes of all the delay lines. Therefore it is possible to produce multichannel reverberation with a single FDN structure. The decay time of an FDN response can be defined in the same way as for parallel comb filters, that is, gains are added in context with each delay line, and their values are solved from Eq. (5).

The frequency-dependent reverberation time caused by air absorption is implemented by adding low-order low-pass filters in context with the feedback-connected delay lines. Fig. 14(a) shows a block diagram of a first-order IIR all-pole filter typically used for this purpose. The magnitude response of the filter [see Fig. 14(b)] implies that the attenuation of the feedback connection is greater for high frequencies, which results in decreasing reverberation time as a function of frequency. The gains g of the low-pass filter can be optimized to match a desired magnitude response curve computed, for example, from a frequency-dependent reverberation time of a real room impulse response.

Fig. 15 presents the reverberator structure we used in the DIVA system [103]. This reverberator contains n parallel feedback loops, where a comb-allpass filter is in each loop. It is a simplification of an FDN structure, and can also be presented in an FDN form containing $2 \cdot n$ delay lines. The comb-allpass filters in the feedback loops, denoted by $A_n(z)$ in Fig. 15, are added to produce an increasing reflection density, thus reducing the per-

ceived fluttering when computation of only a few delay lines can be afforded. The filters $H_n(z)$ implement the frequency-dependent reverberation time. Each contains a simple all-pole first-order low-pass filter depicted in Fig. 14(a), followed by a gain b_n derived from Eq. (5). The gains b_n define the reverberation time at the zero frequency, whereas the low-pass filters cause stronger attenuation at high frequencies. The gains b_n and the gains g_n in the low-pass filters are computed so that the reverberation time decreases as a function of frequency, in a similar way as for all the low-pass filters. We have ended up using a reverberator consisting of 4, 6, or 8 feedback loops, depending on the available computational resources. The lengths of the delay lines in the loops are chosen to be mutually incommensurate in samples to avoid reflections occurring at the same time,

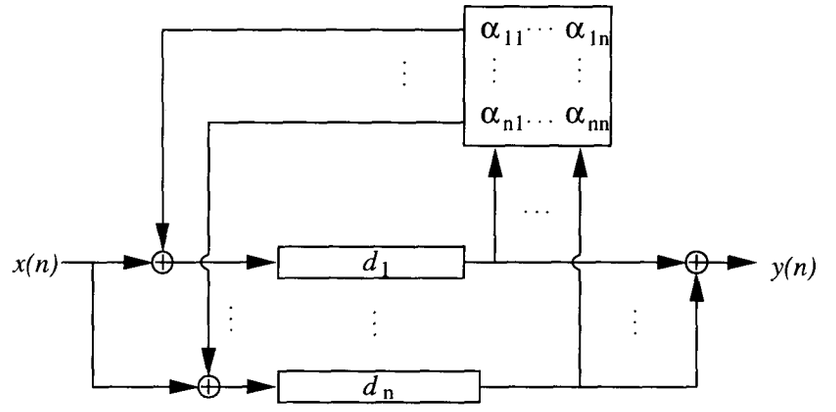
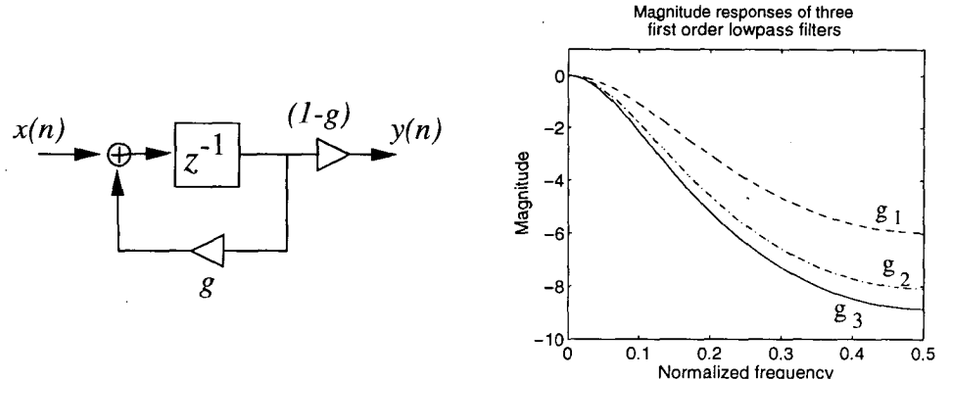


Fig. 13. General structure of FDN. Feedback to each delay line is obtained by calculating a linear combination of delay line output. A scalar multiplication is performed between the output vector of the delay lines and the nth row of the feedback matrix, and the result is fed to the input of the nth delay line.



(a) (b)

Fig. 14. (a) Block diagram of first-order all-pole low-pass filter often used in feedback loops of a reverberator to implement frequency-dependent reverberation time caused by air absorption. (b) Typical magnitude responses of this filter placed in loops with slightly different delay line lengths. Low-pass filtering increases as a function of gain g. Thus to obtain the same frequency-dependent reverberation time for all delay loops, gain g increases as the delay length increases.

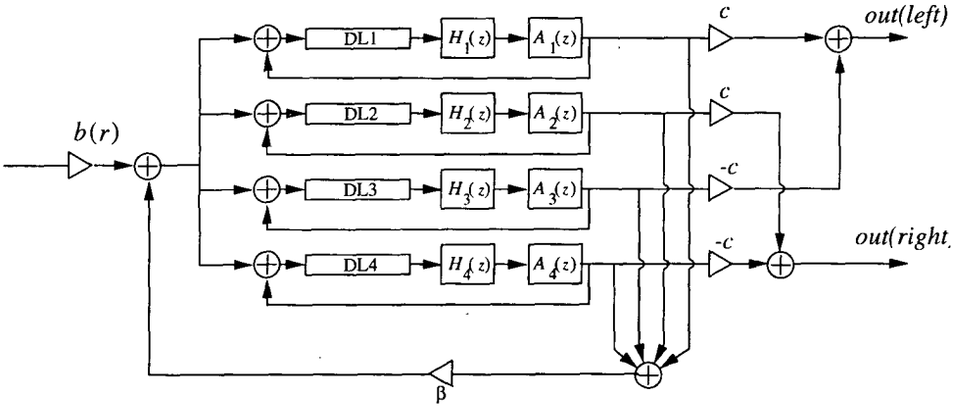


Fig. 15. Block diagram of reverberator used in DIVA system. $A(z)$ and $H(z)$ present comb-allpass and low-pass filters and are explained in Section 2.7.

and strong coloration caused by coinciding modes in the frequency domain [23]. The direct sound is fed to the late reverberation module after the propagation delay and air absorption filtering. The input to the reverberator may also be modified by processing the direct sound with a filter that simulates the diffuse field transfer function of the room [10].

3 LISTENER MODELING AND THREE-DIMENSIONAL SOUND REPRODUCTION

In this chapter methods for producing real-time three-dimensional audio from the modeling data presented in previous chapters are discussed. The main focus is on the binaural technique suitable for headphone or loudspeaker reproduction. Multichannel approaches are also discussed.

3.1 Binaural Reproduction

The binaural technique relies on an accurate reproduction of the cues of spatial sound localization. In a static free-field case the three well-known cues of spatial hearing are [26] 1) the interaural time difference (ITD), 2) the interaural level difference (ILD), and 3) the frequency-dependent filtering due to the pinnae, head, and torso of the listener. The combined frequency- (or time-)domain representation of these static localization cues is often referred to as the head-related transfer fundtion (HRTF). Dynamic and non-free-field cues such as head movements or room reflections are also important factors in sound localization. When these effects are properly reproduced in an immersive virtual reality environment, more convincing and accurate spatial sound rendering will result than with traditional free-field HRTF reproduction.

An HRTF represents a free-field transfer function from a fixed point in a space to a point in the test person's ear canal [29], [30]. HRTF measurements may be carried out for an open or a blocked ear canal. In the DIVA real-time environment design we have used HRTFs based on measurements carried out on human subjects [104] (blocked ear canal) and on dummy heads [105], [104].

Let us consider a binaural system for localizing one virtual source. A monophonic time-domain signal $x_m(n)$ is filtered with two HRTFs $H_l(z)$ and $H_r(z)$ to create a perception of a virtual source [Fig. 16(a)]. The binaural filter design problem therefore refers to approximating two ideal HRTF responses $H_l(z)$ and $H_r(z)$ by digital filters $\hat{H}_l(z)$ and $\hat{H}_r(z)$.

In the case of loudspeaker reproduction, the loudspeaker-to-ear transfer functions have to be taken into account in order to design crosstalk canceling filters. In Fig. 16(b) a symmetric listening situation is outlined, and $H_i(z)$ and $H_c(z)$ (i—ipsilateral, c—contralateral) represent the corresponding loudspeaker-to-ear transfer functions. A convenient solution to crosstalk canceling implementation has been proposed by Cooper and Bauck [39]. The method uses a shuffler structure, and in a symmetrical listening situation only two filters are needed. According to [39], the crosstalk canceling filters, although not minimum phase, are of joint minimum phase, that is, they have a common excess phase which is close to a frequency-independent delay. The delay-normalized crosstalk canceling filters are then minimum phase. Thus the shuffling filters may be defined by their magnitude only, and the phase may be calculated using minimum-phase reconstruction.

Many authors have proposed the use of simplified crosstalk cancelers, based, for example, on spherical head models or simple filter structures [39], [106]. In the DIVA system the performance of crosstalk canceling filters based on human HRTF measurements was, however, found better. The well-known problem of a "sweet spot" in loudspeaker listening of binaurally processed audio reduces the usability of crosstalk canceling techniques to systems with a limited amount of listeners and fixed listening positions. An excellent and thorough discussion of crosstalk canceling methods and related filter design issues is presented in [106].

3.2 HRTF Preprocessing

In most cases the measured HRTFs have to be preprocessed in order to account for the effects of the loudspeaker and the microphone (and headphones for binaural reproduction) that were used in the measurement (see, for example, [29], [30]). Simple deconvolution may be applied to eliminate the effects of the measurement chain, but proper care must be taken so that the resulting filters are realizable [29]. At this point the HRTFs are readily usable, but further equalization may be applied in order to obtain a generalized set of filters.

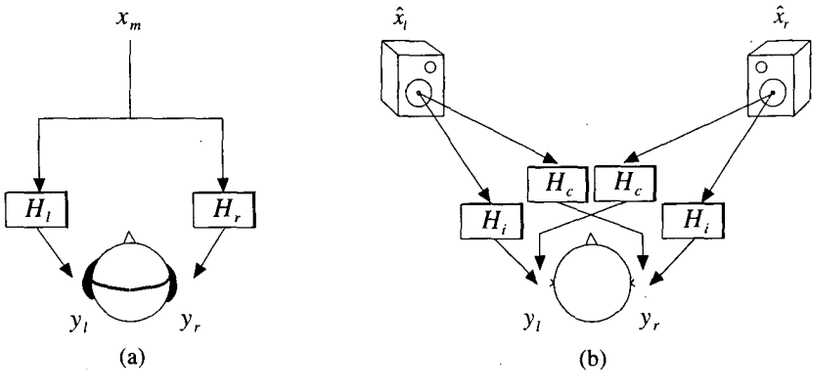


Fig. 16. Audio processing. (a) Binaural. (b) Crosstalk canceled binaural.

Such equalization methods are free-field equalization and diffuse-field equalization [27]. For efficient implementation, the HRTFs can also be divided into an angle-independent part and a directional transfer function (DTF) [33]. The angle-independent features can then be modeled as a general filter for all directions, and as a consequence the DTFs can be designed using lower order models than those used for full HRTFs. Furthermore, functional reconstruction methods such as principal components analysis (PCA) may be used to create an orthonormal transformation of the HRTF database based on singular value decomposition (SVD) [107], [33]. It has been reported that the five first principal components explain approximately 95% of the variance in the HRTF database [33]. The PCA method may prove attractive for HRTF reconstruction due to its low database memory requirements (only the basis filters and weighting functions need to be stored) and is most suitable in systems with multiple sources or listeners, or when using complex ambient phenomena (early reflections, air absorption) [108], [109]. In the following we concentrate on binaural filter design and implementation for room acoustics simulation using a more traditional technique (shown in Fig. 17). This method is based on separate processing of direct sound, early reflections, and late reverberation [24], [89].

3.3 HRTF Filter Design

An attractive property of HRTFs is that they are nearly of minimum phase [109]. The excess phase that is the result of subtracting the original phase response from its minimum-phase counterpart has been found to be approximately linear. This suggests that the excess phase can be implemented separately as an all-pass filter or, more conveniently, as a pure delay. In the case of binaural synthesis, the ITD part of the two HRTFs may then be modeled as a separate delay line, and minimum-phase HRTFs may be used for synthesis. This traditional method is depicted in Fig. 17. Further attractions of minimum-phase systems in binaural simulation are that 1) the energy of the impulse response is optimally concentrated in the beginning, allowing for shortest filter lengths for a specific amplitude response, and 2) due to the previous property, minimum-phase filters are better behaved in dynamic interpolation. Several independent publications have stated that minimum-phase reconstruction does not have any perceptual consequences [33], [110]. This information is crucial in the design and implementation of digital filters for three-dimensional sound.

HRTF filtering constitutes a major part of the computation in binaural virtual acoustic rendering. Therefore it is desirable to seek very efficient filter design and implementation methods for HRTFs [37]. An important topic in HRTF modeling is to take into account the behavior of the human ear. The most important observation here is that whereas filter design normally is carried out on a uniform frequency scale, the human auditory system processes information rather on a logarithmic amplitude and nonlinear frequency scale. Both of these factors should be incorporated in perceptually valid HRTF filter design. Approximations or smoothing of raw HRTF data using auditory criteria have been proposed in the literature by various authors [36], [111]–[118]. The approaches to approximate the nonlinear frequency resolution of the ear can be divided into three categories [37]:

1) Auditory smoothing of the responses prior to filter design

2) Use of a weighting function in filter design

3) Warping of the frequency axis prior to or during the filter design.

An overview of HRTF preprocessing and filter design techniques is given in [119], [37] and only summarized briefly here.

Frequency warping has been shown to be an attractive preprocessing method for HRTF filter design [37]. Warping refers to resampling the spectrum of a transfer function on a warped frequency scale according to nonlinear resolution (such as critical band [120] or ERB [121]). The IIR design method used by Jot et al. [36] involves warping of the frequency scale prior to filter design. This method was originally proposed for use in audio digital filter design in Smith [122] and refined in [36]. The warping is achieved by using the bilinear

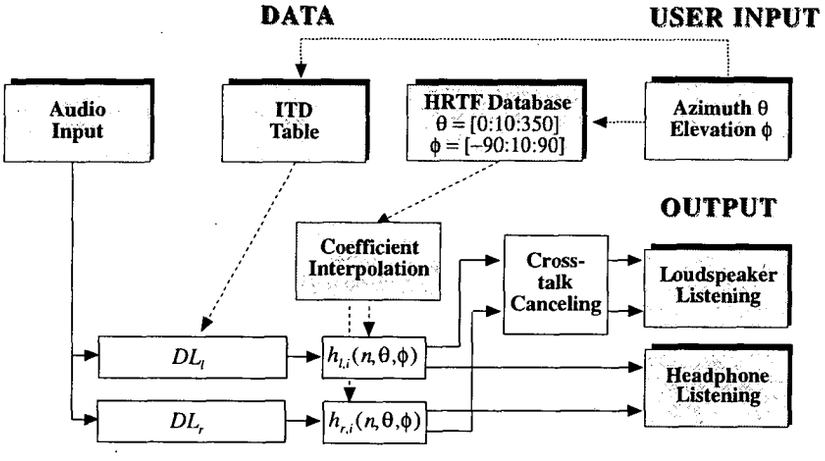


Fig. 17. General implementation framework of dynamic three-dimensional sound using nonrecursive filters. HRTF modeling is carried out using pure delays to represent ITD, and minimum-phase reconstructed head-related impulse responses.

transform implemented with a first-order all-pass filter. The resulting filters have considerably better low-frequency resolution with a tradeoff on high-frequency accuracy, which is tolerable according to the psychoacoustic theory. Warped HRTF design methodology and comparisons with other design techniques have been presented in [117], [37].

3.3.1 HRTF Filters in the DIVA System

In the DIVA system various HRTF filter design aspects have been tested. From the viewpoint of computational efficiency, FIR models appear to be most attractive for dynamic interactive auralization purposes, although warped IIR filters have the best static performance [37]. A least-squares design method using auditory presmoothing (for example, on the ERB scale) has been found to produce satisfactory results. Listening tests that motivate the use of a nonlinear frequency resolution have been carried out in [117], [37]. The conclusions from these studies of HRTF filter design show that the needed filter order for an FIR filter for satisfactory spatial reproduction is approximately 30–35 at the sampling rate of 32 kHz (in the case of nonindividualized HRTFs). In the case of IIR design, the warping method has been mastered, resulting in very low-order filters (orders 15–18).

The ITD implementation follows the principles shown in Fig. 17. Minimum-phase HRTF filters are used and ITD is implemented as a delay line. The horizontal-plane ITDs were derived from human subject HRTFs using linear approximation of interaural excess phase differences [36], [106]. The ITD measurement data have been found to fit well to a spherical-head-based ITD model (discussed in [26], for example). The elevation dependency of the ITD has been taken into account by adding a scaling term to the basic ITD equation

$$\mathrm{ITD} = \frac{a(\sin\theta + \theta)}{2c}\cos\phi \tag{7}$$

where a is the radius of the head, θ is the azimuth angle, ϕ is the elevation angle, and c is the speed of sound. Another approximation method, which includes the elevation angle, has been proposed in [123], but we found a simple cosine dependency of the elevation angle to be accurate enough for our purposes. An example of ITD modeling is shown in Fig. 18.

3.4 Multichannel Reproduction

A natural choice to create a two- or three-dimensional auditory space is to use multiple loudspeakers in the reproduction. With this concept the problems in retaining spatial auditory information are reduced to the placement of the N loudspeakers and panning of the audio signals according to the direction [124]. The problem of multiple-loudspeaker reproduction and implementation of panning rules can be formulated in a form of vector base amplitude panning (VBAP) [45] or by using decoded three-dimensional periphonic systems such as Ambisonics [43], [44]. The VBAP concept introduced in Pulkki [45] gives the possibility of using an arbitrary loudspeaker placement for three-dimensional amplitude panning. In the DIVA system the VBAP method has been implemented for multichannel reproduction [125]. The rapid progress of multichannel surround sound systems for home and theater entertainment during the past decade has opened wide possibilities also

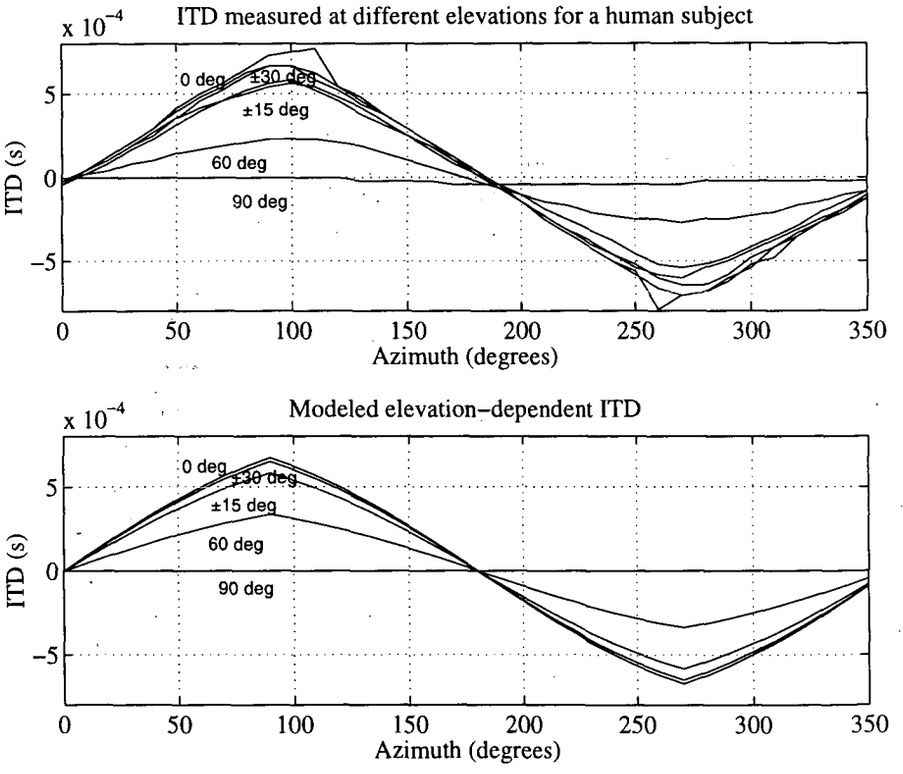


Fig. 18. Measured and simulated ITDs at various elevation angles as a function of azimuth angle.

for multiloudspeaker auralization. Digital multichannel audio systems for movie theaters and home entertainment offer three-dimensional spatial sound that either has been decoded from two-channel material (such as Dolby ProLogic) or uses discrete multichannel decoding (such as Dolby Digital). The ISO/MPEG-2 AAC standard offers 5.1 discrete transparent quality channels at a rate of 320 kbit/s (compression rate 1:12) [54]. Another multichannel compression technique that is already widespread in the market is Dolby Digital which provides similar compression rates to MPEG-2 AAC [54].

A general problem with multichannel ($N > 2$) sound reproduction is the amount of required hardware. On the other hand, intuitively, the listening area should be larger and the localization effect more stable than with two-channel binaural loudspeaker systems. Another advantage of multichannel reproduction over binaural systems is that in multichannel reproduction no listener modeling is required, and thus it is computationally less demanding.

4 INTERACTIVE AURALIZATION

The acoustic response perceived by the listener in a space varies according to the source and receiver positions and orientations. For this reason an interactive auralization model should produce output that depends on the dynamic properties of the source, receiver, and environment. In principle there are two different ways to achieve this goal. The methods presented in the following are called "direct room impulse response rendering" and "parametric room impulse response rendering."

The direct room impulse response rendering technique is based on binaural room impulse responses (BRIRs), which are obtained a priori, either from simulations or from measurements. The BRIRs are defined at a certain grid of listening points. The auralized sound is produced by convolving dry excitation signals with the BRIRs of both ears. In interactive movements this convolution kernel is formed by interpolating the BRIRs of neighboring listening points. Using hybrid frequency–time-domain convolution, it is possible to render long impulse responses with low delay and moderate computational requirements [126]. This method is suitable for static auralization purposes. The setbacks of the method in dynamic simulation are, however, the vast memory requirements for BRIR storage, and the fact that the source, room, and receiver parameters cannot be extracted from the measured BRIR. This means that changes in any of these features will require a new set of BRIR measurements or simulations.

A more robust way for dynamic real-time auralization is to use a parametric room impulse response rendering method. In this technique the BRIRs at different positions in the room are not predetermined. The responses are formed in real time by interactive simulation. The actual rendering process is performed in several parts. The initial part consists of direct sound and early reflections, both of which are time and place variant. The latter part of rendering represents the diffuse reverberant field, which can be treated as a time-invariant filter. In practice this means that the late reverberator can be predetermined, but the direct sound and early reflections are auralized according to parameters obtained by a real-time room acoustic model. The parameters for the artificial late reverberation algorithm are found from measurements or from room acoustic simulations.

The parametric room impulse response rendering technique can be further divided into two categories: the physical approach and the perceptual approach. The physical modeling approach relates acoustical rendering to the visual scene. This involves modeling individual sound reflections off the walls, modeling sound propagation through objects, simulating air absorption, and rendering late diffuse reverberation, in addition to the three-dimensional positional rendering of the source locations. Virtual acoustic rendering can also be approached from a nonphysical viewpoint, investigating the perception of spatial audio and room acoustical quality. This process is termed the perceptual approach to acoustic environment modeling [10]. In the following, the physical approach of parametric virtual acoustics will be described in more detail.

In the DIVA system the main emphasis is in interactive auralization using the physical-model-based parametric room impulse response rendering method (see Fig. 10). By interactive it is implied that users have the possibility to move around in a virtual hall and listen to the acoustics of the room at arbitrary locations in real time.

Performance issues play an important role in the creation of a real-time system. Everything must be done with respect to available computational resources. To make convincing auralization the latency, that is, the time from the user's action to the corresponding change in auralized output, must be small enough. Another problem is how computational resources are distributed in an efficient way to attain the best available quality for auralization.

In the DIVA system we have separate processes for image-source calculation and auralization. The image-source method is used to calculate the direct sound and early reflections. That data are used as parameters for the auralization.

4.1 Auralization in the DIVA System

The structure of the auralization process in the DIVA system is presented in Fig. 19. The audio input is fed to a delay line. This corresponds to the propagation delay from the sound source and each image source to the listener. In the next phase all of the sources are filtered with filter blocks $T_{0 \ldots N}$. This contains the following filters:

- Source directivity filter
- Surface material filters, which represent the filtering occurring at the corresponding reflections (not applied to the direct sound)

- The $1/r$ law distance attenuation
- Air absorption filter.

The signals produced by $T_{0\ldots N}$ are filtered with listener model filters $F_{0\ldots N}$, which make binaural spatialization. To this output is summed the output of the late reverberator R, which is described in Section 2.7.

In a simulation there may be more than one sound source. The image-source method calculates auralization parameters for each direct sound. If the sound sources are close to each other, the sources are grouped together and they produce common image sources. Otherwise each sound source creates its own image sources. If there are N sound sources, each reproducing different sound material, the delay line in Fig. 19 must also be replicated N times.

Furthermore the need for accurate modeling of air absorption and material filtering in the simulation can be questioned. From Fig. 11 it can be seen that the distance attenuation is a more dominant factor than the air absorption, and it is therefore likely that very detailed modeling of air absorption, especially in higher order reflections, is not needed. Material filtering of first-order reflections is very important (because the user can move very close to the boundaries), but the complexity of higher order reflections can be reduced.

4.2 Auralization Parameters

In the DIVA system the listener's position in the virtual room is determined by the GUI (see Fig. 2). The GUI sends the movement data to the room acoustics simulator, which calculates the visible image sources in the space under study. To calculate the image sources the model needs the following information:

- Geometry of the room
- Materials of the room surfaces
- Location and orientation of each sound source
- Location and orientation of the listener.

The orientations of listener and source in the previous list are relative to the room coordinate system. The image-source model calculates the positions and relative orientations of real and image sources with respect to the listener. Data of each visible source are sent to the auralization process. These auralization parameters are:

- Distance from listener
- Azimuth and elevation angles with respect to listener
- Source orientation with respect to listener
- Set of filter coefficients which describe the material properties in reflections.

In the auralization process the parameters affect the coefficients of filters in filter blocks $T_{0\ldots N}$ and $F_{0\ldots N}$ and the pickup point from the input delay line in Fig. 19.

The number of auralized image sources depends on the available computing capacity. In our real-time solution parameters of 10–20 image sources are passed forward.

4.3 Updating the Auralization Parameters

The auralization parameters change whenever the listener moves in the virtual space. The update rate of the auralization parameters must be high enough to ensure that the quality of auralization is not degraded. According to Sandvad [127] rates above 10 Hz should be used. In the DIVA system an update rate of 20 Hz is typically applied.

In a changing environment there are a couple of different possibilities that may cause recalculations of auralization parameters. The main principle in the updating process is that the system must respond within a tolerable latency to any change in the environment. That is reached by gradually refining the calculation. In the first phase only the direct sound is calculated and its parameters are passed to the auralization process. If there are no other changes waiting to be processed, first-order reflections are calculated, then second order, and so on.

In Table 1 the different cases concerning image-source updates are listed. If the sound source moves, all image sources must be recalculated. The same also applies to the situation when reflecting walls in the environment move. Whenever the listener moves, the visibilities of all image sources must be validated. The locations of the image sources do not vary, and therefore there is no need to recalculate them. If the listener turns without

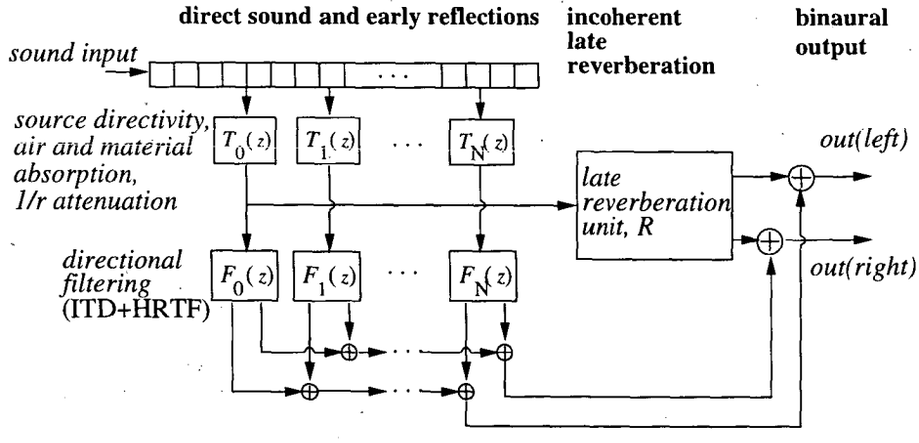


Fig. 19. Example structure of auralization process in DIVA system for headphone listening.

changing position, there are no changes in the visibilities of the image sources. Only the azimuth and elevation angles must be recalculated.

During listener or source movements there are often situations in which some image sources abruptly become visible while others become invisible. This is due to the assumption that sources are infinitely small points and lack diffraction in our acoustic model. The changes in visibilities must be auralized smoothly to avoid discontinuities in the output signal, causing audibly disturbing clicks. The most straightforward method is to fade in the new ones and fade out the ones that become invisible.

Lower and upper limits for the duration of fades are determined by auditory perception. If the time is too short, the fades are observed as clicks. In practice the upper limit is dictated by the rate of updates. In the DIVA system the fades are performed according to the update rate of all auralization parameters. In practice 20 Hz has been found to be a good value.

4.4 Interpolation of Auralization Parameters

In an interactive simulation the auralization parameters change whenever there is a change in the listener's location or orientation in the modeled space. There are various methods of how the changes can be auralized. The topic of interpolating and commuting filter coefficients in auralization systems is discussed, for example, in Jot et al. [36]. The methods described next are applicable if the update rate is high enough, for example, 20 Hz, as in the DIVA system. Otherwise more advanced methods, including prediction, should be used if the latencies are required to be tolerable.

The main principle in all the parameter updates is that the change be performed so smoothly that the listener cannot distinguish the exact update time.

4.4.1 Updating Filter Coefficients

In the DIVA system coefficients of all the filters are updated immediately each time a new auralization parameter set is received. The filters for each image source include:

- Sound source directivity filter
- Air absorption filter
- HRTF filters for both ears.

For the source directivity and air absorption the filter coefficients are stored with such dense grid that there is no need to do interpolation between data points. Instead, the coefficients of the closest data point are utilized. The HRTF filter coefficients are stored in a table with a grid of azimuth $\theta_{grid} = 10°$ and elevation $\phi_{grid} = 15°$ angles. This grid is not dense enough that the coefficients of the nearest data point could be used. Therefore the coefficients are calculated by bilinear interpolation from the four nearest available data points. Since the HRTFs are minimum-phase FIRs, this interpolation can be done [14]. The interpolation scheme for point E located at azimuth angle θ and elevation ϕ is

$$h_E(n) = (1 - c_\theta)(1 - c_\phi)h_A(n) + c_\theta(1 - c_\phi)h_B(n) + c_\theta c_\phi h_C(n) + (1 - c_\theta)c_\phi h_D(n) \quad (8)$$

where h_A, h_B, h_C, and h_D are h_E's four neighboring data points, as illustrated in Fig. 20; c_θ is the azimuth interpolation coefficient $(\theta \bmod \theta_{grid})/\theta_{grid}$, and n goes from 1 to the number of taps of the filter. The elevation interpolation coefficient is obtained similarly, $c_\phi = (\phi \bmod \phi_{grid})/\phi_{grid}$.

4.4.2 Interpolation of Gains and Delays

All the gains and delays are linearly interpolated and changed at every sound sample between two updates. These interpolated parameters for each image source are:

- Distance attenuation gain ($1/r$ law)
- Fade-in and fade-out gains
- Propagation delay
- ITD.

The interpolation in different cases is illustrated in

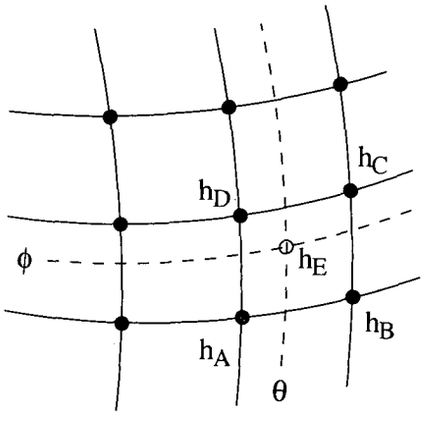


Fig. 20. HRTF filter coefficients corresponding to point h_E at azimuth θ and elevation ϕ are obtained by bilinear interpolation from measured data points h_A, h_B, h_C, and h_D.

Table 1. Required recalculations of image sources in interactive system.

	Recalculate Locations	Recheck Visibilities	Update Orientations
Change in room geometry	×	×	×
Movement of sound source	×	×	×
Turning of sound source			×
Movement of listener		×	×
Turning of listener			×

Figs. 21, 22, and 23. In all of the examples the update rate of the auralization parameters and thus also the interpolation rate is 20 Hz, that is, all interpolations are done within the period of 50 ms. Linear interpolation of the gain factor is straightforward. This technique is illustrated in Fig. 21, where the gain is updated at 0, 50, 100, and 150 ms from values of A_0 to A_3.

The interpolation of delays, namely, the propagation delay and ITD, deserves a more thorough discussion. Each time the listener moves closer to or further from the sound source, the propagation delay changes. In terms of implementation it means a change in the length of a delay line. In the DIVA system the interpolation of delays is done in two steps. The technique applied is presented in Fig. 22. The figure represents a sampled signal in a delay line. The delay changes linearly from D_1 to the new value of D_2 such that the interpolation coefficient τ_1 goes linearly from 0 to 1 during the 50-ms interpolation period and D represents the required delay at each instant. In the first step the interpolated delay D is rounded so that two neighboring samples are found (samples s_1 and s_2 in Fig. 22). In the second step a first-order FIR fractional delay filter with coefficient τ_2 is used to obtain the final interpolated value (s_{out} in Fig. 22).

An accurate implementation of fractional delays would need a higher order filter [128]. The linear interpolation is found to be good enough for our purposes, although it introduces some low-pass filtering. To minimize the low-pass filtering the fractional delays are applied only when the listener moves. At other times the sample closest to the exact delay is used to avoid low-pass filtering. This same technique is applied with the ITDs.

Fig. 23 gives a practical example of the interpolation of delay and gain. There are two updates at 50 and 100 ms. By examining the waveforms one can see that without interpolation there is a discontinuity in the signal while the interpolated signal is continuous.

The applied interpolation technique also enables the Doppler effect in fast movements [49] of the source or the listener. Without interpolation of the delay each update would likely produce a transient sound. A constant update rate of the parameters is essential to produce a natural sounding Doppler effect. Otherwise some fluctuation is introduced to the perceived sound. This corresponds to a situation where the observer moves at alternating speed.

5 IMPLEMENTATION OF THE DIVA VIRTUAL ORCHESTRA SYSTEM

The complete DIVA system was demonstrated for the first time as an interactive installation at the SIGGRAPH'97 conference [47]. The system in Fig. 2

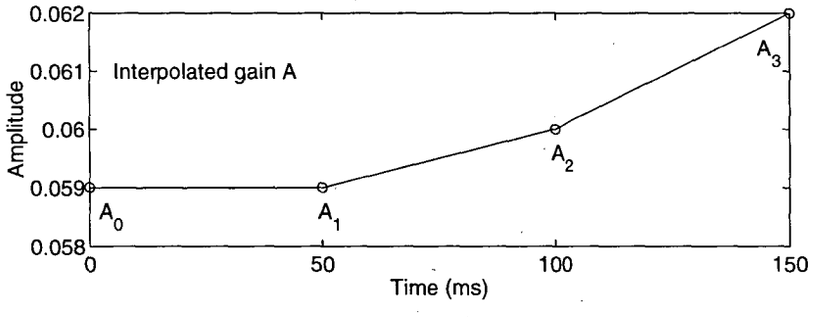


Fig. 21. Interpolation of amplitude gain A ($= 1/r$). Interpolation is done by first-order Lagrange interpolation between key values of gain A, which are marked by circles.

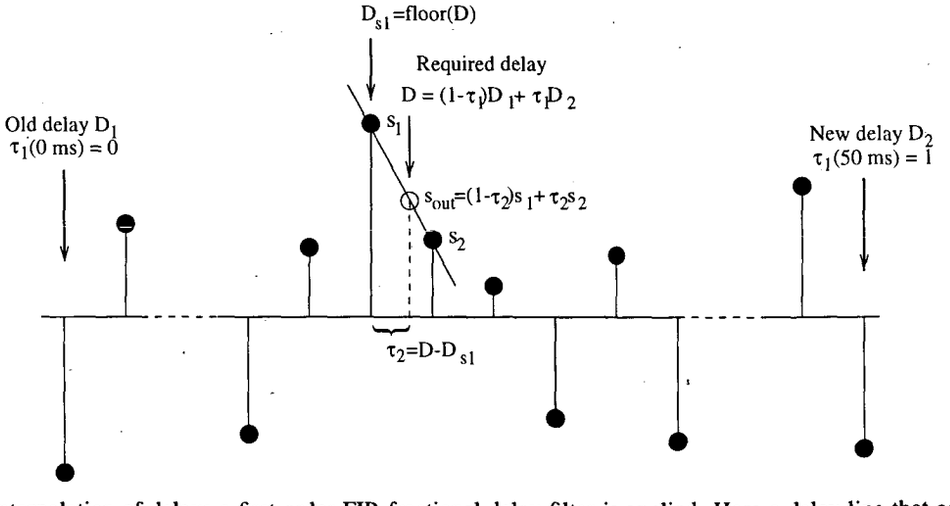


Fig. 22. In the interpolation of delays a first-order FIR fractional delay filter is applied. Here a delay line that contains samples (. . . , s_1, s_2, . . .) and output value s_{out} corresponding to the required delay D is found as a linear combination of s_1 and s_2.

represents the actual setup used there. During the conference over 700 attendees conducted the virtual orchestra and many more visitors listened to the auralized performance with headphones. All experimenters were enthusiastic and gave positive feedback. The orchestra consisted of four virtual musicians: a flutist, a guitarist, a bass player, and a drummer. Fig. 24 is a screenshot of the orchestra playing in a subway station.

The hardware setup for presentation consisted of three workstations (two Silicon Graphics Octanes and one Silicon Graphics O2) and a magnetic tracking device (Ascension Motion Star) connected with an Ethernet network. All the communication was implemented using datagram sockets.

All the software was written using the C++ programming language, and the GUI was based on the OpenGL graphics library. The listener controlled the GUI with a mouse interface. The magnetic tracker was reserved for the conductor.

The sound synthesis was done with digital waveguide models of instruments designed at HUT [72], [129], [130], except that one MIDI synthesizer was required to produce drum sounds.

5.1 Latency

The effects of the update rate of the auralization parameters, latency, and the spatial resolution of the HRTFs on the perceived quality have been studied in Sandvad [127] and Wenzel [131], [132], among others. From the perceptual point of view the most significant parameters are the update rate and latency. The two

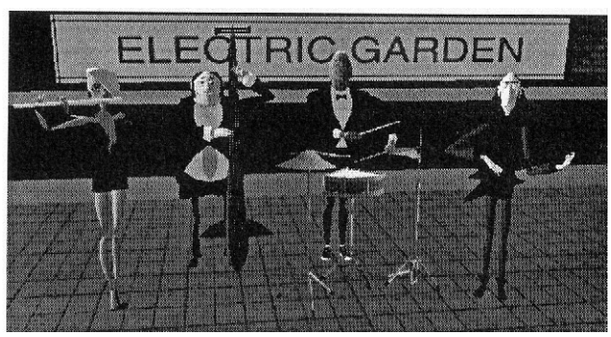


Fig. 24. DIVA virtual orchestra consisting of a flutist, a guitarist, a bass player, and a drummer playing their favorite tune "Kalinka."

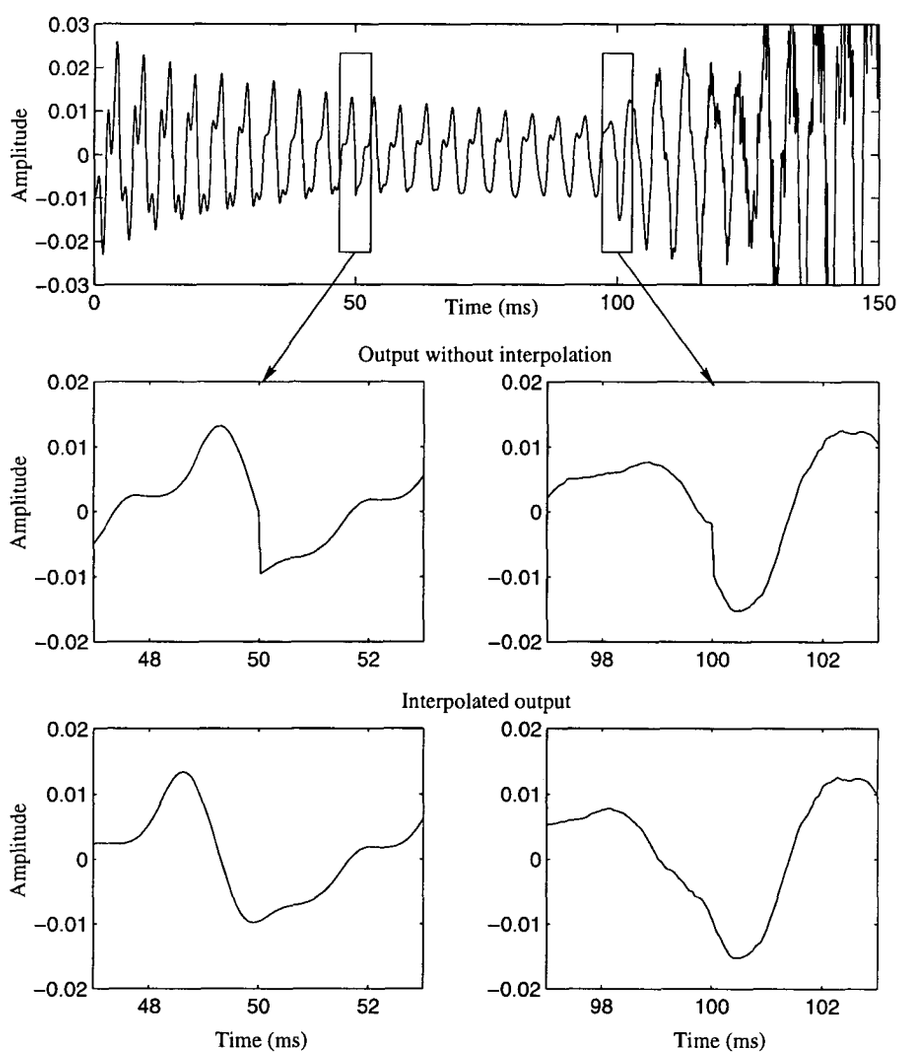


Fig. 23. Need for interpolation of delay can be seen at updates occurring at 50 and 100 ms. In the first focused set there is a clear discontinuity in the signal at those times. The lower set is the continuous signal obtained with interpolation.

are not completely independent variables, since a slow update rate always introduces additional time lag. These studies focus on the accuracy of localization, and Sandvad [127] states that the latency should be less than 100 ms. In the DIVA system the dynamic localization accuracy is not a crucial issue, the most important factor being the latency between visual and aural outputs when either the listener or some sound source moves, thus causing a change in the auralization parameters. According to observations made with the DIVA system this time lag can be slightly larger than 100 ms without a noticeable drop in perceived quality. Note that this statement holds only for updates of auralization parameters in this particular application; in other situations the synchroneity between visual and aural outputs is much more critical, such as lip synchroneity in facial animations or if more immersive user interfaces such as head-mounted displays are used (see, for example, [133], [134]).

The major components of latency in the auralization of the DIVA system are the processing and data transfer delays and bufferings. The total accumulation chain of these is shown in Fig. 25. The latency estimates presented in the following are based on simulations made with a configuration similar to that described. The numbers shown in Fig. 25 represent typical values, not the worst-case situations.

5.1.1 Delays in Data Transfers

There are three successive data transfers before a user's movement is heard as a new soundscape. The transfers are:

- GUI sends data of user's action to image-source calculation
- Image-source calculation sends new auralization parameters to auralization unit
- Auralization sends new auralized material to sound reproduction.

The two first data transfers are realized by sending one datagram message through the Ethernet network. The typical duration of one transfer is 1–2 ms. Occasionally much longer delays of up to 40 ms may occur. Some messages may even get lost or duplicated due to the communication protocol chosen. Fortunately in an unoccupied network these cases are rare. The third transfer is implemented as a memory to memory copy instruction inside the computer and the delay is negligible.

5.1.2 Buffering

In the auralization unit the processing of audio is buffered. The reason for such audio buffering lies in the system performance. Buffered reading and writing of audio sample frames is computationally cheaper than doing these operations sample by sample. Another reason is the UNIX operating system. Since the operating system is not designed for strict real-time systems and the system is not running in a single-user mode, there may be other processes which occasionally require resources. Currently an audio buffering for reading, processing, and writing is done with an audio block size of 50 ms. The latency introduced by this buffering is between 0 and 50 ms due to the asynchronous updates of auralization parameters.

In addition to buffered processing of sound material, the sound reproduction must also be buffered due to the same reasons described. Currently an output buffer of 100 ms is used. At worst this can introduce an additional 50-ms latency, when processing is done with a 50-ms block size.

5.1.3 Delays Caused by Processing

In the DIVA system there are two processes, image-source calculation and auralization, which contribute to the latency occurring after the GUI has processed a user's action.

The latency caused by the image-source calculation depends on the complexity of the room geometry and the number of required image sources. As a case study, a concert hall with about 500 surfaces was simulated, and all the first-order reflections were searched. A listener movement causing visibility check for the image sources took 5–9 ms, depending on the listener's location.

The processing time of one audio block in auralization must be on the average less than or equal to the length of the audio block. Otherwise the output buffer underflows, and this is heard as a disturbing click. Thus the delay caused by actual auralization is less than 50 ms in the current DIVA system.

In addition the fades increase the latency. Each change is fully applied only after a complete interpolation period of 50 ms in the DIVA system.

5.1.4 Total Latency

Altogether in a typical situation the DIVA system runs smoothly and produces continuous output with 110–160-ms average latency, as illustrated in Fig. 25. However, in the worst case the latency is more than 200 ms, which may cause a buffer underflow, resulting in a discontinuity in the auralized sound. If less latency is required, the buffers, which cause most of the delays, can be shortened. However, then the risk of discontinuous sound is increased.

The latency depends much on the underlying hardware platform including both computers and the network. With dedicated servers and network the system can run with remarkably shorter delays since the audio buffers can be kept shorter.

6 CASE STUDY: MARIENKIRCHE

In this section an auralization case study is presented. It concerns the implementation of virtual acoustics in the Marienkirche concert hall located in Neubrandenburg, Germany. Originally the building was an old gothic cathedral, but it was left in ruins during World War II. Currently it is being rebuilt as a concert hall [135]. The seating capacity of the hall is approximately 1200 places.

There were two separate goals in the modeling. The first goal was to create an interactive real-time auralization software, the second was to produce a demonstration video [136], which was computed as a batch job.

6.1 Acoustic Model of the Hall

The geometric model of the hall is represented in Fig. 26. The model contains nearly 500 polygons and 14 different surface materials. The same model was used for both the real-time and the off-line tasks.

In a complex space, as illustrated in Fig. 26, only a few of all possible image sources are visible. To obtain the average amount of visible image sources we have calculated the amount of visible ones in 72 cases (4 source and 18 listener points). From that simulation we obtained the result that on the average only about 7 first-order and about 17 second-order image sources are simultaneously visible.

6.2 Parametrization of the Acoustic Model

In the design of air absorption filters the assumed temperature was 20°C and the humidity was 60%. The filters were computed with 1-m spacing from 0 to 60 m. In Fig. 11 filters at distances of 1, 10, 20, 30, 40, and 50 m are illustrated (note that value for humidity is 20% in Fig. 11). Filters were used according to the distance of the direct sound and the image sources, as described in Section 4.4.

The material filters used were IIRs of third order in batch simulation and IIRs of first order in real-time simulation. Two examples of the magnitude responses of the filters are given in Fig. 12. For the first-order reflections 14 material filters were needed. The second-order reflections need 105 filters according to Eq. (3).

To find parameters for the late reverberation unit (in Fig. 15) the ray-tracing algorithm was used. Based on these simulations the reverberation time (RT_{60}) at low frequencies was set to 2.3 s. Calculation of the feedback coefficient of the low-pass filters (the IIR filters inside the late reverberation unit) was then made using Eq. (5). The delay line lengths (DL1–DL4 in Fig. 15) were chosen so that the shortest delay line was slightly shorter than the delay corresponding to the most distant image source. This means that the first output from the late reverberation unit overlaps somewhat with the early reflections represented by the image sources. To avoid overlapping between outputs of different delay lines all delay line lengths used were prime numbers.

6.3 Validation of Model

The validation of the model is difficult because no single objective parameter exists for good concert hall acoustics. In preparation of the demonstration the most important way of validation was informal listening of the auralized result. Based on that the model was refined until all listeners were satisfied with the result.

At present the actual hall is still under construction, so measuring the acoustical attributes of the hall is not possible. The simulation results in Table 2 were compared against the design criteria of the hall and with recommended values presented in Beranek [137]. The simulation data were averaged from two omnidirectional sources and four listener points shown in Fig. 26. There is good agreement between simulated and recommended values. This shows one reason why the auralization of the hall sounds natural.

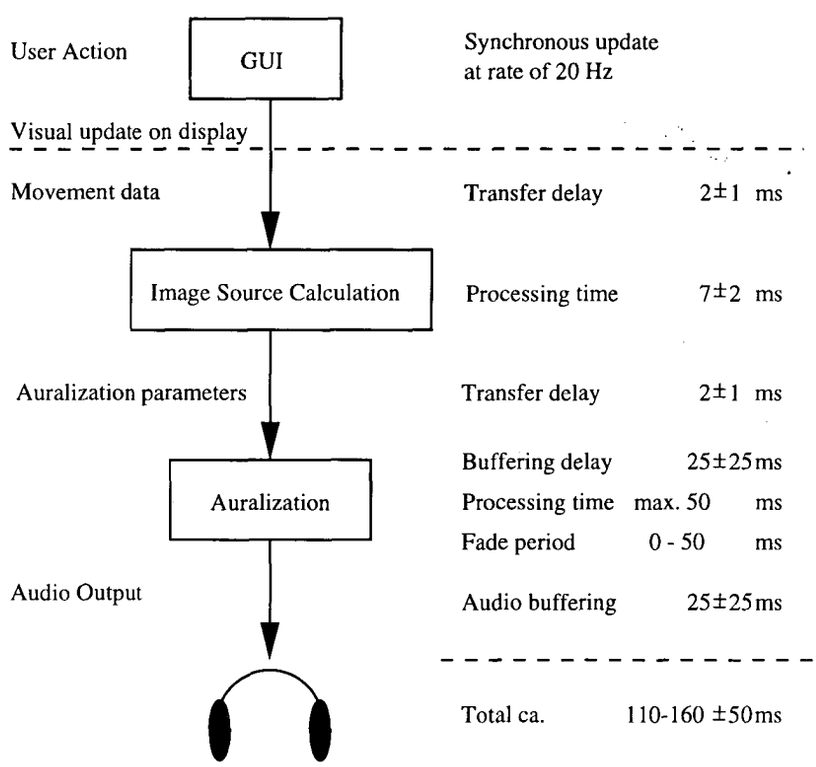


Fig. 25. Various components in DIVA system introduce latency to system. The most significant ones are caused by bufferings.

6.4 Performance

Table 3 presents how the auralization filters of Fig. 19 can be divided according to the source–medium–receiver model described in the Introduction. In a typical simulation the computationally most laborious part is the spatialization, although each part of the model can be refined such that it reserves the major part of the

Table 3. Division of different auralization filters according to source–medium–receiver model.

	Source: Sound Source	Medium: Room Acoustics	Receiver: Listener
$T_{0\ldots N}$	×	×	
R		×	(×)
$F_{0\ldots N}$			×

Table 2. Calculated acoustical attributes from computer model of Marienkirche.*

	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz
RT (s)	2.2	2.2	2.1	2.1	1.9	1.4
EDT (s)	2.1	2.0	2.0	1.9	1.6	1.1
C_{80} (dB)	−1.2	−0.5	−0.7	−0.5	0.4	2.4
ASW (corr.)	0.38	0.30	0.32	0.24	0.30	0.22
LEV (corr.)	0.08	0.08	0.08	0.09	0.06	0.06

* All values are averaged from eight impulse responses (two omnidirectional sources and four listener points).

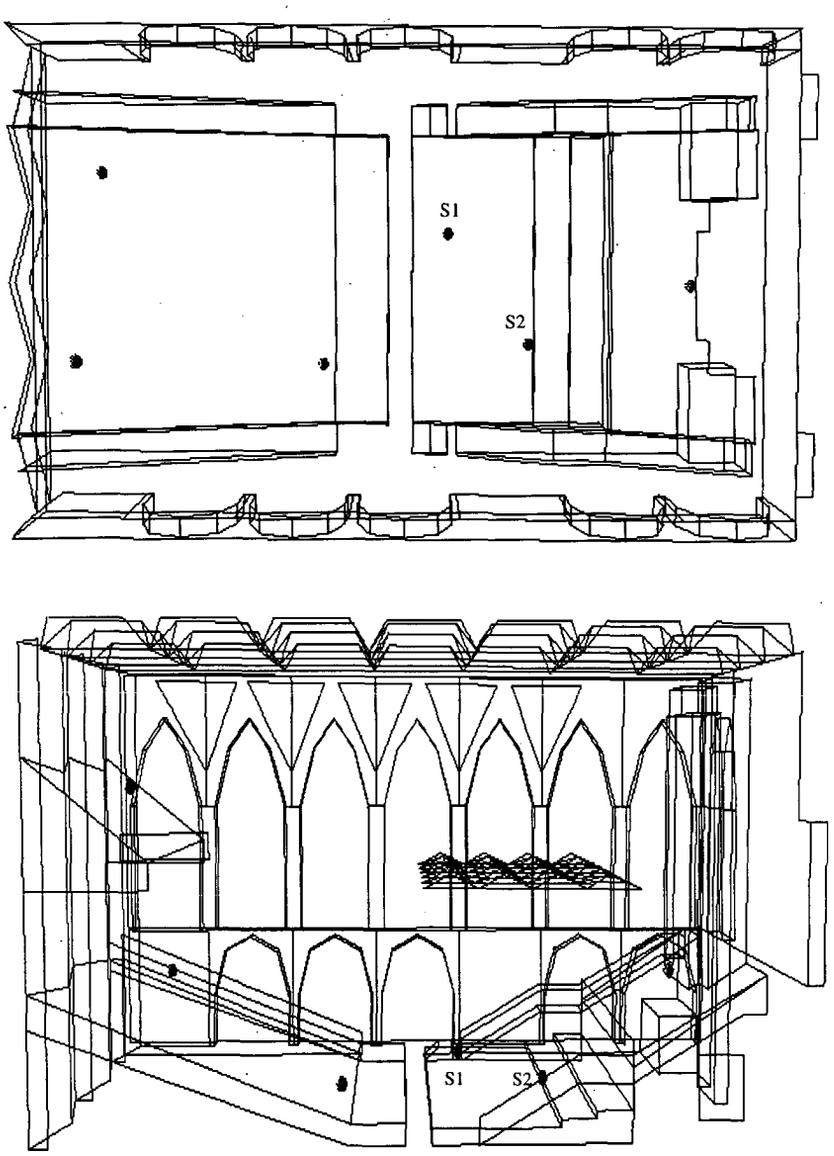


Fig. 26. In case study, virtual acoustics were designed for Marienkirche concert hall. Geometric model of hall consists of nearly 500 polygons. Model is seen from above and from one side. Source locations S1 and S2 and listening points are marked.

available resources.

In the real-time simulation all first-order reflections were auralized. The system runs on a sampling rate of 32 kHz. Using a Silicon Graphics Octane workstation the direct sound and 8 image sources can be auralized. To obtain an accurate spatialization for direct sound, HRTF filters with 30-tap FIR approximation were used. For image sources the computational load can be reduced with two simplifications. First the image sources are grouped (as proposed by Heinz [86]) so that, for example, the HRTFs of (azimuth = 10°, elevation = 0°) can be used for angles (5° . . . 14°, −7° . . . 7°). In that case the interpolation of HRTFs is not needed. Second, when minimum-phase FIRs for HRTFs are used, the filtering can be done, for example, with 10 first taps only (when 30-tap FIRs are designed). This simplification is possible since almost all energy of a minimum-phase FIR is in the beginning of the impulse response of the filter. Both of these simplifications can be accepted if the spatialization of the direct sound is done accurately. By reducing the quality of the spatialization the number of auralized image sources could be increased such that all first- and second-order reflections could be auralized.

In the demonstration video [136] in addition to the first-order reflections, all second-order reflections were also auralized. The spatialization was done using HRTFs with 60 taps for direct sound as well as for image sources. Also the sampling rate was higher, namely, 48 kHz. The computation took approximately five times the actual duration of the soundtrack.

The distribution of the computing time was examined by profiling the software. Results were gathered of a simulation run where the listener was in continuous movement for 1 minute. During the run the direct sound was visible all the time, and on the average there were nine visible image sources. The most time-consuming process in the auralization was the HRTF filtering, which took about 50% of the processor capacity. Another computationally laborious part of the calculation was the interpolations, which used about 25% of processor capacity. The remaining 25% was divided between filterings of material and air absorption (10%), diffuse late reverberation algorithms (about 8%), and updates (7%).

7 FUTURE WORK

The current room acoustics model of the DIVA system is based on geometrical room acoustics. That technique is valid at high frequencies. At low frequencies the diffraction and diffusion are remarkable, and in the DIVA system those are modeled only by the diffuse late reverberation algorithm. For non-real-time simulations a frequency-domain hybrid might also be a possible solution. Using the DIVA system this can be implemented by combining an FDTD model for low-frequency simulations with the current system.

The DIVA system consists of many different parts, which all affect the observed quality of the audio output. Therefore listening tests should be carried out to tell which parts of the system require most enhancements. Factors to be tested include:

- Number of auralized image sources
- Accuracy of material filters
- Accuracy of spatialization filters (HRTFs)
- Synchroneity mismatch between visual and aural outputs
- Directivity of image sources.

The listening tests of dynamic virtual acoustic environments carried out by other research groups concentrate on the accuracy of localization (see, for example, [127], [131], [138]), while our goal is to achieve as natural a listening experience as possible.

We are currently building a virtual environment laboratory similar to the CAVE system [139], where large-screen projectors show visual stereoscopic images on each wall of the room and three-dimensional audio can be heard through either headphones or loudspeakers using multichannel reproduction. The user may freely move within the room and control the virtual world with a position tracker. Conjectured future applications include the use of the system in architectural acoustics design.

8 CONCLUSIONS

The general auralization process consists of modeling three separate parts: the sound source, the room acoustics, and the listener. In this paper we discussed a few modeling techniques of each component. The main emphasis has been on real-time modeling.

We have developed the software system DIVA for producing interactive virtual audiovisual performances in real time. The whole processing chain from sound synthesis through sound propagation in a room to spatialization at the listener's ears is implemented and combined with synchronized animation of virtual musicians conducted by a human conductor.

Room acoustics is simulated using a time-domain hybrid model in which the direct sound and early reflections are obtained by the image-source method and the late reverberation is modeled using a recursive digital filter structure.

The listener can move freely in the virtual space. The direct sound and early reflections are each spatialized using several alternative solutions. The best results are obtained by using accurate HRTFs, but simpler approximations of the interaural level and time differences (ILD and ITD) can also be used.

The interpolation of the parameters of each reflection (such as propagation delay, distance attenuation, ITD) enables interactive movement of the listener so that the auditory scene sounds natural. In fact if the listener moves fast, the listener can hear a realistic Doppler effect that is quite impressive.

In addition to the virtual orchestra the technologies of the DIVA system can be used for various purposes such as games, multimedia, and virtual reality applications and architectural acoustics design.

9 ACKNOWLEDGMENT

The authors would like to thank Prof. Tapio Takala, Prof. Matti Karjalainen, Dr. Vesa Välimäki, the personnel of the DIVA project, and the personnel of the Laboratory of Acoustics and Audio Signal Processing at the Helsinki University of Technology for support of the project. The authors also wish to thank Dr. Tapio Lahti and Mr. Henrik Möller (Akukon Oy) for providing the geometrical concert hall models. Furthermore we are very grateful to Akukon Oy and Pekka Salminen Architects for inviting us to cooperate in the Marienkirche project. This project has been partially financed by the Academy of Finland, Nokia Research Center, and the Technology Development Centre of Finland (TEKES).

10 REFERENCES

[1] ISO/IEC JTC1/SC29/WG11 IS 14496, "Information Technology–Coding of Multimedia Objects (MPEG-4)," (1999).

[2] ISO/IEC JTC/SC24 IS 14772-1, "Information Technology—Computer Graphics and Image Processing—The Virtual Reality Modeling Language (VRML97)" (1997 Apr.). URL: http://www.vrml.org/Specifications/VRML97/.

[3] SUN, Inc., "JAVA 3D API Specification 1.1" (1998 Dec.). URL: http://java.sun.com/products/java-media/3D/forDevelopers/j3dguide/j3dTOC.doc.html.

[4] W. F. Dale, "A Machine-Independent 3D Positional Sound Application Programmer Interface to Spatial Audio Engines," in *Proc. AES 16th Int. Conf. on Spatial Sound Reproduction* (Rovaniemi, Finland, 1999 Apr.), pp. 160–171.

[5] Interactive Audio Special Interest Group. URL: http://www.iasig.org.

[6] E. Wenzel, "Spatial Sound and Sonification," presented at the International Conference on Auditory Display (ICAD'92). Also in *Auditory Display: Sonification, Audification, and Auditory Interface, SFI Studies in the Sciences of Complexity*, Proc. XVIII, G. Kramer, Ed. (Addison-Wesley, Reading, MA, 1994).

[7] D. Begault, *3-D Sound for Virtual Reality and Multimedia* (Academic Press, Cambridge, MA, 1994).

[8] S. Foster, E. Wenzel, and R. Taylor, "Real-Time Synthesis of Complex Acoustic Environments," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'91)* (New Paltz, NY, 1991).

[9] B. Shinn-Cunningham, H. Lehnert, G. Kramer, E. Wenzel, and N. Durlach, "Auditory Displays," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. Gilkey and T. Anderson, Eds. (Lawrence Erlbaum, Mahwah, NJ, 1997), pp. 611–663.

[10] J.-M. Jot, "Real-Time Spatial Processing of Sounds for Music, Multimedia and Interactive Human–Computer Interfaces," *Multimedia Sys.* (Special Issue on Audio and Multimedia), vol. 7, no. 1, pp. 55–69 (1999).

[11] H. Lehnert and J. Blauert, "Principles of Binaural Room Simulation," *Appl. Acoust.*, vol. 36, pp. 259–291 (1992).

[12] J. P. Vian and J. Martin, "Binaural Room Acoustics Simulation: Practical Uses and Applications," *Appl. Acoust.*, vol. 36, pp. 293–305 (1992).

[13] J. Martin, D. Van Maercke, and J. P. Vian, "Binaural Simulation of Concert Halls: A New Approach for the Binaural Reverberation Process," *J. Acoust. Soc. Am.*, vol. 94, pp. 3255–3264 (1993).

[14] J. Huopaniemi, "Virtual Acoustics and 3-D Sound in Multimedia Signal Processing," Ph.D. thesis, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing (1999).

[15] M. Kleiner, B.-I. Dalenbäck, and P. Svensson, "Auralization—An Overview," *J. Audio Eng. Soc.*, vol. 41, pp. 861–875 (1993 Nov.).

[16] A. Krokstad, S. Strøm, and S. Sørsdal, "Calculating the Acoustical Room Response by the Use of a Ray Tracing Technique," *J. Sound Vib.*, vol. 8, pp. 118–125 (1968).

[17] A. Kulowski, "Algorithmic Representation of the Ray Tracing Technique," *Appl. Acoust.*, vol. 18, pp. 449–469 (1985).

[18] J. B. Allen and D. A. Berkley, "Image Method for Efficiently Simulating Small-Room Acoustics," *J. Acoust. Soc. Am.*, vol. 65, pp. 943–950 (1979).

[19] J. Borish, "Extension of the Image Model to Arbitrary Polyhedra," *J. Acoust. Soc. Am.*, vol. 75, pp. 1827–1836 (1984).

[20] H. Kuttruff, "Sound Field Prediction in Rooms," in *Proc. 15th Int. Congr. on Acoustics (ICA'95)*, vol. 2 (Trondheim, Norway, 1995 June), pp. 545–552.

[21] D. Botteldooren, "Finite-Difference Time-Domain Simulation of Low-Frequency Room Acoustic Problems," *J. Acoust. Soc. Am.*, vol. 98, pp. 3302–3308 (1995).

[22] L. Savioja, J. Backman, A. Järvinen, and T. Takala, "Waveguide Mesh Method for Low-Frequency Simulation of Room Acoustics," in *Proc. 15th Int. Congr. on Acoustics (ICA'95)*, vol. 2 (Trondheim, Norway, 1995 June), pp. 637–640.

[23] M. R. Schroeder, "Natural-Sounding Artificial Reverberation," *J. Audio Eng. Soc.*, vol. 10, pp. 219–223 (1962).

[24] J. A. Moorer, "About This Reverberation Business," *Comput. Music J.*, vol. 3, pp. 13–28 (1979).

[25] W. Gardner, "Reverberation Algorithms," in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds. (Kluwer Academic, Boston, MA, 1997), pp. 85–131.

[26] J. Blauert, *Spatial Hearing. The Psychophysics of Human Sound Localization*, 2nd ed. (MIT Press, Cambridge, MA, 1997).

[27] H. Møller, "Fundamentals of Binaural Technology," *Appl. Acoust.*, vol. 36, pp. 171–218 (1992).

[28] G. Kendall, "A 3-D Sound Primer: Directional Hearing and Stereo Reproduction," *Comput. Music J.*, vol. 19, no. 4, pp. 23–46 (1995 Winter).

[29] F. L. Wightman and D. J. Kistler, "Headphone Simulation of Free-Field Listening. I: Stimulus Synthe-

sis," *J. Acoust. Soc. Am.*, vol. 85, pp. 858–867 (1989).

[30] H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen, "Head-Related Transfer Functions of Human Subjects," *J. Audio Eng. Soc.*, vol. 43, pp. 300–321 (1995 May).

[31] G. S. Kendall and W. L. Martens, "Simulating the Cues of Spatial Hearing in Natural Environments," in *Proc. 1984 Int. Computer Music Conf.* (Paris, France, 1984), pp. 111–125.

[32] F. Asano, Y. Suzuki, and T. Sone, "Role of Spectral Cues in Median Plane Localization," *J. Acoust. Soc. Am.*, vol. 88, pp. 159–168 (1990).

[33] D. J. Kistler and F. L. Wightman, "A Model of Head-Related Transfer Functions Based on Principal Components Analysis and Minimum-Phase Reconstruction," *J. Acoust. Soc. Am.*, vol. 91, pp. 1637–1647 (1992).

[34] M. A. Blommer and G. H. Wakefield, "On the Design of Pole–Zero Approximations Using a Logarithmic Error Measure," *IEEE Trans. Signal Process.*, vol. 42, pp. 3245–3248 (1994 Nov.).

[35] J. Sandvad and D. Hammershøi, "Binaural Auralization: Comparison of FIR and IIR Filter Representation of HIRs," presented at the 96th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 42, p. 395 (1994 May), preprint 3862.

[36] J. M. Jot, O. Warusfel, and V. Larcher, "Digital Signal Processing Issues in the Context of Binaural and Transaural Sterephony," presented at the 98th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 43, p. 396 (1995 May), preprint 3980.

[37] J. Huopaniemi, N. Zacharov, and M. Karjalainen, "Objective and Subjective Evaluation of Head-Related Transfer Function Filter Design," *J. Audio Eng. Soc.*, vol. 47, pp. 218–239 (1999 Apr.).

[38] M. Schroeder and B. Atal, "Computer Simulation of Sound Transmission in Rooms," in *IEEE Conv. Rec.*, pt. 7 (1963), pp. 150–155.

[39] D. H. Cooper and J. L. Bauck, "Prospects for Transaural Recording," *J. Audio Eng. Soc.*, vol. 37, pp. 3–19 (1989 Jan./Feb.).

[40] K. B. Rasmussen and P. M. Juhl, "The Effect of Head Shape on Spectral Stereo Theory," *J. Audio Eng. Soc.*, vol. 41, pp. 135–142 (1993 Mar.).

[41] W. Gardner, "Transaural 3-D Audio," MIT Media Lab Perceptual Computing, Tech. Rep. 342 (1995).

[42] M. J. Walsh and D. J. Furlong, "Improved Spectral Stereo Head Model," presented at the 99th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 43, p. 1093 (1995 Dec.), preprint 4128.

[43] M. A. Gerzon, "Periphony: With-Height Sound Reproduction," *J. Audio Eng. Soc.*, vol. 21, pp. 2–10 (1973 Jan./Feb.).

[44] D. Malham and A. Myatt, "3-D Sound Spatialization Using Ambisonic Techniques," *Comput. Music J.*, vol. 19, no. 4, pp. 58–70 (1995).

[45] V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *J. Audio Eng. Soc.*, vol. 45, pp. 456–466 (1997 June).

[46] T. Takala, R. Hänninen, V. Välimäki, L. Savioja, J. Huopaniemi, T. Huotilainen, and M. Karjalainen, "An Integrated System for Virtual Audio Reality," presented at the 100th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 44, p. 644 (1996 July/Aug), preprint 4229.

[47] DIVA Group: J. Hiipakka, R. Hänninen, T. Ilmonen, H. Napari, T. Lokki, L. Savioja, J. Huopaniemi, M. Karjalainen, T. Tolonen, V. Välimäki, S. Välimäki, and T. Takala, "Virtual Orchestra Performance," in *Visual Proc. of SIGGRAPH'97* (Los Angeles, CA, 1997), p. 81, ACM SIGGRAPH.

[48] T. Lokki, J. Hiipakka, R. Hänninen, T. Ilmonen, L. Savioja, and T. Takala, "Real-Time Audiovisual Rendering and Contemporary Audiovisual Art," *Organised Sound*, vol. 3, no. 3 (1999).

[49] T. Takala and J. Hahn, "Sound Rendering," Comput. Graphics, SIGGRAPH'92, no. 26, pp. 211–220 (1992).

[50] J. Hahn, J. Geigel, J. W. Lee, L. Gritz, T. Takala, and S. Mishra, "An Integrated Approach to Sound and Motion," *J. Visualiz. and Comput. Animation*, vol. 6, no. 2, pp. 109–123 (1995).

[51] T. Ilmonen, "Tracking Conductor of an Orchestra Using Artificial Neural Networks," Master's thesis, Helsinki University of Technology (1999).

[52] R. Hänninen, "LibR—An Object-Oriented Software Architecture for Realtime Sound and Kinematics," Licentiate thesis, Helsinki University of Technology (1999).

[53] ISO/IEC JTC1/SC29/WG11 IS 14496-3 (MPEG-4), "Information Technology—Coding of Audiovisual Objects. Part 3: Audio" (1999).

[54] K. Brandenburg and M. Bosi, "Overview of MPEG Audio: Current and Future Standards for Low-Bit-Rate Audio Coding," *J. Audio Eng. Soc.*, vol. 45, pp. 4–21 (1997 Jan./Feb.).

[55] J. O. Smith, "Physical Modeling Synthesis Update," *Comput. Music J.*, vol. 20, pp. 44–56 (1996 Summer).

[56] J. Huopaniemi, M. Karjalainen, V. Välimäki, and T. Huotilainen, "Virtual Instruments in Virtual Rooms—A Real-Time Binaural Room Simulation Environment for Physical Models of Musical Instruments," in *Proc. Int. Computer Music Conf. (ICMC'94)* (Aarhus, Denmark, 1994 Sept.), pp. 455–462.

[57] M. Karjalainen, J. Huopaniemi, and V. Välimäki, "Direction-Dependent Physical Modeling of Musical Instruments," in *Proc. 15th Int. Congr. on Acoustics (ICA'95)* (Trondheim, Norway, 1995 June), pp. 451–454.

[58] J. Meyer, *Acoustics and the Performance of Music* (Verlag das Musikinstrument, Frankfurt/Main, Germany, 1978).

[59] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments* (Springer, New York, 1991).

[60] J. Flanagan, "Analog Measurements of Sound Radiation from the Mouth," *J. Acoust. Soc. Am.*, vol. 32, pp. 1613–1620 (1960 Dec.).

[61] J. Huopaniemi, K. Kettunen, and J. Rahkonen, "Measurement and Modeling Techniques for Directional Sound Radiation from the Mouth," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'99)* (New Paltz, NY, 1999 Oct.).

[62] D. H. Cooper, "Calculator Program for Head-Related Transfer Function," *J. Audio Eng. Soc. (Personal Calculator Programs)*, vol. 30, pp. 34–38 (1982 Jan./Feb.).

[63] W. M. Rabinowitz, J. Maxwell, Y. Shao, and M. Wei, "Sound Localization Cues for a Magnified Head: Implications from Sound Diffraction about a Rigid Sphere," *Presence*, vol. 2, no. 2, pp. 125–129 (1993).

[64] R. Duda and W. Martens, "Range-Dependence of the HRTF of a Spherical Head," *J. Acoust. Soc. Am.*, vol. 104, pp. 3048–3058 (1998 Nov.).

[65] M. R. Schroeder, "Digital Simulation of Sound Transmission in Reverberant Spaces," *J. Acoust. Soc. Am.*, vol. 47, no. 2, pt. 1, pp. 424–431 (1970).

[66] M. R. Schroeder, "Computer Models for Concert Hall Acoustics," *Am. J. Phys.*, vol. 41, pp. 461–471 (1973).

[67] A. Pietrzyk, "Computer Modeling of the Sound Field in Small Rooms," in *Proc. AES 15th Int. Conf. on Audio Acoustics and Small Spaces* (Copenhagen, Denmark, 1998 Oct. 31–Nov. 2), pp. 24–31.

[68] H. Kuttruff, *Room Acoustics*, 3rd ed. (Elsevier, Essex, UK, 1991).

[69] R. Lyon and R. DeJong, *Theory and Application of Statistical Energy Analysis*, 2nd ed. (Butterworth-Heinemann, Newton, MA, 1995).

[70] D. Jaffe and J. O. Smith, "Extensions of the Karplus–Strong Plucked String Algorithm," *Comput. Music J.*, vol. 7, no. 2, pp. 56–69 (1983 Summer). reprinted in *The Music Machine*, C. Roads, Ed. (MIT Press, Cambridge, MA, 1989), pp. 481–494.

[71] J. O. Smith, "Physical Modeling Using Digital Waveguides," *Comput. Music J.*, vol. 16, no. 4, pp. 74–87 (1992 Winter).

[72] V. Välimäki and T. Takala, "Virtual Musical Instruments—Natural Sound Using Physical Models," *Organised Sound*, vol. 1, no. 2, pp. 75–86 (1996).

[73] J. O. Smith, "Principles of Digital Waveguide Models of Musical Instruments," in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds. (Kluwer Academic, Boston, MA, 1997), chap. 10, pp. 417–466.

[74] S. Van Duyne and J. O. Smith, "Physical Modeling with the 2-D Digital Waveguide Mesh," in *Proc. Int. Computer Music Conf. (ICMC'93)* (Tokyo, Japan, 1993 Sept.), pp. 40–47.

[75] L. Savioja, M. Karjalainen, and T. Takala, "DSP Formulation of a Finite Difference Method for Room Acoustics Simulation," in *Proc. IEEE Nordic Signal Processing Symp. (NORSIG'96)* (Espoo, Finland, 1996 Sept.), pp. 455–458.

[76] S. Van Duyne and J. O. Smith, "The Tetrahedral Digital Waveguide Mesh," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'95)* (New Paltz, NY, 1995 Oct.).

[77] F. Fontana and D. Rocchesso, "Physical Modeling of Membranes for Percussion Instruments," *Acustica* united with *Acta Acustica*, vol. 84, pp. 529–542 (1998 May/June).

[78] L. Savioja and V. Välimäki, "Improved Discrete-Time Modeling of Multi-Dimensional Wave Propagation Using the Interpolated Digital Waveguide Mesh," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'97)*, vol. 1 (Munich, Germany, 1997 Apr. 19–24), pp. 459–462.

[79] L. Savioja and V. Välimäki, "Reduction of the Dispersion Error in the Triangular Digital Waveguide Mesh Using Frequency Warping," *IEEE Signal Process. Lett.*, vol. 6, no. 3, pp. 58–60 (1999 Mar.).

[80] L. Savioja and V. Välimäki, "Reduction of the Dispersion Error in the Interpolated Digital Waveguide Mesh Using Frequency Warping," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'99)*, vol. 2 (Phoenix, AZ, 1999 Mar. 15–19), pp. 973–976.

[81] A. Kulowski, "Error Investigation for the Ray Tracing Technique," *Appl. Acoust.*, vol. 15, pp. 263–274 (1982).

[82] D. van Maercke and J. Martin, "The Prediction of Echograms and Impulse Responses within the Epidaure Software," *Appl. Acoust.*, vol. 38, no. 2–4 (Special Issue on Computer Modelling and Auralisation of Sound Fields in Rooms), pp. 93–114 (1993).

[83] G. M. Naylor, "ODEON—Another Hybrid Room Acoustical Model," *Appl. Acoust.*, vol. 38, no. 2–4 (Special Issue on Computer Modelling and Auralisation of Sound Fields in Rooms), pp. 131–143 (1993).

[84] B. M. Gibbs and D. K. Jones, "A Simple Image Method for Calculating the Distribution of Sound Pressure Levels within an Enclosure," *Acustica*, vol. 26, no. 1, pp. 24–32 (1972).

[85] H. Lee and B. H. Lee, "An Efficient Algorithm for the Image Model Technique," *Appl. Acoust.*, vol. 24, pp. 87–115 (1988).

[86] R. Heinz, "Binaural Room Simulation Based on an Image Source Model with Addition of Statistical Methods to Include the Diffuse Sound Scattering of Walls and to Predict the Reverberant Tail," *Appl. Acoust.*, vol. 38, no. 2–4 (Special Issue on Computer Modelling and Auralisation of Sound Fields in Rooms), pp. 145–159 (1993).

[87] D. van Maercke, "Simulation of Sound Fields in Time and Frequency Domain Using a Geometrical Model," in *Proc. 12th Int. Congr. on Acoustics (ICA'86)*, vol. 2 (Toronto, Ont., Canada, 1986 July), paper E11-7.

[88] M. Vorländer, "Simulation of the Transient and Steady-State Sound Propagation in Rooms Using a New Combined Ray-Tracing/Image-Source Algorithm," *J. Acoust. Soc. Am.*, vol. 86, pp. 172–178 (1989).

[89] F. R. Moore, "A General Model for Spatial Processing of Sounds," *Comput. Music J.*, vol. 7, no. 3, pp. 6–15 (1983 Fall).

[90] M. Tamminen, "The EXCELL Method for Efficient Geometric Access to Data," *Acta Polytechnica Scandinavica, Math. and Comput. Sci. Ser.*, no. 34

(1981).

[91] H. Samet, *The Design and Analysis of Spatial Data Structures* (Addison-Wesley, Reading, MA, 1990).

[92] H. Bass and H. J. Bauer, "Atmospheric Absorption of Sound: Analytical Expressions," *J. Acoust. Soc. Am.*, vol. 52, pp. 821–825 (1972).

[93] ISO 9613-1, "Acoustics—Attenuation of Sound during Propagation Outdoors—Part 1: Calculation of the Absorption of Sound by the Atmosphere," International Standards Organization, Geneva, Switzerland (1993).

[94] J. Huopaniemi, L. Savioja, and M. Karjalainen, "Modeling of Reflections and Air Absorption in Acoustical Spaces—A Digital Filter Design Approach," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'97)* (New Paltz, NY, 1997 Oct. 19–22).

[95] G. Naylor and J. Rindel, *Odeon Room Acoustics Program, Version 2.5, User Manual*, Technical University of Denmark, Acoustics Laboratory, Publ. 49 (1994).

[96] T. Lahti and H. Møller, "The Sigyn Hall, Turku—A Concert Hall of Glass," in *Proc. Nordic Acoustical Meeting (NAM'96)* (Helsinki, Finland, 1996 June), pp. 43–48.

[97] J.-M. Jot, "Etude et réalisation d'un spatialisateur de sons par modèles physique et perceptifs," Ph.D. thesis, Ecole Nationale Supérieure des Télécommunications, Télécom Paris 92 E 019 (1992 Sept.).

[98] M. R. Schroeder, "An Artificial Sterephonic Effect Obtained from a Single Audio Signal," *J. Audio Eng. Soc.*, vol. 6, pp. 74–79 (1985 Apr.).

[99] J. Stautner and M. Puckette, "Designing Multi-Channel Reverberators," *Comput. Music J.*, vol. 6, pp. 569–579 (1982).

[100] R. Vermeulen, "Stereo-Reverberation," *J. Audio Eng. Soc.*, vol. 6, pp. 124–130 (1958 Apr.).

[101] W. Gardner, "Virtual Acoustic Room," Master's thesis, MIT, Cambridge, MA (1992).

[102] D. Rocchesso and J. O. Smith, "Circulant and Elliptic Feedback Delay Networks for Artificial Reverberation," *IEEE Trans. Speech Audio Process.*, vol. 5, pp. 51–63 (1997 Jan.).

[103] R. Väänänen, V. Välimäki, and J. Huopaniemi, "Efficient and Parametric Reverberator for Room Acoustics Modeling," in *Proc. Int. Computer Music Conf. (ICMC'97)* (Thessaloniki, Greece, 1997 Sept.), pp. 200–203.

[104] K. A. Riederer, "Repeatability Analysis of Head-Related Transfer Function Measurements," presented at the 105th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 46, p. 1036 (1998 Nov.), preprint 4846.

[105] W. Gardner and K. Martin, "HRTF Measurements of a KEMAR," *J. Acoust. Soc. Am.*, vol. 97, pp. 3907–3908 (1995).

[106] W. Gardner, "3-D Audio Using Loudspeakers," Ph.D. thesis, MIT Media Lab., Cambridge, MA (1997 Sept.). Revised version published by Kluwer Academic, Boston, MA (1998).

[107] W. Martens, "Principal Components Analysis and Resynthesis of Spectral Cues to Perceived Direction," in *Proc. Int. Computer Music Conf. (ICMC'87)* (1987), pp. 274–281).

[108] J. Abel and S. Foster, "Method and Apparatus for Efficient Presentation of High-Quality Three-Dimensional Audio including Ambient Effects," US patent 5,802,180 (1998 Sept.).

[109] S. Mehrgardt and V. Mellert, "Transformation Characteristics of the External Human Ear," *J. Acoust. Soc. Am.*, vol. 61, pp. 1567–1576 (1977).

[110] A. Kulkarni, S. K. Isabelle, and H. S. Colburn, "On the Minimum-Phase Approximation of Head-Related Transfer Functions," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'95)* (New Paltz, NY, 1995 Oct.).

[111] J. Köring and A. Schmitz, "Simplifying Cancellation of Cross-Talk for Playback of Head-Related Recordings in a Two-Speaker System," *Acustica* united with *Acta Acustica*, vol. 179, pp. 221–232 (19930.

[112] A. Kulkarni and H. S. Colburn, "Efficient Finite-Impulse-Response Filter Models of the Head-Related Transfer Function," *J. Acoust. Soc. Am.*, vol. 97, p. 3278 (1995).

[113] A. Kulkarni and H. S. Colburn, "Infinite-Impulse-Response Filter Models of the Head-Related Transfer Function," *J. Acoust. Soc. Am.*, vol. 97, p. 3278 (1995).

[114] J. Huopaniemi and M. Karjalainen, "HRTF Filter Design Based on Auditory Criteria," in *Proc. Nordic Acoustical Meeting (NAM'96)* (Helsinki, Finland, 1996).

[115] K. Hartung and A. Raab, "Efficient Modeling of Head-Related Transfer Functions," *Acta Informatica*, vol. 82 (suppl. 1), S88 (1996).

[116] J. Mackenzie, J. Huopaniemi, V. Välimäki, and I. Kale, "Low-Order Modelling of Head-Related Transfer Functions Using Balanced Model Truncation," *IEEE Signal Process. Lett.*, vol. 4, no. 2, pp. 39–41 (1997 Feb.).

[117] J. Huopaniemi and M. Karjalainen, "Review of Digital Filter Design and Implementation Methods for 3-D Sound," presented at the 102nd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 45, p. 413 (1997 May), preprint 4461.

[118] S. Wu and W. Putnam, "Minimum Perceptual Spectral Distance FIR Filter Design," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'97)*, vol. 1 (Los Alamitos, CA, 1997), pp. 447–450. (Institute of Electrical and Electronics Engineers, IEEE Computer Society Press).

[119] J. Huopaniemi and J. O. Smith, "Spectral and Time-Domain Preprocessing and the Choice of Modeling Error Criteria for Binaural Digital Filters," in *Proc. AES 16th Int. Conf. on Spatial Sound Reproduction* (Rovaniemi, Finland, 1999 Apr.), pp. 301–312.

[120] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models* (Springer, Heidelberg, Germany, 1990).

[121] B. Moore and B. Glasberg, "Suggested Formulae for Calculating Auditory-Filter Bandwidths and Excitation Patterns," *J. Acoust. Soc. Am.*, vol. 74, pp. 750–753 (1983 Sept.).

[122] J. O. Smith, "Techniques for Digital Filter Design and System Identification with Application to the Violin," Ph.D. thesis, Stanford University, Stanford, CA (1983 June).

[123] V. Larcher and J. M. Jot, "Techniques d'interpolation de filtres audionumériques: Application à la reproduction spatiale des sons sur écouteurs," in *Proc. CFA: Congrès Français d'Acoustique* (1997 Apr.).

[124] M. A. Gerzon, "Panpot Laws for Multispeaker Stereo," presented at the 92nd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 40, p. 447 (1992 May), preprint 3309.

[125] V. Pulkki and T. Lokki, "Creating Auditory Displays with Multiple Loudspeakers Using VBAP: A Case Study with DIVA Project," in *Proc. Int. Conf. on Auditory Display (ICAD'98)* (Glasgow, UK, 1998 Nov. 1–4).

[126] W. G. Gardner, "Efficient Convolution without Input–Output Delay," *J. Audio Eng. Soc.*, vol. 43, pp. 127–136 (1994).

[127] J. Sandvad, "Dynamic Aspects of Auditory Virtual Environments," presented at the 100th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 44, p. 644 (1996 July/Aug.), preprint 4226.

[128] T. I. Laakso, V. Välimäki, M. Karjalainen, and U. K. Laine, "Splitting the Unit Delay—Tools for Fractional Delay Filter Design," *IEEE Signal Process. Mag.*, vol. 13, no. 1, pp. 30–60 (1996 Jan.).

[129] V. Välimäki, M. Karjalainen, Z. Janosy, and U. K. Laine, "A Real-Time DSP Implementation of a Flute Model," in *Proc. 1992 IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 2 (San Francisco, CA, 1992 Mar.), pp. 249–252.

[130] V. Välimäki, J. Huopaniemi, M. Karjalainen, and Z. Janosy, "Physical Modeling of Plucked String Instruments with Application to Real-Time Sound Synthesis," *J. Audio Eng. Soc.*, vol. 44, pp. 331–353 (1996 May).

[131] E. Wenzel, "Analysis of the Role of Update Rate and System Latency in Interactive Virtual Acoustic Environments," presented at the 103rd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 45, pp. 1017, 1018 (1997 Nov.), preprint 4633.

[132] E. Wenzel, "Effect of Increasing System Latency on Localization of Virtual Sounds," in *Proc. AES 16th Int. Conf. on Spatial Sound Reproduction* (Rovaniemi, Finland, 1999 Apr.), pp. 42–50.

[133] Deutsche Telekom, "Investigations on Tolerable Asynchronism between Audio and Video," Doc. 11A/DTAG1, Question ITU-R 35-2/11 (1995 Apr.).

[134] M. P. Hollier and A. N. Rimell, "An Experimental Investigation into Multi-Modal Synchronisation Sensitivity for Perceptual Model Development," presented at the 105th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 46, p. 1033 (1998 Nov.), preprint 4790.

[135] H. Möller and T. Lahti, "Acoustical Design of the Marienkirche Concert Hall, Neubrandenburg" (Abstract), *J. Acoust. Soc. Am.*, vol. 105, pp. 928–929 (1999 Feb.).

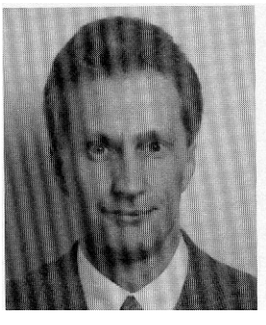
[136] T. Takala, E. Rousku, T. Lokki, L. Savioja, J. Huopaniemi, R. Väänänen, V. Pulkki, and P. Salminen, "Marienkirche—A Visual and Aural Demonstration Film," in *Electronic Art and Animation Catalogue (SIGGRAPH'98)* (Orlando, FL, 1998 July 19–24), p. 149. Presented at SIGGRAPH'98 Computer Animation Festival (Electronic Theater).

[137] L. Beranek, *Concert and Opera Halls—How They Sound* (Acoustical Society of America, New York, 1996).

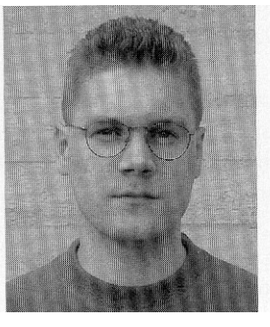
[138] E. M. Wenzel, "What Perception Implies about Implementation of Interactive Virtual Acoustic Environments," presented at the 101st Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 44, p. 1165 (1996 Dec.), preprint 4353.

[139] C. Cruz-Neira, D. Sandin, T. DeFanti, R. Kenyon, and J. Hart, "The Cave—Audio Visual Experience Automatic Virtual Environment," *Commun. ACM*, vol. 35, no. 6, pp. 64–72 (1992 June).

THE AUTHORS



L. Savioja



T. Lokki

R. Väänänen

Lauri Savioja was born in Turku, Finland, in 1966. He studied computer science and acoustics and received the degrees of M.Sc. in technology (1991) and Licentiate of Science in technology (1995) from the Department of Computer Science, Helsinki University of Technology (HUT), Espoo, Finland.

Currently he works for the Laboratory of Telecommunications Software and Multimedia and the Laboratory

of Acoustics and Audio Signal Processing in the university. He is pursuing a Ph.D. degree in technology. His research interests include room acoustics, physical modeling of musical instruments, and virtual reality.

Mr. Savioja is a member of the AES and the Acoustical Society of Finland. His home page on the World Wide Web is http://www.tcm.hut.fi/~las/. Email: Lauri.Savioja@hut.fi.

•

Tapio Lokki was born in Helsinki, Finland, in 1971. He studied acoustics and audio signal processing at Helsinki University of Technology (HUT), and received an M.Sc. degree in electrical engineering in 1997. Since 1998 he has been a Ph.D. candidate at the Telecommunications Software and Multimedia Laboratory at HUT.

Mr. Lokki's research activities include 3-dimensional sound, virtual acoustic environments, auralization, and virtual reality. His main interests are in real-time applications. His nonprofessional activities include cookery, sports, and the "newer French horn music" in Retuperän WBK, thought to be the world's best student orchestra at HUT.

Mr. Lokki is the secretary of the Acoustical Society of Finland. His home page on the World Wide Web is http://www.tcm.hut.fi/~ktlokki/. Email: Tapio.Lokki@hut.fi.

•

Riitta Väänänen was born in Helsinki, Finland, in 1970. She received an M.Sc. degree in electrical engineering from the Helsinki University of Technology (HUT) in 1997. She is currently a Ph.D. student at the HUT majoring in acoustics and audio signal processing. Ms. Väänänen has worked as a research assistant and a research scientist at the Laboratory of Acoustics and Audio Signal Processing in HUT since 1996. Her research activities include room reverberation modeling and modeling of sound sources and acoustic environments in interactive virtual reality systems.

The biography for Jyri Huopaniemi was published in the January/Febuary issue of the *Journal*. Email: jyri.huopaniemi@research.nokia.com.