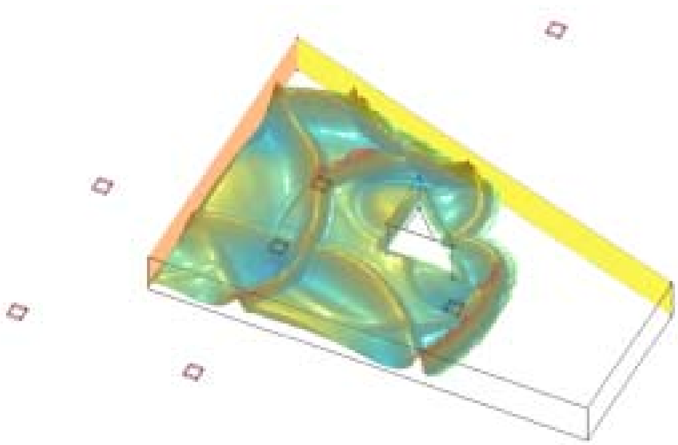
Helsinki University of Technology
Publications in Telecommunications Software and Multimedia
Teknillisen korkeakoulun tietoliikenneohjelmistojen ja multimedian julkaisuja
Espoo 2002 TML-A5

Physically-based Auralization

– Design, Implementation, and Evaluation

Tapio Lokki



TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY
TECHNISCHE UNIVERSITÄT HELSINKI
UNIVERSITE DE TECHNOLOGIE D'HELSINKI

Helsinki University of Technology
Publications in Telecommunications Software and Multimedia
Teknillisen korkeakoulun tietoliikenneohjelmistojen ja multimedian julkaisuja
Espoo 2002 TML-A5

Physically-based Auralization

– Design, Implementation, and Evaluation

Tapio Lokki

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Department of Computer Science and Engineering, for public examination and debate in Auditorium T1 at Helsinki University of Technology (Espoo, Finland) on the 8th of November, 2002, at 12 noon.

Helsinki University of Technology
Department of Computer Science and Engineering
Telecommunications Software and Multimedia Laboratory

Teknillinen korkeakoulu
Tietotekniikan osasto
Tietoliikenneohjelmistojen ja multimedian laboratorio

Distribution:
Helsinki University of Technology
Telecommunications Software and Multimedia Laboratory
P.O.Box 5400
FIN-02015 HUT
Tel. +358-9-451 2870
Fax. +358-9-451 5014

ISBN 951-22-6157-X
ISSN 1456-7911

Otamedia Oy
Espoo 2002

ABSTRACT

Author Tapio Lokki
Title Physically-based Auralization
– Design, Implementation, and Evaluation

The aim of this research is to implement an auralization system that renders audible a 3D model of an acoustic environment. The design of such a system is an iterative process where successive evaluation of auralization quality is utilized to further refine the model and develop the rendering methods. The work can be divided into two parts corresponding to design and implementation of an auralization system and evaluation of the system employing objective and subjective criteria.

The presented auralization method enables both static and dynamic rendering. In dynamic rendering positions and orientations of sound sources, surfaces, or a listener can change. These changes are allowed by modeling the direct sound and early reflections with the image-source method. In addition, the late reverberation is modeled with a time-invariant recursive digital filter structure. The core of the thesis deals with the processing of image sources for auralization. The sound signal emitted by each image source is processed with digital filters modeling such acoustic phenomena as sound source directivity, distance delay and attenuation, air and material absorption, and the characteristics of spatial hearing. The digital filter design and implementation of these filters are presented in detail. The traditional image-source method has also been extended to handle diffraction in addition to specular reflections.

The evaluation of quality of the implemented auralization system was performed by comparing recorded and auralized soundtracks subjectively. The compared soundtracks were prepared by recording sound signals in a real room and by auralizing these signals with a 3D model of the room. The auralization quality was assessed with objective and subjective methods. The objective analysis was based on both traditional room acoustic criteria and on a simplified auditory model developed for this purpose. This new analysis method mimics the behavior of human cochlea. Therefore, with the developed method, impulse responses and sound signals can be visualized with similar time and frequency resolution as human hearing applies. The evaluation was completed subjectively by conducting listening tests. The utilized listening test methodology is explained and the final results are presented. The results show that the implemented auralization system provides plausible and natural sounding auralizations in rooms similar to the lecture room employed for evaluation.

UDC 534.84, 004.383.3
Keywords auralization, room acoustic modeling, digital signal processing, virtual reality, 3D sound, spatial sound evaluation, binaural technology

PREFACE

This work is a result of research carried out at the Telecommunications Software and Multimedia Laboratory, Helsinki University of Technology, Espoo, during the years 1998-2002.

My warmest thanks go to Prof. Lauri Savioja, my thesis supervisor, for his encouragement and guidance during these years. Without the collaboration and daily discussions with Lauri, this thesis has never been completed. In addition, I am grateful to Prof. Tapio "Tassu" Takala and Prof. Matti Karjalainen for their support and enthusiasm for teaching me different ways to make research.

The work reported in this thesis was not done alone. I am very grateful to all my co-authors Mr. Jarmo Hiipakka, Dr. Jyri Huopaniemi, Ms. Hanna Järveläinen, Prof. Matti Karjalainen, Dr. Ville Pulkki, Prof. Lauri Savioja, Prof. Peter Svensson, and Ms. Riitta Väänänen. Their comments and help with diverse problems have without doubt improved the quality of this thesis. My thanks go also to all listening test attendees for their patience and valuable input to this work.

Furthermore, I would like to thank the pre-examiners of my thesis, Dr. Bengt-Inge Dalenbäck and Dr. Jean-Marc Jot for positive feedback and contructive comments. In addition, I wish to express my gratitude to Mr. Nick Zacharov for his help in improving the English of this thesis.

Special thanks go to the personnel of Telecommunications Software and Multimedia Laboratory as well as to the personnel of Acoustics Laboratory for providing an exciting, fun, and relaxed research environment.

Finally, I am deeply intebted to my wife Jutta Jaakkola for her love, support, and patience during these years.

Last but not least, I am grateful to the organizations that have found this work important to support it financially. For this, I thank the Helsinki Graduate School in Computer Science and Engineering (HeCSE), TEKES (The National Technology Agency), Tekniikan edistämissäätiö, and the Nokia Foundation.

Otaniemi, Espoo, 27th September 2002

Tapio Lokki

TABLE OF CONTENTS

LIST OF PUBLICATIONS

This thesis summarizes the following articles and publications, referred to as [P1]-[P7]:

[P1] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen. Creating interactive virtual acoustic environments. *Journal of the Audio Engineering Society*, vol. 47, no. 9, pages 675–705, September 1999.

[P2] T. Lokki, J. Hiipakka, and L. Savioja. A framework for evaluating virtual acoustic environments. *Presented at the 110*$^{\text{th}}$ *Convention of the Audio Engineering Society*, Amsterdam, The Netherlands, May 12–15 2001, preprint no. 5317.

[P3] T. Lokki, U.P. Svensson, and L. Savioja. An efficient auralization of edge diffraction. In *Proceedings of the Audio Engineering Society 21*$^{\text{st}}$ *International Conference: Architectural Acoustics & Sound Reinforcement*, pages 166–172, St. Petersburg, Russia, June 1–3 2002.

[P4] T. Lokki and M. Karjalainen. Analysis of room responses, motivated by auditory perception. *Journal of New Music Research*, vol. 31, no. 2, pages 163–169, 2002.

[P5] T. Lokki. Objective comparison of measured and modeled binaural room responses. In *Proceedings of the 8*$^{\text{th}}$ *International Congress on Sound and Vibration*, pages 2481–2488, Hong Kong, China, July 2–6 2001.

[P6] T. Lokki and H. Järveläinen. Subjective evaluation of auralization of physics-based room acoustic modeling. In *Proceedings of the 7*$^{\text{th}}$ *International Conference on Auditory Display*, pages 26–31, Espoo, Finland, July 29 – August 1 2001.

[P7] T. Lokki and V. Pulkki. Evaluation of geometry-based parametric auralization. In *Proceedings of the Audio Engineering Society 22*$^{\text{nd}}$ *International Conference: Virtual, Synthetic and Entertainment Audio*, pages 367–376, Espoo, Finland, June 15–17 2002.

LIST OF SYMBOLS

a	gain coefficient
$A(z)$	comb-allpass filter
$AI_K(z)$	air absorption filter for image source
$AI_M(z)$	air absorption filter for edge source
c	speed of sound
DL	delay line
$D_K(z)$	sound source directivity filter for image source
$D_M(z)$	sound source directivity filter for edge source
$DF_M(z)$	diffraction filter
f_s	sampling frequency
$F_k(z)$	listener model filter block
g	distance attenuation gain coefficient
$h_\mathrm{diffr}(t)$	impulse response of an edge
$H_\mathrm{diffr}(z)$	edge diffraction filter
$H_{\mathrm{L},K}(z)$	diffuse field equalized HRTF filter for image source (left ear)
$H_{\mathrm{R},K}(z)$	diffuse field equalized HRTF filter for image source (right ear)
$H_{\mathrm{L},M}(z)$	diffuse field equalized HRTF filter for edge source (left ear)
$H_{\mathrm{R},M}(z)$	diffuse field equalized HRTF filter for edge source (right ear)
$H_{\mathrm{dir},i}(\omega,\theta,\phi)$	measured frequency response of sound source
i	integer variable
k	integer variable
K	integer variable
l	edge point-to-receiver distance
$L(z)$	lowpass filter
m	source-to-edge point distance
n	integer variable
M	integer variable
$M_K(z)$	reflection filter for image source
$M_M(z)$	reflection filter for edge source
$\vec{n}$	normal vector
r	distance
R	late reverberation algorithm
$R(X,Y,Z)$	position of receiver
$S(X,Y,Z)$	position of source
$S_\mathrm{d}(z)$	diffuse field filter of sound source directivity
$S_\mathrm{d}(\omega,\theta,\phi)$	power spectrum of diffuse field filter of sound source
t	time variable
$T_k(z)$	auralization filter block
z	z-transform variable
Z_apex	apex point of an edge
$Z_0(X,Y,Z)$	start point of an edge
$Z_1(X,Y,Z)$	end point of an edge

α	angle
$\beta_{\pm\pm}$	directivity function
δ	unit impulse
γ	angle
ν	wedge index
ϕ	elevation angle
θ	azimuth angle
θ_{R}	angle
θ_{S}	angle
θ_{w}	wedge angle
ω	angular frequency

LIST OF ABBREVIATIONS

3D	Three-dimensional
ANOVA	Analysis of Variance
API	Application Programming Interface
B&K	Brüel & Kjær
BRIR	Binaural Room Impulse Response
C50	Clarity Index
DIVA	Digital Interactive Virtual Acoustics
DSP	Digital Signal Processing
EDT	Early Decay Time
ERB	Equivalent Rectangular Bandwidth
FIR	Finite Impulse Response
FDN	Feedback Delay Network
GUI	Graphical User Interface
HD	High Definition
HRIR	Head Related Impulse Response
HRTF	Head Related Transfer Function
IIR	Infinite Impulse Response
IRCAM	Institut de Recherche et Coordination Acoustique/Musique
ITD	Interaural Time Difference
ITU-R	International Telecommunication Union, Radiocommunication Assembly
ISO	International Standardization Organization
MPEG	Moving Picture Experts Group
NASA	National Aeronautics and Space Administration
PC	Personal Computer
SPSS	Statistical Package for Social Sciences
T30	Reverberation Time
VBAP	Vector Base Amplitude Panning
VR	Virtual Reality

This thesis concerns the design, implementation, and evaluation of an auralization system. Auralization or "rendering audible" is an analogous term for visualization when creating images from 3D models. The focus is in design methods employed in the Digital Interactive Virtual Acoustic (DIVA) auralization system and in the assessment of its quality.

The research field related to auralization is multidisciplinary, thus the implementation of the system requires understanding and knowledge of room acoustics, digital signal processing, and psychoacoustics. On a general level the research problem is to model and simulate the sound propagation from sound sources to the ear drums of a listener through the modeled space. The research problems related to the design of an auralization system lie to a great extent in the areas of room acoustic modeling, digital filter design, and 3D sound reproduction as illustrated in Figure 1.1. The goal of the design is authentic auralization, i.e., creation of a virtual auditory environment that is indistinguishable from a real auditory environment [17].

The term auralization has been defined by Kleiner et al. [86] as follows: "*Auralization is the process of rendering audible, by physical or mathematical modeling, the sound field of a source in a space, in such a way as to simulate the binaural listening experience at a given position in the modeled space.*" The most straightforward method to realize auralization is to measure binaural room impulse responses (BRIR) and convolve them with an anechoic signal. This method is called the direct room impulse response rendering method [P1] and it enables authentic auralization. However, measurements can not be performed in virtual spaces, e.g., in a computer model of a hypothetical concert hall. In such cases room acoustic modeling is utilized to predict the BRIRs for convolution.

The auralization system presented in this thesis applies the parametric room impulse response rendering method [P1]. It enables a more robust way for dynamic rendering in which the position of sound sources, surfaces or a listener can change during the rendering process. In this technique the BRIRs are not calculated before the actual auralization process. Instead, a set of either perceptually- or physically-based parameters for the auralization process is defined. The presented

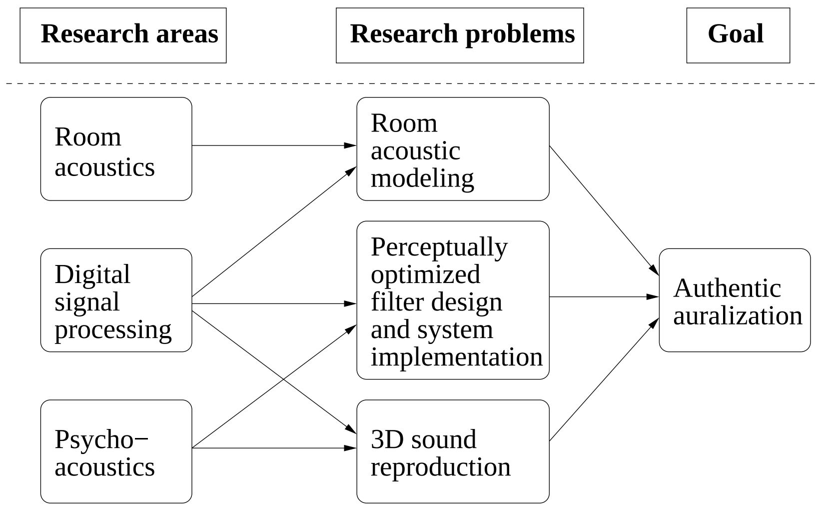


Figure 1.1: Research areas and problems involved in the design of an auralization system.

auralization system is based on the room acoustic modeling and it renders audible the modeled 3D geometry of a space. This physically-based approach is also called geometry-based auralization.

The modeling methods presented in this thesis are mainly derived from physics. Conceptually, this approach is straightforward since by modeling physical principles of sound propagation in air and reflections from boundaries as correctly as possible a high-quality auralization should be achieved. However, the final judgment of the authenticity of the auralization is made by human auditory perception. Therefore, another approach to realize auralization is to design the rendering from the perception point of view. This perceptual approach can be performed without any knowledge of the room geometry and perceptually optimized efficient algorithms are often used [73]. In many applications in computer music and professional audio, perceptually-based auralization produces relevant and accurate enough results.

Room acoustic modeling, required for physically-based auralization, can be realized with different methods. In principle, the most accurate results can be achieved with the wave-based methods that numerically solve the wave equation in a modeled space [P1]. The more practical modeling methods, especially from auralization point of view, are ray-based methods, such as the ray-tracing [88] and the image-source method [7, 21]. In this thesis the room acoustic modeling is realized with the image-source method which can be seen conceptually as a decomposition of the sound field into elementary waves [164]. This concept can be explained as follows.

In free space a sound source emits a spherical wavefront, i.e., an elementary wave. It is a wavefront propagating homogeneously in all directions emitted by a point source. The amplitude of emitted sound is inversely proportional to the distance from the sound source. The reflections can be modeled as new sound sources since each reflection creates a new elementary wave. Therefore it is possible to calculate reflections in a recursive manner so that for each reflection a new sound source is created. Finally, the geometry of a modeled space is represented only with a sound source and secondary sources, representing reflections, and the final sound field in the listening position is a superposition of elementary waves emitted by these sources. However, some of these secondary sources are not visible to a listening point due to occlusion by surfaces. For this reason validity of all sources is verified with a visibility check.

The traditional image-source method neglects such phenomena as diffraction and diffusion. However, they can be modeled by representation as secondary sources that emit elementary waves. In conclusion, the acoustic space can be represented by different types of secondary sources, such as image sources, edge sources, and surface sources [164]. The auralization system implemented in this thesis consists of image sources and edge sources. The surface sources, which correspond to diffuse reflections are not implemented, although they are an important phenomenon in room acoustics [29].

Figure 1.2 illustrates the concept of sound field decomposition. Each reflection from a wall is replaced with an image source and each corner (except convex rectangular corners) is replaced with an edge source. All these secondary sources emit a wavefront shown inside the geometry. With the concept of secondary sources each elementary wave can be easily filtered with frequency dependent acoustic phenomena such as sound source directivity, distance delay and attenuation, air, material, and wall absorption which are all included to the simulation in Figure 1.2. Auralization of elementary waves includes suitable processing for applied 3D sound reproduction method. In the case of multi-channel loudspeaker reproduction elementary waves are positioned to correct directions by applying

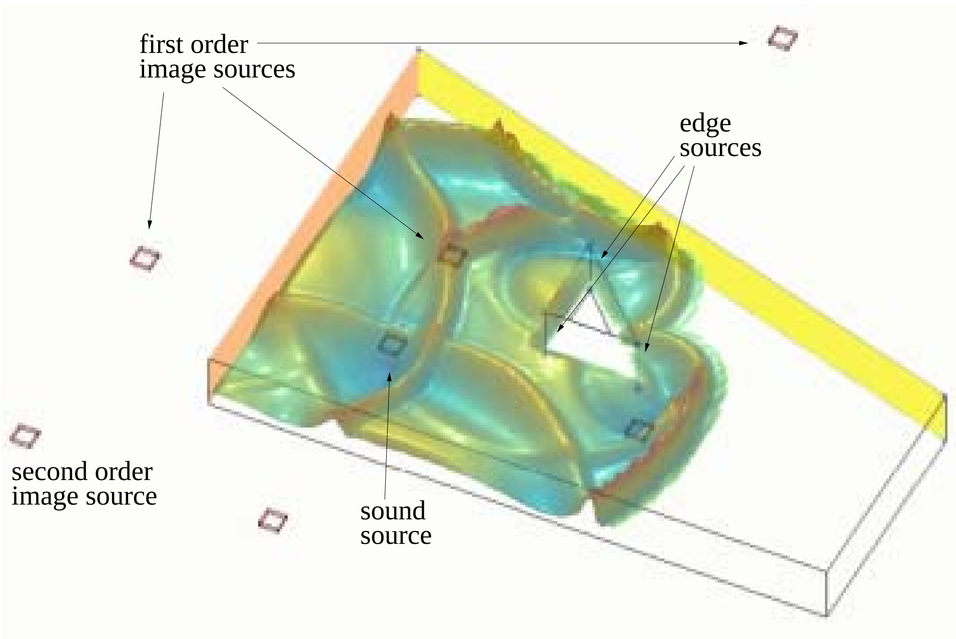


Figure 1.2: An example of sound field decomposition into elementary waves with the image-source method. The illustration is created by computing an impulse response in each pixel and by plotting the time moment of the 680^{th} sample which corresponds to 14.2 ms in time.

such methods as vector base amplitude panning (VBAP) [133, 135] or Ambisonic processing [50, 102]. When reproduction is handled with two channels, loudspeakers or headphones, the modeling of human spatial hearing is needed so that elementary waves emanate from the correct angle.

1.1 Aims of the Thesis

The aim of this thesis is to describe the creation and validation of the DIVA auralization system. The realized system renders a virtual auditory environment based on the information of room geometry, physical material data, and the positions of a sound source and a listener. The auralization is implemented in a manner that enables both static and dynamic rendering either in real-time or in non-real-time applications. Another aim is to evaluate the quality of the auralizations with objective and subjective criteria.

The design of an auralization system is an iterative process where the individual parts are refined and upgraded step by step to obtain as authentic an auralization as possible. The evaluation of the quality is complicated since the optimal quality cannot be defined. The quality of auralization in this case is defined with perceptual criteria, the auralization is good enough if it cannot be distinguished from the corresponding recording in which an anechoic signal is recorded in the real space. The evaluation is performed both subjectively with listening tests and objectively by studying monaural signals with a simple auditory model.

A Case Study

The quality of the realized auralizations is evaluated by comparing the signals recorded in a real room and the signals auralized with the 3D model of the same room. For this case study, the lecture room “T3” of the Computer Science building at the Helsinki University of Technology was chosen, because the room geometry could be modeled quite accurately and the room was easily available for

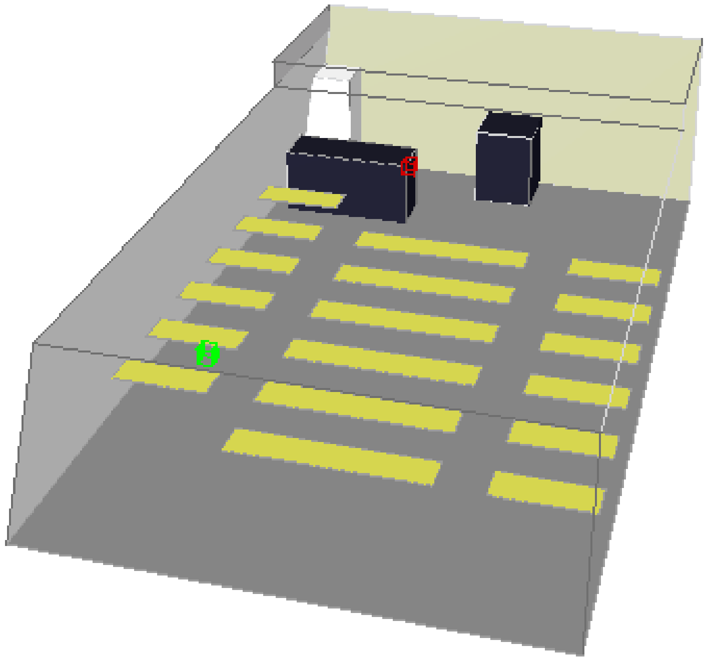


Figure 1.3: The geometry of the studied lecture room.

recordings and measurements. The dimensions of this lecture room are 12.0 m x 7.3 m x 2.6 m. The modeled geometry is depicted in Figure 1.3. The sound source utilized in recordings was a loudspeaker, the directivity of which was measured. The recordings were made with a real-head recording technique so that spatial characteristics of sound were captured and movements of the listener were possible.

1.2 Applications

Auralization systems can be employed, for example, in room acoustic design, in psychoacoustics and sound reproduction research, as well as in multimedia applications. Auralization of 3D models of spaces could be a powerful tool in the design of acoustically demanding spaces, such as lecture rooms, mixing studios, and concert halls. In addition, acoustic consultants can use auralization to convince their customers of the acoustical design of spaces.

In psychoacoustics research, physically-based auralization can be used to create different kinds of acoustical spaces in which some specific feature can be adjusted. In this way, researchers can concentrate, e.g., on the perception of reflections from different directions. Another research area which can benefit from physically-based auralization is multichannel sound reproduction research. By modeling the reproduction room where a loudspeaker setup under study is arranged the important loudspeaker-room interaction can be included to simulations.

Rendering audible 3D models is beneficial in a number of multimedia applications. Maybe the biggest application area for auralization is in animation and the computer game industries where realistic virtual soundscapes enhance the total perceptual quality of the applications. Finally, it should be noted that the signal processing concepts presented in this thesis are not optimal for all applications. However, the presented parametric rendering structure can be tailored to the needs of the target application.

1.3 Organization of the Thesis

This thesis is organized as follows. Chapter 2 gives an overview of the research and development related to auralization systems. In Chapter 3, the implemented auralization system is presented and details of the digital signal processing concepts and filter designs are explained. Chapter 4 discusses the assessment of the auralization quality and presents the results of objective and subjective evaluation. Finally, Chapter 5 concludes the thesis and a summary of publications and author's contribution is provided in Chapter 6.

2 RELATED RESEARCH

Significant research has been carried out in the field of room acoustic modeling, digital filter design, and 3D sound reproduction, which are all needed in the design of an auralization system. Being such a broad field, consisting of several subtopics of acoustics, it is impossible to cover all related research within the scope of this thesis. Despite this fact, this chapter gives a brief overview of existing auralization systems, in relation to the presented work.

2.1 Auralization and Computational Room Acoustics

The predecessors of auralization systems have been artificial reverberation algorithms. Since the pioneering work of Schroeder [154, 155, 156] digital reverberators have been developed for professional audio and music industry. The design goal in reverberators is similar than in auralization systems; to model decaying sound field by producing a dense pattern of reflections. Other fundamental work in the field of reverberator design has been completed, e.g., by Moore [119], Moorer [120], and Stautner and Puckette [161].

Reverberation algorithms as well as auralization systems can be designed from a physical or perceptual point of view. The physical approach seeks to simulate exactly the propagation of sound from the source to the listener for a given room while the perceptual approach endeavors to reproduce only the perceptually salient characteristics of reverberation [47]. While most of the realized auralization systems are based on physical approach all of them include some perceptual features, used to simplify and optimize calculation.

Auralization systems have been studied by several research groups. In the following different research groups are mentioned and references to the most important publications are given. First, the research groups whose main achievements are in computational modeling of room acoustics are listed. It should be noted that almost all work has been performed during the last 15 years and many institutes published ideas and algorithms at the same time:

- One of the first attempts to model room acoustics and realize binaural auralizations of concert halls was realized in the end of 1980's in the Centre Scientifique et Technique du Bâtiment, France, [104, 105, 178, 179, 180, 181]. In addition, they were one of the first groups proposing separate modeling of early reflections and statistical late reverberation.

- Room acoustic modeling research has been carried out in the Technical University of Aachen a decade ago, Germany, [52, 89, 90, 91, 182]. Recently, they have applied auralization in studies of sound insulation in buildings [184, 185, 186].

- Much research on modeling of reflections as well as basic research on auralization methods has been conducted in the Chalmers University of Technology, Sweden, since the beginning of 1990's [24, 25, 26, 27, 28, 29, 30, 31, 32, 86, 87]. Recently, they have studied edge diffraction modeling successfully [163, 170, 171]. A room acoustic modeling program *CATT Acoustic*[1], including auralization, has been developed by Dr. Dalenbäck in close cooperation with people from the Chalmers University of Technology.

[1] http://www.catt.se/

- In the Technical University of Denmark, research on modeling of concert hall acoustics and auralization [99, 121, 122, 139, 140, 141, 142, 143] during the last decade has led to a room acoustic modeling program called *Odeon*[2].

- In the University of Parma, Italy, research on concert hall acoustics and sound systems inside cars and automotive acoustics [36, 37, 38, 39, 40] has been executed since mid 1990's. In addition, they have produced a commercial room acoustics modeling software called *Ramsete*[3].

- In relation to computer graphics, room acoustic modeling methods and real-time sound rendering algorithms have been developed since 1997 in the Princeton University and the Bell Laboratories, USA, [42, 43, 172, 173, 174, 175].

- Research on artificial reverberators and auralization have been carried out since 1992 in the Massachusetts Institute of Technology, USA, [22, 44, 46, 47, 48, 49, 158].

- A room acoustic simulation program *EASE* including auralization has been developed during the last decade by the Acoustic Design Ahnert[4], Germany, [3, 4, 5, 6]. Other room acoustics prediction programs that contain auralization features are *RayNoise*[5] and *Bose Auditioner*[6].

The following research groups have had significant impact on the fields of 3D sound recording and reproduction:

- In the Ruhr-Universität Bochum, Germany, a great deal of basic research on human spatial hearing and room acoustic modeling has been done over 30 years [17, 18, 19, 33, 95, 96].

- Binaural technology and HRTF measurement methods have been developed extensively since 1992 in the Aalborg University, Denmark [109, 110, 111, 112, 113, 114, 115, 116, 145, 146].

- In the Institute of Sound and Vibration Research, United Kingdom, 3D sound reproduction technology and signal processing algorithms have been developed during the last decade [54, 80, 81, 83, 84, 123, 124, 125, 169].

- In the NASA Ames Research Center, USA, the research has been driven by the interest to directional hearing and real-time systems since 1990 [13, 14, 41, 187, 188, 189, 192]. In addition, Dr. Begault published the first book concerning 3D sound in 1994 [12]. Recently, they have been building a real-time auralization tool, SLAB [108, 190, 191], for interactive spatial sound research.

The following research has been performed mainly for some specific application, although the listed publications contain a great deal of basic research and implementation details:

- A spatializer tool for musicians with perceptual modeling approach has been developed since the beginning of 1990's in the Institut de Recherche et Coordination Acoustique/Musique (IRCAM) [23, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79]. In particular, they have developed signal processing algorithms for efficient rendering and for real-time systems.

[2] http://www.dat.dtu.dk/~odeon/
[3] http://www.ramsete.com/
[4] http://www.ada-acousticdesign.de/
[5] http://www.lms.be/
[6] http://www.bose.co.uk/

- During recent years researchers at the Studer Professional Audio AG, Switzerland, have implemented tools for professional audio, such as virtual reference listening room [128] and binaural room scanning system for digital mixing console [55, 56, 100].

- Applications of auralization in virtual reality (VR) installations has been created in Germany in the Fraunhofer Institute for Computer Graphics [8, 9] and German National Research Center for Information Technology [35].

- In the beginning of 1990's before the era of powerful desktop computers the convolution of impulse responses and anechoic signals were realized with such devices as *Convolvotron* [41, 192] by Crystal River Engineering and *HURON* [107, 137] by Lake Technologies.

Furthermore, in the Helsinki University of Technology we have been developing auralization techniques since 1994 by concentrating on room acoustic modeling research [61, 63, 97, 136, 147, 148, 149, 150, 151, 152, 167] and spatial hearing and 3D sound reproduction research [57, 58, 59, 62, 133, 134, 138].

Despite the extensive research performed in the areas of room acoustic modeling, digital filter design, and 3D sound reproduction, research problems still exist to be solved. In this thesis the dynamic rendering and auralization quality are in particular addressed.

2.2 Evaluation of Auralization Quality

Although quite many auralization systems have been developed, the evaluation of their quality has not been reported. This might be a consequence of the facts that such evaluation is laborious and the absolute quality is hard to define. Pellegrini [128] suggests that an authentic auditory virtual environment cannot be reached and the quality should only be estimated with the concept of plausibility. This means that all required quality features should be tested with a given specific application and in this way it can be defined whether the performance of auralization is plausible with the specific application. The auralization system presented in this thesis is not meant for any specific application and its evaluation employing usability studies with a certain task was not appropriate.

In this thesis the evaluation of the quality of auralization has been realized by comparing recordings and auralizations [P2]. This idea has already been proposed by Borish in 1984 [20] and also Kleiner et al. [86] discuss of such a comparison. However, only one such evaluation has been reported so far by Pompetzki and Blauert [17, 131, 132]. Their study, realized ten years ago in the Ruhr-Universität Bochum, was performed in a big lecture hall and to obtain binaural responses they applied artificial head for impulse response measurements. As a sound source they used a dodecahedron loudspeaker array with a fairly uniform radiation characteristics [132]. As a result, they claim that with proper geometrical modeling and with in-situ measured absorption characteristics of walls a reasonably authentic perception was achieved with speech signals.

2.3 Current Trends

While the research of auralization has grown out of concert hall acoustics, the largest application area is currently associated with entertainment. In particular computer games is currently the driving force in the development for efficient rendering algorithms. In modern computer games 3D graphics is always present and more and more often 3D sound is utilized to enhance the quality of the gaming experience. In fact, manufacturers and researchers have defined some guidelines and recommendations that should be included in modern sound applications and application programming interfaces (API) [126, 127]. These guidelines are intended for game developers and sound card producers so that similar auralization methods, as presented in this thesis, can be efficiently used with standard PCs.

In addition to computer games, other multimedia applications might benefit the realistic auralization of the modeled audiovisual scene. For such applications, MPEG-4 standard [1] that specifies coding and object-based presentation of audio and visual content has been defined. The audio part of the MPEG-4 standard contains algorithms and definitions for spatial processing of sound [153, 176].

3 THE DIVA AURALIZATION SYSTEM

The history of the DIVA auralization system dates back to the beginning of 1990's. The first article concerning sound rendering [166] introduced the ideas how to make dynamic sound rendering. Since these days the system has been developed more or less actively. The earlier versions of the DIVA auralization have been part of larger multimedia systems such as the Virtual Orchestra [53, 97] and the Marienkirche demonstration film [168][P1]. The DIVA auralization system has already been the inspiration for two doctoral thesis [57, 147].

In this chapter the current implementation of the DIVA auralization system is briefly described. The methods and algorithms are mostly presented in [57, 98, 147, 177][P1-P3].

3.1 Overview of the System

To enable both static and dynamic rendering, also in real-time, the DIVA auralization system utilizes a parametric room impulse rendering method [P1]. In this technique the impulse response needed for auralization is not calculated, but a parametric representation of it is created for the auralization process. The whole process can be divided into three levels, as proposed by Borish [20] and Blauert [17]:

1. definition of the scene,
2. calculation of sound propagation and reflections in the space,
3. auralization and audio signal processing.

In the following each level is discussed separately. Figure 3.1 clarifies the division and presents how the room acoustic modeling part is divided into two parts.

Definition of the Scene
The acoustic model of a scene contains 3D geometry and information about the material properties. In the DIVA system polygonal modeling is applied and each polygon is associated with a material described by absorption coefficients. They depend on frequency and direction, but in practice they are given independent of direction in octave bands. In addition, the model contains location and orientation of the sound source(s) and the listener as well as the directivity information of the sound source(s). In dynamic rendering, this data can be read from a file or from an input device, such as a mouse or a head-tracker device.

Room Acoustic Simulation
The room acoustic modeling, required for auralization, is divided into two parts. The early sound [15], i.e., direct sound and early reflections are modeled with the image-source method [7, 21] and the late reverberation is modeled with an efficient perceptually motivated algorithm [177]. This division enables efficient dynamic rendering. Each individual component of early sound is time and position variant while the latter part of rendering represents the diffuse reverberant field, which can be treated as a time-invariant filter. Naturally, such time-invariant reverberation is valid only with diffuse sound fields, which is typical, e.g. in concert halls.

As illustrated in Figure 1.2, early sound is modeled by decomposing the sound field into the elementary waves. These waves are represented by image and edge

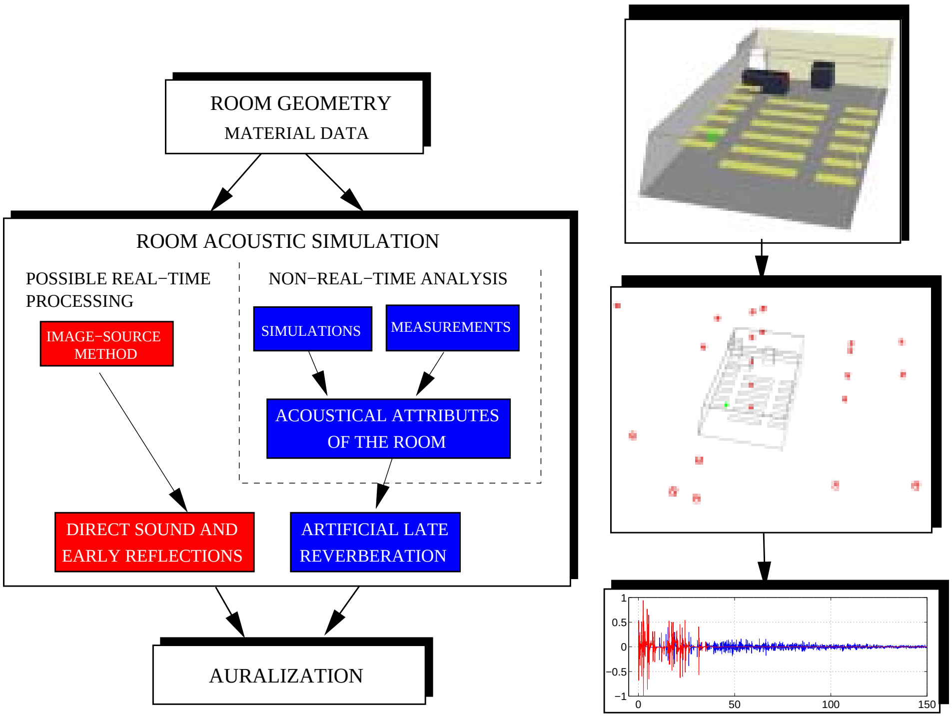


Figure 3.1: Three levels of the modeling process and a hybrid method for room acoustic modeling [P1][98].

sources calculated with the image-source method. The image sources are found by reflecting the original source against all surfaces. The reflection process can be recursively extended to find higher order image sources. In a typical room geometry most of the image sources are invalid since the corresponding reflection paths are not realizable. In addition, some of the image sources are not visible at the receiver point because of occlusion. Thus a visibility check is needed and it is implemented as follows. For each image source the reflection path from the source to the receiver through all the reflecting surfaces is formed and possible intersections of the path with other surfaces are calculated. To reduce the number of required intersection calculations a spatial directory with adaptive spatial subdivision and hash-based addressing is used. A more detailed description of applied algorithms for searching valid image and edge sources is out of the scope of this thesis, but algorithms are explained in [136, 147].

The input data for the calculation of image and edge sources is the room geometry and the material data. Based on the location and orientation of the sound source(s) and the listener the process provides the following auralization parameters for each visible image and edge source:

- order of reflection,
- orientation (azimuth and elevation angles) of sound source,
- distance from the listener,
- set of filter coefficients which describe the material properties in reflections,
- incoming direction of the sound (azimuth and elevation angle in relation to the listener),
- in case of an edge source, parameters needed for edge response calculation (see Section 3.2).

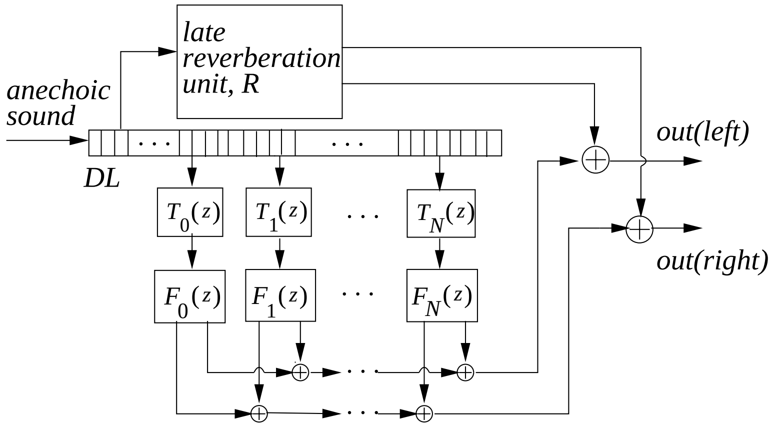


Figure 3.2: The signal processing structure of the DIVA auralization system [149][P1]. In this example the output is binaural.

The parameters of late reverberation are calculated off-line as illustrated in Figure 3.1. This allows tuning of the reverberation time and other features of the reverberation according to the modeled space.

Auralization and Signal Processing

Image-source calculation provides the auralization parameters which are finally converted to signal processing parameters. The reason for this two level process is the fact, that in a dynamic rendering the auralization parameters do not need to be updated for every audio sample. However, the signal processing parameters have to be defined on a sample by sample basis. For efficiency, they are picked from pre-calculated tables or created by interpolating the auralization parameters.

The final auralization process is implemented as a signal processing structure, presented in Figure 3.2. The signal processing structure contains a long delay line DL which is fed with anechoic sound to be processed. The distance of the image source from the listener defines the pick-up point to the filter block $T_k(z)$ where $k = 0, 1, 2 \dots N$ is the identifier of the image source ($k = 0$ corresponds to direct sound). Each block $T_k(z)$ modifies sound signal with the sound source directivity filter, distance dependent gain, air absorption filter and material filter (not for direct sound). The incoming direction of the sound is defined with block $F_k(z)$ containing directional filtering or panning depending on the reproduction method. The superimposed outputs of the each filter $F_k(z)$ are finally summed with the outputs of the late reverberation unit R which is a complex recursive algorithm.

In the following section the signal processing issues as well as the applied filter design methods are presented in more detail.

3.2 Implementation of Digital Signal Processing Structure

The processing of each elementary wave emitted by an image or an edge source is subdivided into several filters. This subdivision is shown in Figure 3.3 for one image and for one edge source. The design of each filter and their use in both static and dynamic renderings are presented in articles [P1-P3]. Here, the applied algorithms and filter design methods are briefly reviewed and expanded upon the optimal performance in the case study of the lecture room.

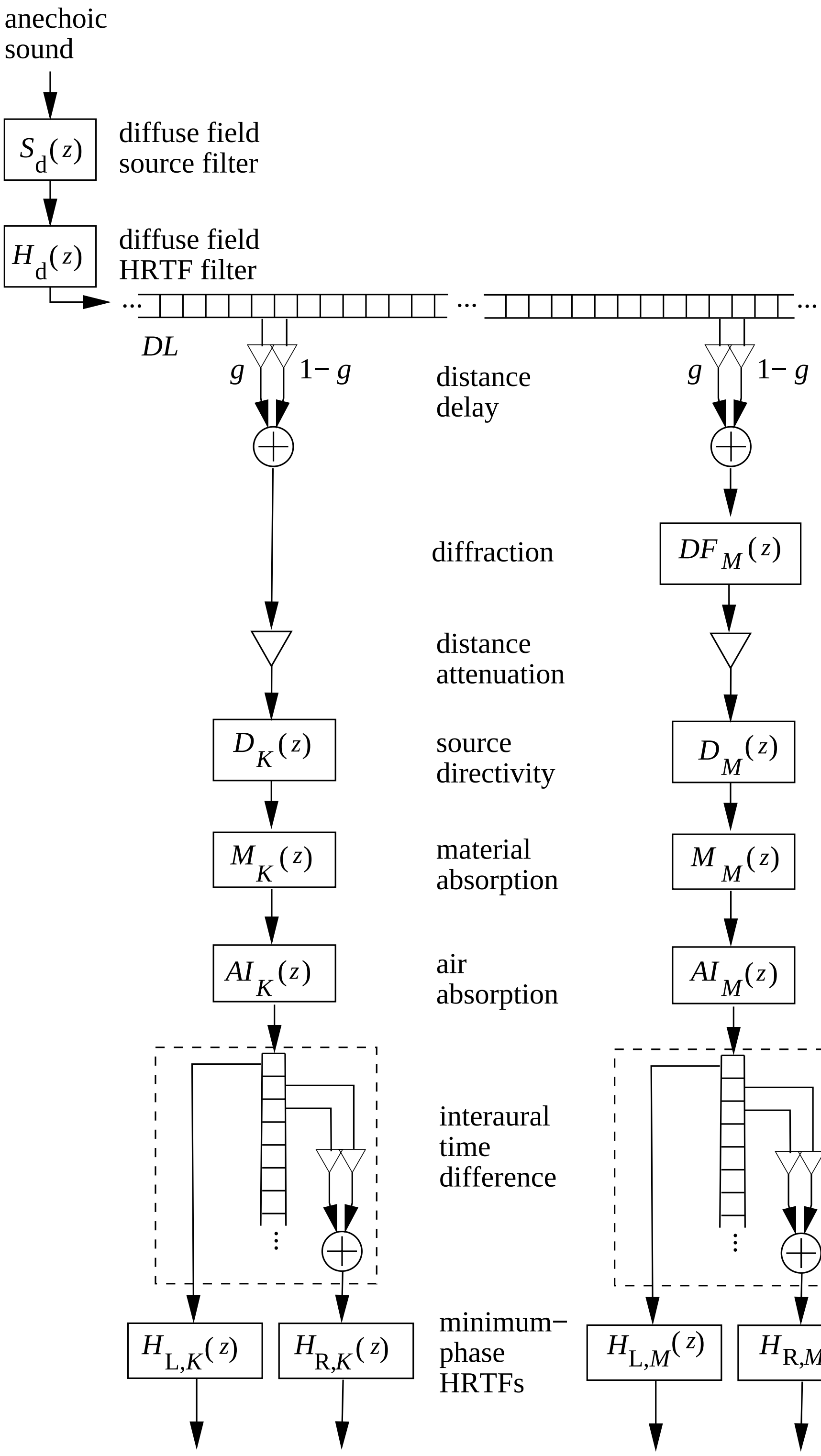


Figure 3.3: Processing of each image (on the left side) and edge (on the right side) source [P3].

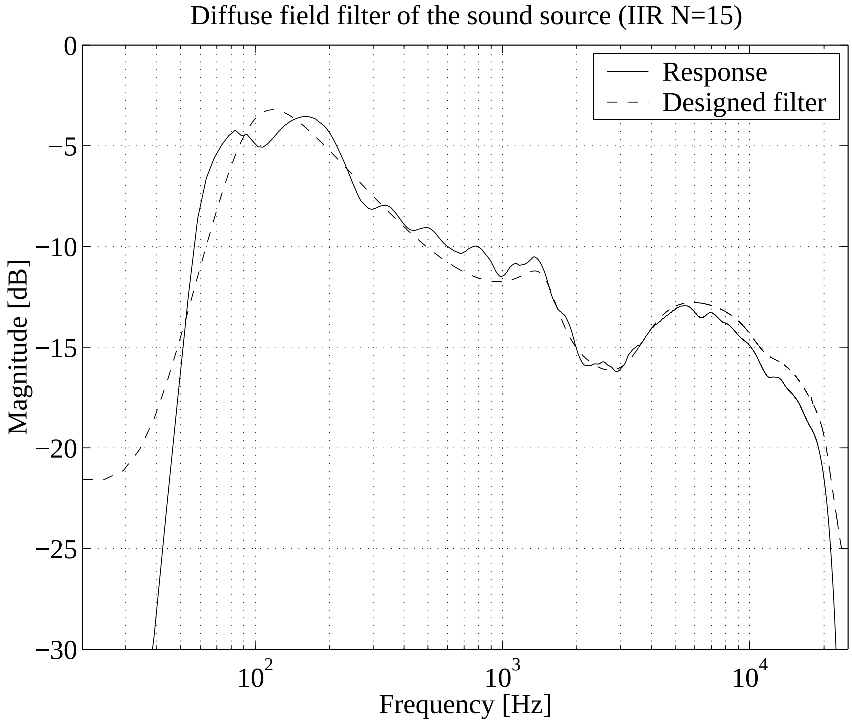


Figure 3.4: Magnitude response of the diffuse field filter of the sound source [P2].

Distance Delay and Attenuation

The anechoic signal to be processed is first fed to a long delay line DL which is implemented as a ring buffer. The pick-up points from DL are defined by distances of image or edge sources from the listener. In dynamic rendering the fractional delays [92] have to be utilized for the distance changes to obtain smooth and continuous output. In Figure 3.3 the simplest possible fractional delay, linear interpolation, implemented with a first order finite impulse response (FIR) filter is depicted. The slight high frequency attenuation caused by fractional delays is discussed in [92][P2].

The frequency independent attenuation according to $1/r$-law, caused by the spherical propagation of sound waves (see, e.g., [129]), is calculated with a single coefficient. In dynamic rendering, the linear interpolation of the distance attenuation coefficient between updates is applied as discussed in [P1].

Sound Source Directivity

The sound source used in this study was a small Genelec 1029A loudspeaker. For modeling purposes the impulse responses of the loudspeaker were measured in anechoic chamber from 24 different azimuth angles (every 15°) and 4 elevation angles (0°, 30°, 60°, and 90°), in total 96 responses. From these measurements a response grid of every 5° in azimuth and elevation was interpolated. The total number of directivity responses was 1368 and they were applied symmetrically for negative elevations, although the loudspeaker radiation characteristics are not strictly symmetrical in upper and lower hemisphere.

The sound source directivity filtering has been realized according to ideas presented previously [60, 76, 82]. The filtering is divided into two filters, the diffuse field filter ($S_{\mathrm{d}}(z)$ in Figure 3.3) and the diffuse field equalized directivity filter ($D_K(z)$ and $D_M(z)$ in Figure 3.3).

The target diffuse field power spectrum for diffuse field filter has been derived from all measured responses ($H_{\mathrm{dir},i}(\omega,\theta,\phi)$) by power-averaging

$$S_{\mathrm{d}}(\omega,\theta,\phi) = \sqrt{\frac{1}{k}\sum_{i=1}^{k} |H_{\mathrm{dir},i}(\omega,\theta,\phi)|^2 \cos\phi}, \tag{3.1}$$

where k is the number of responses, θ is the azimuth angle, ϕ is the elevation angle, and ω is the angular frequency. The actual filter fitting was performed as follows [P2]. First, the magnitude response was pre-processed to respect the nonlinear frequency resolution of the human hearing with frequency dependent smoothing and weighting according to Equivalent Rectangular Bandwidth (ERB) scale [118]. Then the filter design was executed applying a least squares method (`yulewalk.m` function in Matlab [106]). The magnitude response of the applied diffuse field filter, implemented with an infinite impulse response (IIR) filter of order 15, is depicted in Figure 3.4.

The final primary directivity filters ($D_K(z)$ and $D_M(z)$ in Figure 3.3) were designed in a warped frequency domain. The warped frequency scale utilized was an approximation of the Bark scale and the warping was done utilizing the conformal bilinear transform [160]. After filter fitting to magnitude response (with `yulewalk.m` function in Matlab [106]), the warped filter coefficients were unwarped back to the linear frequency scale with an algorithm proposed by Huopaniemi [57]. By this way the direct-form pole-zero IIR filter structure could be applied in auralization.

Filtering the input signal with the diffuse field filter $S_{\mathrm{d}}(z)$ has many advantages. For example, it makes the design of the primary directivity filters $D_K(z)$ and $D_M(z)$ much easier, by flattening the response at low and high frequencies. This is also called diffuse field equalization and it allows the use of lower order filters for the primary directivity filter. In this study, finally the IIR filters of order 6 was applied and three examples of the magnitude responses of filters are plotted in Figure 3.5.

In total 1386 filters were designed and stored in a table. During the auralization process this table was read into the memory during the initializing phase and suitable filter coefficients for each image and edge sources were chosen according to auralization parameters. In dynamic rendering two filters for each image and edge source were computed simultaneously and the smooth output between updates were obtained by cross fading the outputs of these two filters.

Material Absorption

The material absorption is modeled by utilizing the reflection filters $M_K(z)$ and $M_M(z)$ in Figure 3.3. The data for filter design is absorption coefficients available in octave bands, for an example see Table 3.1. When sound is reflected from two or more surfaces the absorption data can be cascaded so that only one filter is needed [61][P1]. The algorithm for realizing cascaded absorption coefficient data with a low-order IIR filter is as follows [61]. First, all possible absorption combinations are calculated and transformed into reflectance data. The resulting amplitudes are transformed into the frequency domain and filter fitting is performed in a warped frequency domain as in the case of sound source directivity filters. In this study, direct-form IIR filters of order 4 were applied. Examples of magnitude responses of applied filters, based on the absorption data in Table 3.1, are presented in Figs. 3.6 and 3.7. All the filters were calculated off-line and filter coefficients were associated to each image and edge source already in the image source calculation.

In the real world the reflection of a sound wave from acoustic boundary material is a complex one. The temporal or spectral behavior of reflected sound as a function of incident angle, the scattering and diffraction phenomena, etc., makes

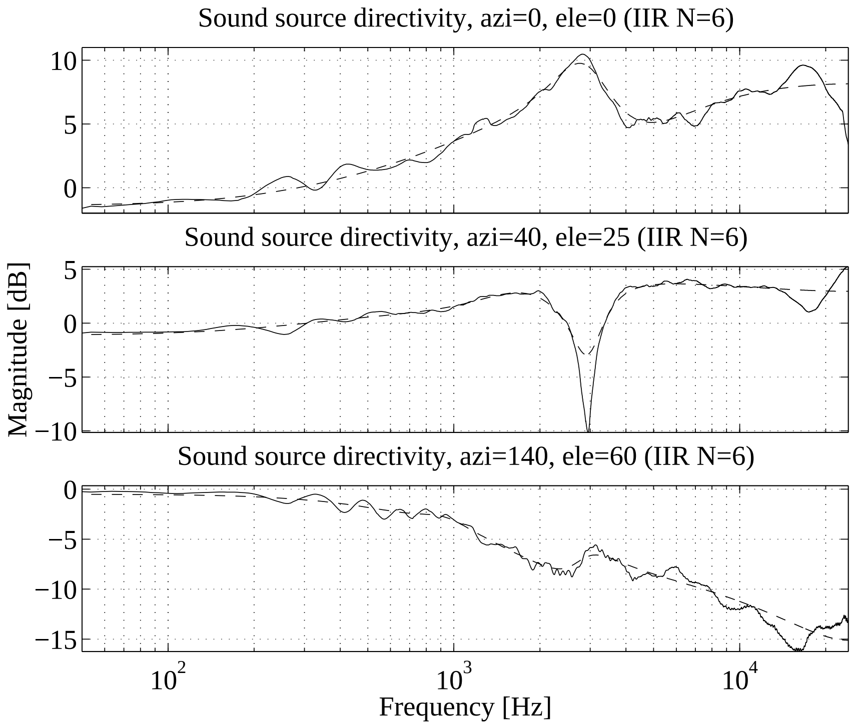


Figure 3.5: Three examples of the magnitude responses of the diffuse field equalized directivity filters. The solid lines are target responses and the dashed lines are designed responses.

Material name	ID	63 Hz	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz	16000 Hz
Rockfon panel	1	0.15	0.30	0.70	0.85	0.90	0.90	0.85	0.60	0.35
Wooden table	2	0.20	0.15	0.10	0.10	0.10	0.07	0.07	0.07	0.07
Concrete wall	3	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.02
Wooden table	4	0.02	0.02	0.05	0.10	0.15	0.25	0.30	0.30	0.30

Table 3.1: The absorption coefficients of employed materials in octave bands.

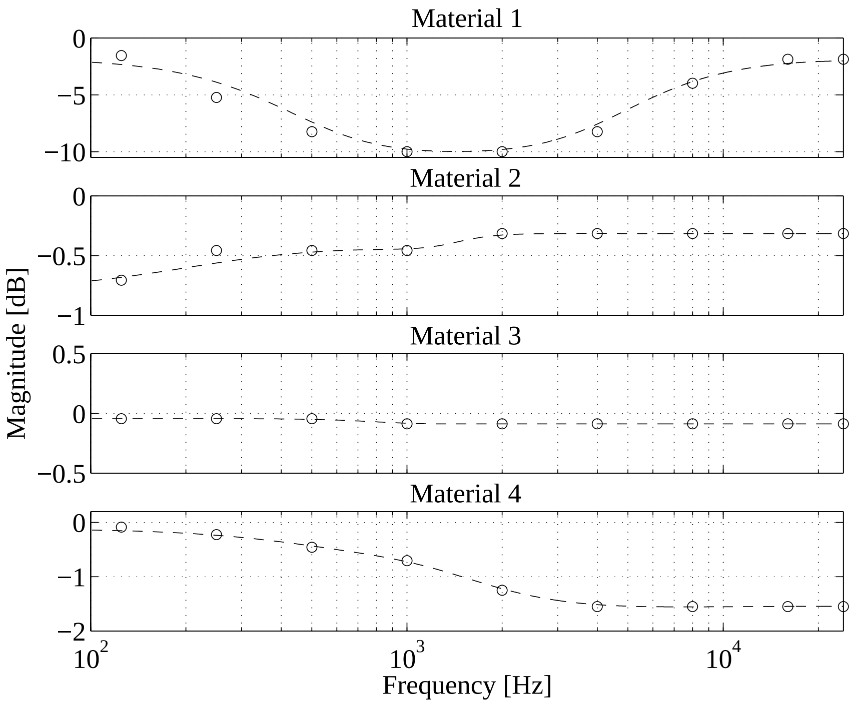


Figure 3.6: Magnitude responses of reflection filters for materials in the Table 3.1. The 'o' marks are target points and the dashed lines are magnitude responses of the designed filters.

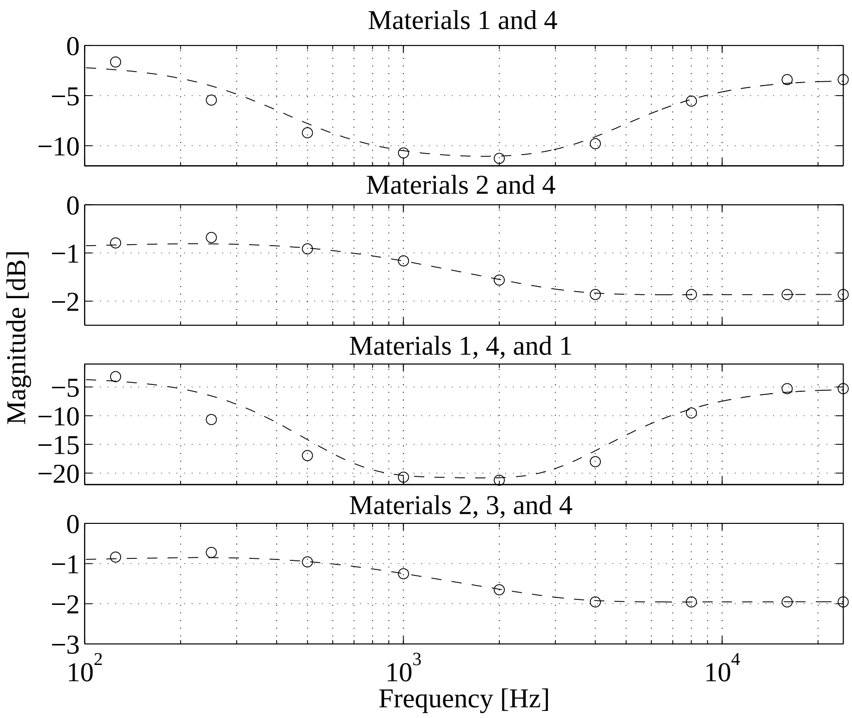


Figure 3.7: Four examples of combined (for two second order and two third order reflections) reflection filters. The 'o' marks are target points and the dashed lines are magnitude responses of the designed filters.

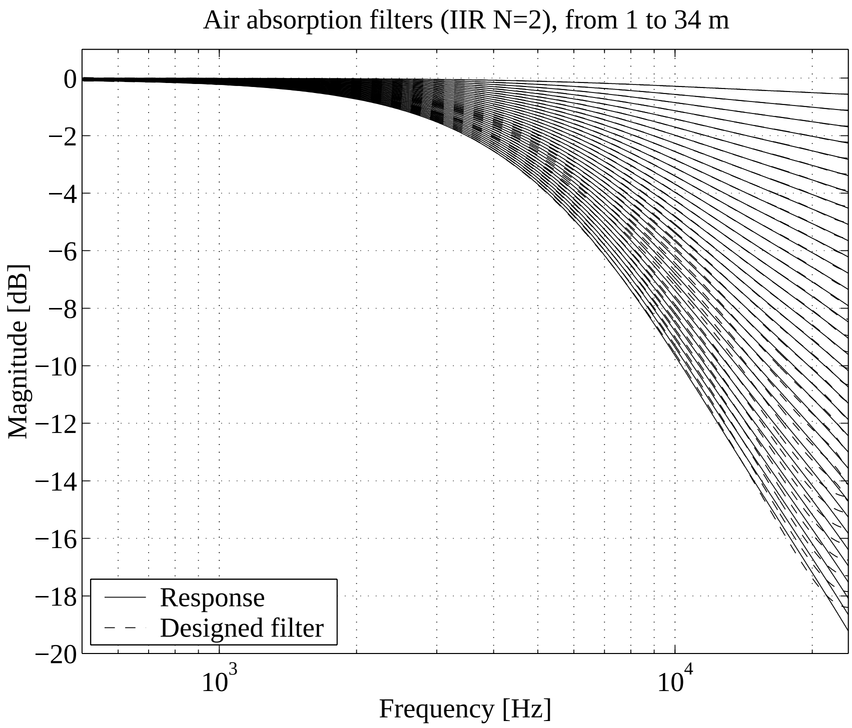


Figure 3.8: Magnitude responses of air absorption filters with one meter steps spacing at distances from 1 to 34 m.

it practically impossible to develop numerical models that are accurate in all aspects. Therefore, it is restricted to model only the angle independent absorption characteristics [P1].

Air Absorption

The absorption of sound in air depends mainly on the distance, temperature, and humidity. The equations for calculation of air absorption are standardized [2] and they were applied to calculate the target responses as explained in [57, 61][P1]. The filters were fitted to minimum-phase complex frequency responses using function `invfreqz.m` in Matlab [106]. Examples of the magnitude responses of designed filters, IIRs of order 2, are depicted in Figure 3.8. In dynamic rendering the filter coefficients need not to be interpolated, despite the use of recursive filters. Instead, filter coefficients can be changed without audible defects if the grid of filters is dense enough. In this case one meter steps spacing has been found dense enough.

Edge Diffraction

Svensson et al. [163] have derived a mathematical solution for calculating the impulse response for an edge of a finite length. With this analytical solution the edge diffraction is modeled within the DIVA auralization system. The auralization parameters for an edge source are:

- wedge angle θ_{w},
- position of source $S(X, Y, Z)$,
- position of receiver $R(X, Y, Z)$,
- start and end point of the edge $Z_0(X, Y, Z)$; $Z_1(X, Y, Z)$,

	edge length	m via Z_{apex}	l via Z_{apex}	$180\pi\theta_{\text{w}}$	$180\pi\theta_{\text{S}}$	$180\pi\theta_{\text{R}}$
edge1	6.90 m	1.72 m	2.30 m	270.0°	166.0°	245.0°
edge2	1.40 m	8.37 m	6.04 m	359.0°	3.7°	9.5°
edge3	2.80 m	3.29 m	1.60 m	359.0°	169.3°	130.4°
edge4	1.20 m	2.48 m	3.67 m	270.0°	226.6°	164.1°
edge5	2.20 m	0.53 m	3.81 m	270.0°	114.0°	107.3°
edge6	8.50 m	2.66 m	5.29 m	99.5°	71.9°	91.6°
edge7	3.00 m	2.61 m	5.36 m	189.5°	71.9°	91.6°

Table 3.2: The data for calculation the seven edge response examples. Note that all θ angles are in degrees, but in Equations (3.2) and (3.3) they are expressed in radians.

- normal vector $\vec{n}$ of a surface.

With this data for each edge, one of which is illustrated in Figure 3.9, the impulse response from a source via edge to a receiver is calculated with the following equations [163][P3]

$$h_{\text{diffr}}(t) = -\frac{\nu}{4\pi}\int_{Z_0}^{Z_1}\delta\left(t-\frac{m+l}{c}\right)\frac{\beta_{++}+\beta_{+-}+\beta_{-+}+\beta_{--}}{ml}dZ, \tag{3.2}$$

$$\beta_{\pm\pm} = \frac{\sin[\nu(\pi\pm\theta_{\text{S}}\pm\theta_{\text{R}})]}{\cosh\left(\nu\cosh^{-1}\frac{1+\sin\alpha\sin\gamma}{\cos\alpha\cos\gamma}\right)-\cos[\nu(\pi\pm\theta_{\text{S}}\pm\theta_{\text{R}})]}. \tag{3.3}$$

The variables can be found in Figure 3.9 where a finite wedge is illustrated. In addition, c is speed of sound, $\nu = \pi/\theta_{\text{w}}$ is the wedge index, m is the source-to-edge point distance, and l is the edge point-to-receiver distance. The integration range is between the two end points of a finite edge.

For efficient auralization a warped IIR filter of order 3 was fitted to calculated magnitude response [P3]. Seven examples of diffracted responses in the studied lecture room, based on the data in Table 3.2, are depicted in Figure 3.10.

The diffraction filter implements the diffraction phenomenon as an impulse response at one point [P3], but in the real life diffraction is not point-like. Sound passes the edge through all points along the edge, however, most of the energy is concentrated to the least-time point of the edge. Based on this, the simplification to apply point-like secondary source is not so severe. In addition, the edge source, being a point source, can be panned to the direction where the least-time point indicates as proposed by Torres et al. [171]. The same principle holds for the sound source directivity, since from the viewpoint of the edge most of the sound energy from the actual source radiates towards the least-time point of the edge [P3]. The situations where this simplification could be most audible would be long horizontal edges that are close to the listener.

In the current implementation the edge diffraction filters are designed between image source calculation and auralization process. Thus, this implementation is not practical for real-time use, but dynamic off-line rendering is straightforward to perform. The same interpolation technique, employing two cross-faded filters, as used for source directivity filters was applied in dynamic rendering.

Binaural Processing

The required directional filtering of each image and edge source (blocks $F_{0\ldots N}(z)$ in Figure 3.2) depends on the available 3D sound reproduction method. In this thesis only binaural reproduction is considered, but in addition the vector base amplitude panning (VBAP) [133] has been implemented within the DIVA system [135].

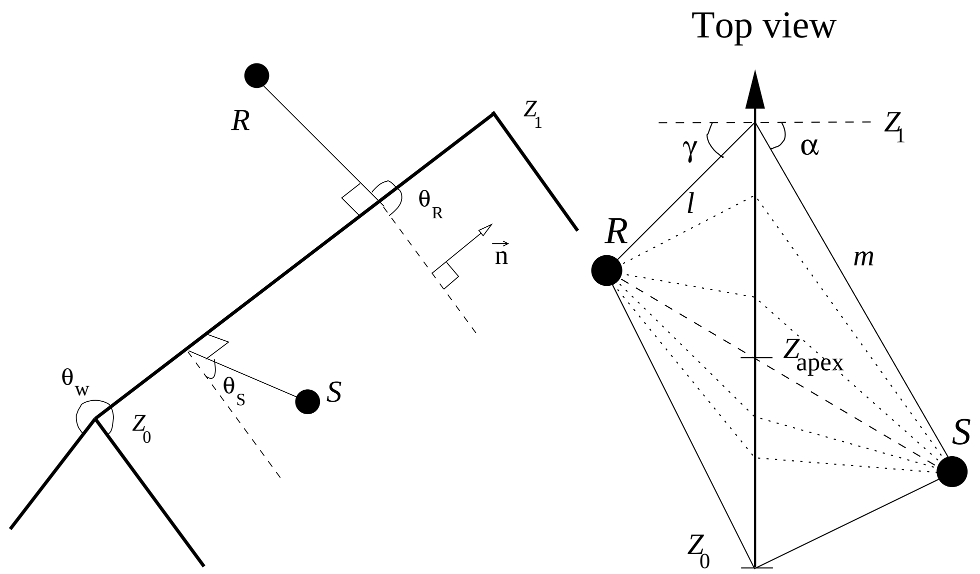


Figure 3.9: Geometry of a finite wedge. On the right sound paths via edge points Z_0 and Z_1 are indicated by the solid lines, the least-time sound path via the apex point Z_{apex} is depicted with dashed line and some other sound paths are illustrated with dotted lines.

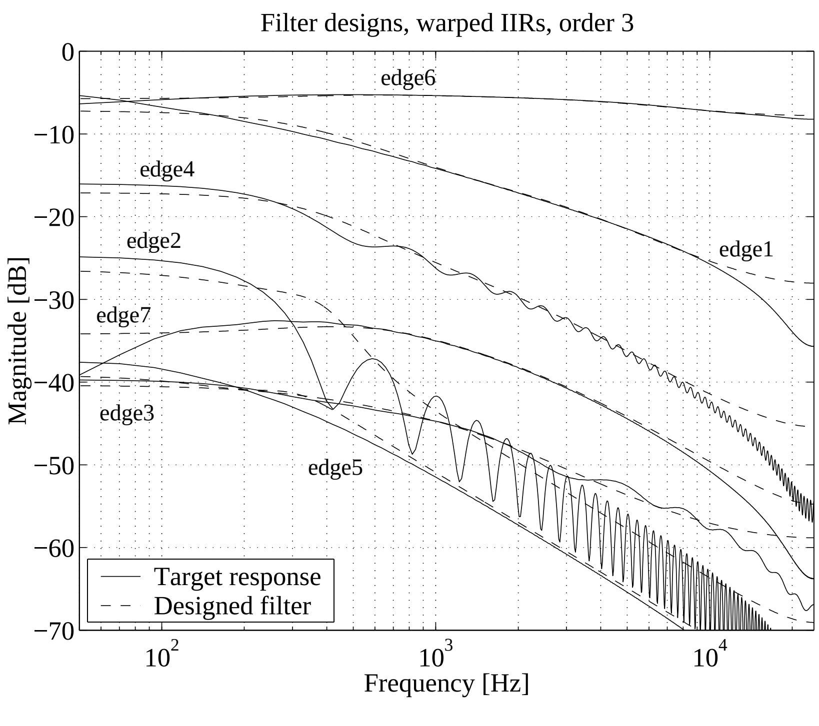


Figure 3.10: Seven examples of frequency responses of diffracted components in the modeled lecture room. The responses were calculated based on the data in Table 3.2. The oscillations of the target responses are caused by the length of the finite edge. The distance attenuation $(1/r)$ is compensated from responses.

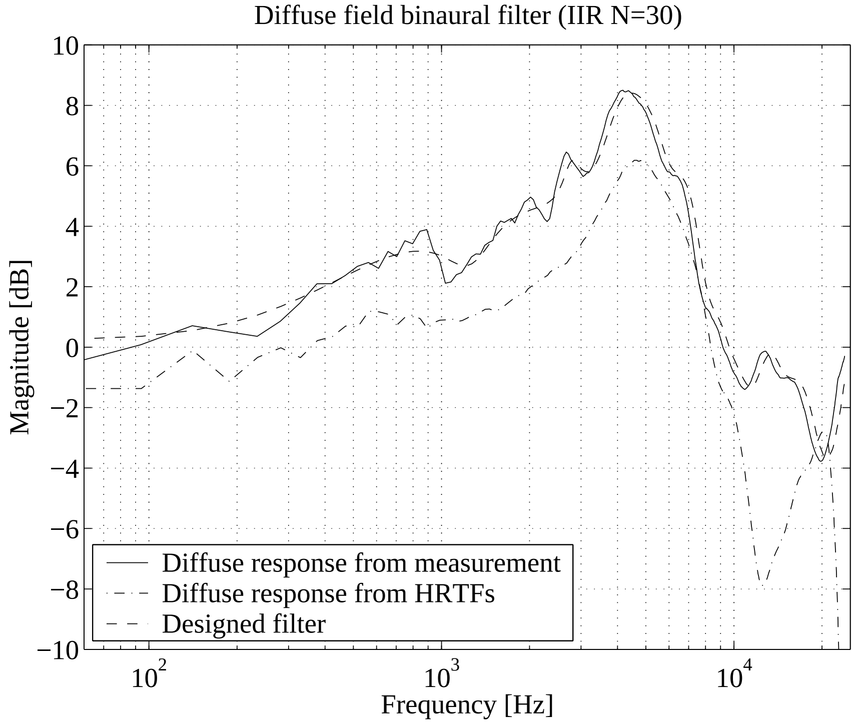


Figure 3.11: Magnitude response of the diffuse field binaural filter applied in auralization.

The headphone reproduction of 3D sound requires the modeling of human directional hearing. The modeling is usually done by measuring the head-related transfer functions (HRTF) and by fitting digital filters to measured responses. For headphone sound reproduction these filters modify sound signals as they would be modified by shoulders, head, and pinnae in the real listening conditions. The human directional hearing and the properties of HRTFs are the most intensively studied areas related to auralization.

In this thesis the applied HRTFs were measured from the author [138]. The measurements were completed in anechoic chamber from seven elevation angles (-30°, -15°, 0°, 15°, 30°, 60°, and 90°) and from 36 azimuth angles (every 10°), in total 252 directions. The filter design was mainly based on the work of Huopaniemi [57][P1].

Binaural filtering was realized with a diffuse field filter ($H_{\mathrm{d}}(z)$ in Figure 3.3) and with diffuse field equalized HRTF filters [P2]. The diffuse field response for filter design was obtained with the method proposed by Larcher et al. [94]. In this method the real-head impulse responses and a monophonic impulse response are measured in a normal room, at the same position. In this study, the diffuse field binaural response was estimated from the period of time after the direct sound and early reflections of the measured responses. The time interval between 50 ms and 75 ms were chosen. The monophonic impulse response was deconvolved from binaural responses and the remaining frequency responses were the diffuse field binaural responses. Finally, the applied diffuse field response was achieved by averaging across several measurements and from both ears. The actual filter design method applied was the same than for the diffuse field filter of the sound source. The utilized diffuse field response and the magnitude response of the designed filter are depicted in Figure 3.11.

The actual angle-dependent binaural filters were divided into interaural time difference (ITD) part (implemented with a pure delay line) and a minimum-phase

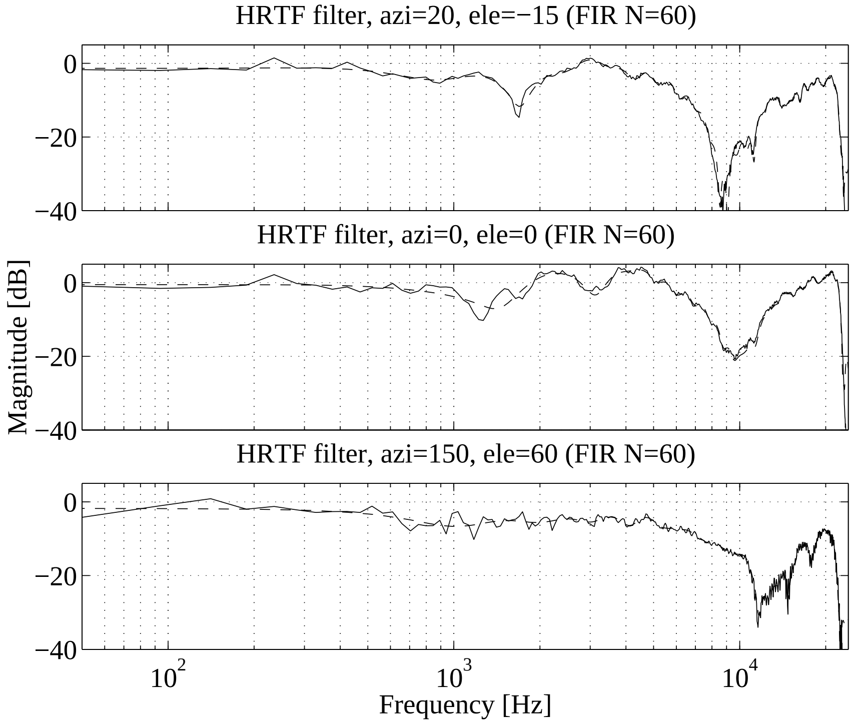


Figure 3.12: Three examples of the magnitude responses of angle-dependent diffuse field equalized HRTF filters. The solid lines are target responses and the dashed lines are the responses of designed filters.

counterpart of diffuse field equalized HRTF as illustrated in Figure 3.3. The ITD was calculated applying a spherical head based ITD model (discussed in, e.g., [17]) with added elevation dependency [P1]. In the design of final angle-dependent binaural filters the measured head-related impulse responses (HRIR) were first compensated with measurement system response, minimum-phased and transformed to the frequency domain, divided with diffuse field magnitude response, and finally the filter fitting was done with the least squares method. The applied filters were FIR filters of order 60 and the magnitude responses of three of them are depicted in Figure 3.12. During the auralization process, from directions where no filter exists the filter coefficients were calculated by bilinear interpolation from the four nearest available data points [P1]. In dynamic rendering, the filter coefficients were changed without interpolation between outputs if the change in direction was small enough, typically not more than a few degrees.

Late Reverberation

The late reverberation in a room is often considered nearly diffuse and the corresponding impulse response exponentially decaying random noise [154]. Under these assumptions the late reverberation does not have to be modeled as individual reflections with certain directions. Therefore, to optimize computation in late reverberation modeling, recursive digital filter structures have been designed, whose responses model the characteristics of real room responses, such as the frequency dependent reverberation time.

The applied late reverberation algorithm [177] contained eight parallel feedback loops, with comb-allpass filters in each loop (see Figure 3.13). The algorithm is a simplification of a feedback delay network (FDN) structure [70, 144] and produces natural sounding late reverberation. The comb-allpass filters in the feedback loops, denoted by $A_{1\dots8}(z)$, are added to produce an increased reflection

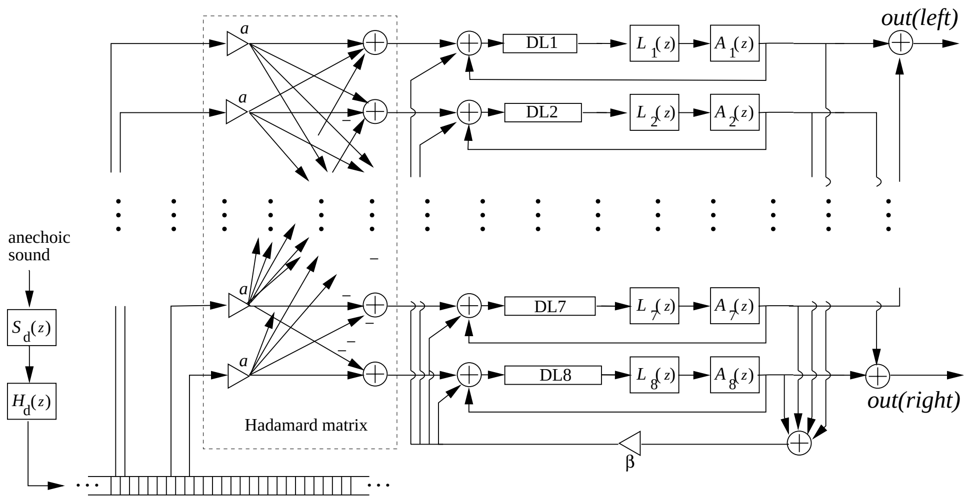


Figure 3.13: Late reverberation algorithm. The filters $H_{1...8}(z)$ are lowpass one-pole IIR filters and $A_{1...8}(z)$ are comb-allpass filters.

	Channels n in reverberator							
	1	2	3	4	5	6	7	8
pick-ups from main DL	881	953	1031	1151	1217	1307	1427	1579
lengths of DLn	1031	1129	1259	1307	1433	1531	1693	1777
delays inside $A_n(z)$	137	157	167	173	181	199	211	227

Table 3.3: The delay line lengths applied in the reverberator. All numbers are samples (f_s = 48 kHz).

density. The filters $L_{1...8}(z)$ implement the frequency dependent reverberation time. Each of them contains a simple one-pole lowpass filter whose parameters are calculated automatically based on the required reverberation time at low and high frequencies [75]. The lengths of the delay lines in the loops should be mutually incommensurate in samples to avoid reflections occurring at the same time, and strong coloration caused by coinciding modes in the frequency domain [154].

The inputs to the reverberator are picked directly from the main propagation delay line DL. The anechoic signal is at this point already filtered by the diffuse field filters of the sound source and HRTFs. In this manner the spectrum of late reverberation is modified with the power spectrums of a sound source and a listener as suggested by Jot et al. [76]. In the applied implementation the fixed pick-up points were arbitrarily chosen so that the first outputs of the late reverberation slightly overlapped with last early reflections. To increase the reflection density and to make reverberation as diffuse as possible the input signals were fed through a Hadamard matrix. The other advantage produced by the Hadamard matrix is that it makes reverberation outputs highly uncorrelated. However, this lack of correlation was not frequency dependent as in real life cases where high correlation at low frequencies occurs. The low frequency correlation can be achieved with a correlation filter [70, 104], but it was not implemented. The applied delay line lengths as well as pick-up points from the delay line DL are collected to Table 3.3.

In the presented implementation the late reverberation was not modeled with the geometry under rendering. On the contrary, it has been considered as diffuse reverberation and modeled with an algorithm that produced natural sounding reverberation. However, with this approach such rooms that have flutter echos or non-exponential decay cannot be modeled. To be more accurate the late reverberation should be modeled, e.g., by applying ray-tracing. In real-time rendering the modeled reverberant tail can be efficiently implemented, for example, by using real-time convolution in the frequency domain [45].

3.3 Considerations About Real-time Auralization

The presented signal processing chain can be used in real time, if a powerful enough computer is available. The need for processing power depends on many factors, such as the number of image sources, sampling rate, latency requirements, etc. However, the filter designs were not optimized from the computational point of view. All filters, except edge diffraction filters were designed before the image source calculation process and stored to data structures. The auralization parameters were used to access these tables for suitable filter selection for each image source. The diffraction filters could not be computed before the image source calculation but their efficient design in real-time with, e.g., neural networks has been proposed in [P3].

The sound source directivity and HRTF filters require the most computational power in the presented implementation. However, much more efficient modeling of HRTFs has been proposed, e.g., in [85, 103]. In this solution, based on principal component analysis, the whole HRTF data set was represented with five basis functions to which filters were fitted. Employing this technique each HRTF filter could be reconstructed with a linear combination of these five filters, i.e., for each image and edge source only five coefficients were needed. Such an implementation is very efficient in cases when a great number of image and edge sources are filtered. Different techniques to HRTF data set decomposition have been compared by Larcher et al. [93]. The same kind of decomposition of sound source directivity data could be applied for efficient implementation of source directivity.

4 EVALUATION OF AURALIZATION QUALITY

In this chapter the evaluation of quality of the DIVA auralization system is presented. Most of the results have been reported earlier [P5-P7] and only the final evaluation results are presented here.

4.1 Background

The evaluation of auralization quality was based on the framework which was first defined in [P2] and then redefined in [P6] and finally in [P7]. In this framework, illustrated in Figure 4.1, the assessment was performed by comparing recorded and auralized soundtracks. The recordings made in the studied lecture room, which was also modeled for auralization purposes, were considered as reference signals. The idea of such a comparison of recorded and auralized signals has been proposed earlier [20, 86, 131], but in this thesis both static and dynamic renderings were compared both with objective and subjective methods.

Evaluation was performed twice objectively [P5, P7] and subjectively with three reported [P6, P7] listening tests and one further informal test. The reason for this iterative assessment is that after each study models and filter designs were refined. In addition, the evaluation methodology has been developed during the whole process since no widely accepted methodology to study auralization quality or quality of spatial sound reproduction is agreed upon within the research community.

Creation of the Soundtracks for Comparison

The reference signals were measured in the studied lecture room, by applying a real-head recording technique [101, 110, 115]. In addition, binaural and monophonic impulse responses were measured in several receiver positions. The sound source was a Genelec 1029A loudspeaker. To capture spatial sound small electret microphones (Sennheiser KE 4-211-2) were positioned at the entrance of ear canals. Besides the binaural recordings, the monophonic reference signals were recorded with a high-quality omnidirectional microphone (B&K 4192).

The auralizations were prepared by computing image sources up to fourth order with and without diffraction modeling. Only the first order edge diffraction[1], meaning that only one of the secondary sources (reflections) could be an edge source, were considered. This decision was based on a finding by Torres et al. [171] that second order diffractions are usually not audible. Both monophonic and binaural soundtracks were computed and background noise was added to each soundtrack (signal-to-noise ratio was about 25 dB), as illustrated in Figure 4.1.

The quantities of visible image and edge sources in the receiver positions r1 and r2 (see Figure 4.1) are collected to Table 4.1. It can be noted that the number of edge sources was high compared to image sources due to the reason that each table in the geometry produces four edge sources. In addition, almost all edges were visible to all other surfaces, providing a huge number of higher order secondary sources.

[1] Sound paths in which one of the four reflections can be a diffraction are considered as first order edge sources. For example, such sound path as diffraction-specular-specular-specular or specular-specular-diffraction-specular can occur.

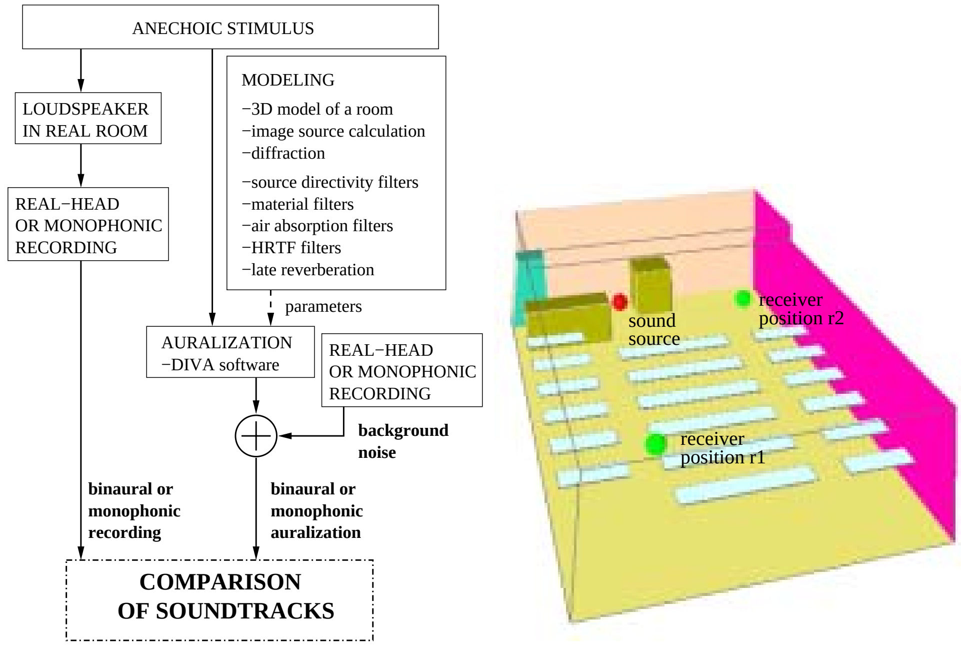


Figure 4.1: The framework utilized in the evaluation of auralization quality. On the right the geometry of the studied lecture room is drawn and the positions of the sound source and two receiver positions, utilized in the final evaluation, are marked. In both receiver positions the listener was looking straight ahead.

	Receiver position r1				Receiver position r2			
Order	IS	ES	All	Total	IS	ES	All	Total
0	0	0	0	1	0	0	0	1
1	6	90	96	97	5	92	97	98
2	13	905	918	1015	11	875	886	984
3	22	4596	4618	5633	18	4389	4407	5391
4	29	15464	15493	21126	28	14881	14909	20300

Table 4.1: Visible image and edge source quantities in the receiver positions r1 and r2. *IS* denotes for the number of image sources, *ES* denotes for edge sources, *All* denotes for all sources together, and *Total* represents the cumulative sum of the direct sound (Order 0), image, and edge sources.

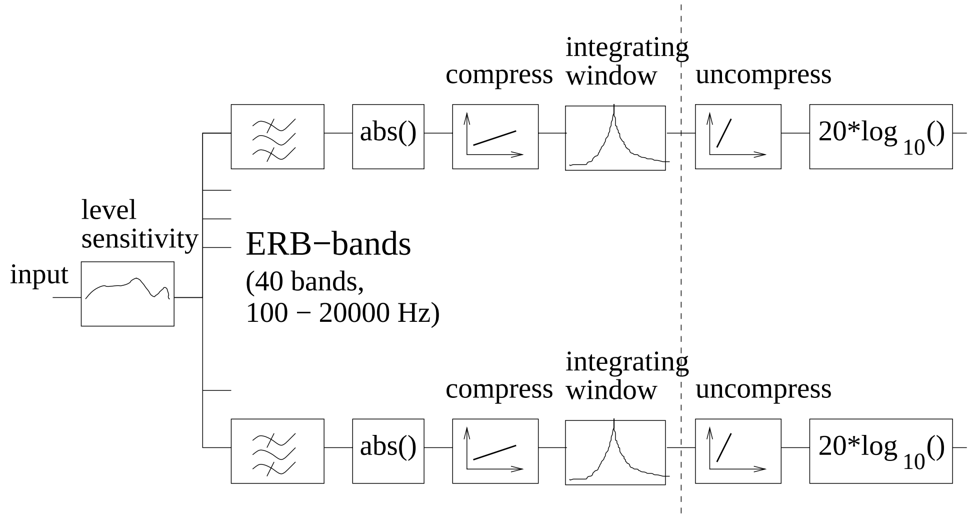


Figure 4.2: Block diagram of the monaural analysis method motivated by auditory perception [P4].

4.2 Methods

The evaluation was performed objectively by calculating different room acoustical parameters and subjectively by organized listening tests. In this section applied methods are introduced.

Objective Methods

Traditionally, the quality of room acoustics has been described with objective parameters [10, 15, 67, 90] that can be calculated from measured impulse responses. In this thesis the following parameters were utilized:

- Reflection density, calculated by computing individual reflections in a 20 ms sliding window (with 10 ms hops) in each of which the peaks within 20 dB from the strongest reflection were counted for defining the density [51]. The algorithm has been implemented by detecting the absolute values of the response which exceed the neighboring values as proposed in [177].

- Reverberation time (T30), early decay time (EDT), and clarity (C50), which are defined in the ISO3382 standard [67].

For more detailed objective analysis a monaural auditory model was developed [P4] and it was applied in analysis of responses separately for each ear. The aim in auditory modeling is to find mathematical models which represent some physiological and perceptual aspects of human hearing. The applied model is not an accurate auditory model, but it mimics the behavior of human cochlea. Therefore, with the proposed method, responses and signals can be visualized with similar time and frequency resolution than human hearing has.

A block diagram of the developed analysis method is presented in Figure 4.2. First block models the level sensitivity of human auditory system with a frequency weighting filter, such as the inverse of the 60 dB equal loudness curve [66]. Then the signal is fed into a gammatone filterbank [159] which divides it into 40 ERB-scale bands [118, 117], simulating the frequency resolution of human ear. After the division into ERB-scale bands absolute signal values are taken. For implementation reasons absolute values are used instead of the half-wave rectification which happens in the hair cells of human ear. The next stage in the analysis is formed by a compression and a sliding window which together roughly simulate the time resolution of the ear [130]. The final step of the analysis is to use a proper mapping for visualization purposes. By uncompressing and taking the logarithm of the rectified and temporally processed signal in each frequency band the decibel values can be depicted in a time-frequency plot. The sone-related loudness scale could be utilized as well.

The proposed analysis method has been an excellent tool in the development of the DIVA auralization system. Modeled responses and auralized signals have been studied in the time-frequency domain and the visualizations intuitively highlight possible audible features of the responses. It has also proved to be helpful for explaining the results of a listening test [P7].

Subjective Methods

To find out subjective perceptual differences between the recorded and the auralized soundtracks several listening tests have been carried out. Different listening test methods were tested as no recommended listening test methodology for testing the auralization quality exists. Finally, the last evaluation was done with a method proposed for audio codec quality testing, since in the comparison of the recorded and the auralized soundtracks the aim was to look into the subjective assessment of small differences. The chosen comparison method, recommended in ITU-R BS.1116-1 [68], was double-blind triple stimulus with hidden reference method. It is intended for use in the assessment of systems which introduce impairments (or differences) so small as to be undetectable without rigorous control of the experimental conditions [68]. The recommendation was not strictly followed in the listening test. For example, requirements regarding the listener selection and exhaustive training of subjects were not possible within the performed test.

The auralization quality has many different aspects and it is multidimensional by nature. Because of the lack of better knowledge on these dimensions only two attributes, namely spatial and timbral differences, were studied in the final listening test. These two attributes have been found to be common terms for subgroups of descriptive attributes of spatial sound [193] and it has been claimed that people can reliably discriminate between spatial and timbral cues when evaluating room acoustics [11].

The listening test was conducted using the GuineaPig2 software [64, 65]. The answering dialog is illustrated in Figure 4.3. Totally 20 subjects (three females and 17 males) participated in the final listening test. All of them reported normal hearing although this was not verified with audiometric tests. The test was done in a standard listening room and the headphone reproduction method was applied with Sennheiser HD-580 headphones.

The listening task was to compare spatial and timbral differences between the recorded and the auralized soundtracks. Subjects were told to quantify sound source location, size of space, and reverberation when considering spatial differences. Similarly, such attributes as color of sound and frequency content were advised to be listened for judging timbral differences. The answering scale was from "very annoying" to "imperceptible" (see Figure 4.3) as recommended in the ITU-R BS.1116-1 [68]. Each answer corresponded to a decimal value from 1.0 to 5.0 in steps of 0.1, score 1.0 being for "very annoying".

Both the recorded and the auralized soundtracks, with durations from 2 to 3 seconds, were played in parallel to a listener who could switch between them (cross fading time was 40 ms). In the applied double-blind triple-stimulus hidden-reference method the reference signal was either signal A or signal B and subjects were forced to grade this hidden reference to be "imperceptible". Then the other signal (A or B) was judged against the reference. In this way reliability of the subjects could be evaluated all the time.

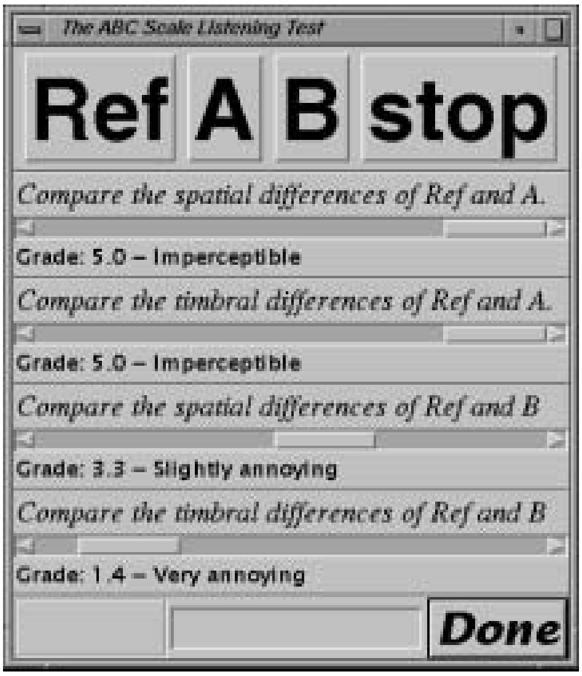


Figure 4.3: The graphical user interface (GUI) utilized in the listening test.

The rendering cases were limited to two static ones so that subjects could comprehensively concentrate on comparison of the auralization quality. While the dynamic rendering usually is beneficial with headphone reproduction, reducing such problems as front-back confusion [14] and in-head localization [34], it was noted in the previous test that dynamic rendering did not help in the comparison of the soundtracks [P6].

All subjects were trained with four tasks which were listened before the test under surveillance of the test supervisor. During the training session subjects learned to utilize the GUI and they also familiarized themselves with the tasks and the grading scales.

The whole listening test contained 36 tasks which were listened to in two groups (first 24 binaural and then 12 monophonic tasks). The playing order of these tasks inside a group was randomized. The tested variables, which were considered to be relevant, were the following:

- three stimuli (sound of a clarinet, sound of a guitar, and snare drum hits),
- two receiver positions (r1 and r2),
- two auralization methods (with and without diffraction).

In addition, twelve of the 24 binaural tasks were lowpass filtered to contain only frequencies below 5 kHz[2], while the other 12 binaural and 12 monophonic tasks contained the whole audible frequency range (f_s = 48 kHz). The total number of different tasks was obtained by combining all the variables (3x2x2x3 = 36 tasks).

Based on previous evaluations [P5, P6], the following hypotheses were considered:

- the binaural soundtracks have differences, especially at frequencies above 5 kHz,

[2]The binaural recordings sounded unnaturally bright and some recording artifacts above 5 kHz were suspected.

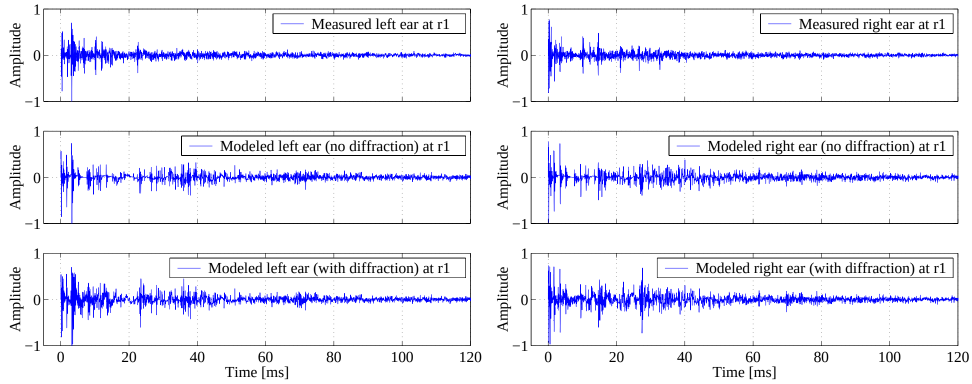


Figure 4.4: Measured and modeled binaural impulse responses at the receiver position r1.

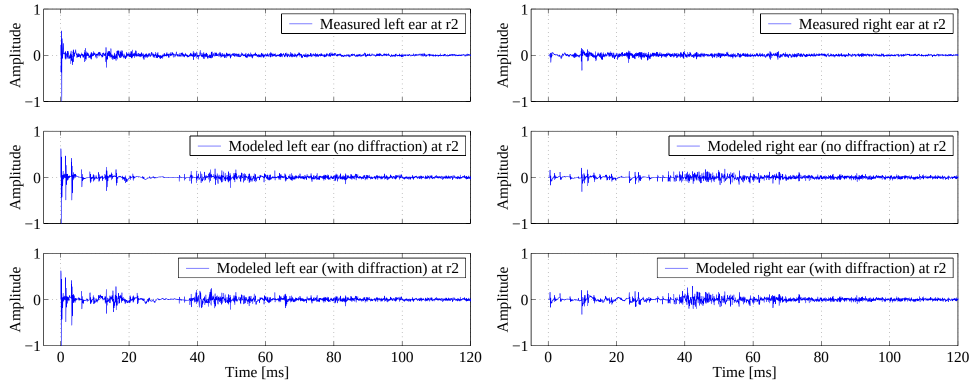


Figure 4.5: Modeled and measured binaural impulse responses at the receiver position r2.

- there are only slight audible differences between the auralized and the recorded monophonic and lowpass filtered binaural soundtracks,
- the inclusion of diffraction modeling is needed in auralization even when the sound source is not in the shadow.

In the following, both objective analysis and the listening test results are presented to test the validity of these hypotheses. In addition, the reasons to study these hypotheses are explained in more detail.

4.3 Objective Analysis

For objective examination of the auralization quality binaural impulse responses were measured and simulated in both receiver positions, besides the recordings and the auralizations of stimulus signals. Unfortunately, such standardized [67] parameters as reverberation time and clarity index are monaural, thus the spatial characteristics of auralization cannot be studied with them. For this reason, the objective analysis was performed only separately for responses of both ears. The measured and modeled BRIRs in both receiver positions are depicted in Figures 4.4 and 4.5.

The objective parameters were calculated from responses and results are plotted in Figure 4.6. The reflection densities in the beginning of the impulse responses seem to be lower in the auralizations than in the measurements. Diffraction modeling slightly raised reflection density but not as much as was expected based on Table 4.1. A possible explanation might be that the diffracted components from the table edges had so little energy that they did not produce peaks within 20 dB from the strongest reflections. The reverberation times indicate that at low frequencies below 400 Hz the auralizations were less reverberant than the measurements, but at high frequencies reverberation times coincide quite well. The EDT values have more deviation than reverberation times and diffraction modeling seems to lower the EDT values at frequencies below 4 kHz. Finally, the clarity values have quite large variance and no clear trends are seen between the measurements and the auralizations. Maybe it should be noted that clarity indices of the auralizations with diffraction modeling are similar to measurements in ipsilateral ears (at r1 right ear and at r2 left ear). In conclusion, it can be stated that with traditional room acoustical parameters no clear picture regarding the nature of differences between the auralizations and the measurements can be seen.

To analyze the impulse responses, plotted in Figures 4.4 and 4.5, by simulating auditory perception, the proposed analysis method [P4] was utilized and the results are depicted in the left columns of Figures 4.7-4.10. In addition, visualizations of the differences between measured and auralized responses are presented in the right columns. These difference plots were computed by subtracting two of the left column plots from each other. In the following list a few findings of these visualizations are pointed out.

- From all difference plots between the measurements and the auralizations it can been seen that the late reverberation (starting around 40 ms after the direct sound) had slightly more energy in the auralizations than in the measurements since the light areas are dominating.
- High frequencies above 6 kHz in the measurements contained much more energy in the beginning of the responses compared to the modeled responses.
- Visualizations in Figures 4.7 and 4.8 indicate that diffraction modeling raised the energy of early sound in the auralizations especially at low frequencies. However, the auralizations without diffraction seem to be closer to the measured ones.
- Contrary to the receiver position r1, in the position r2 (Figures 4.9 and 4.10) the auralizations with diffraction seem to be closer to the measurements, at least in the beginning of the responses.
- Plots in Figure 4.9 show clearly the gap between the early reflections and late reverberation in the auralizations (see also Figure 4.5).

The above observations indicate that some differences between the recordings and the auralizations could be audible. In the following the results of subjective listening test are presented.

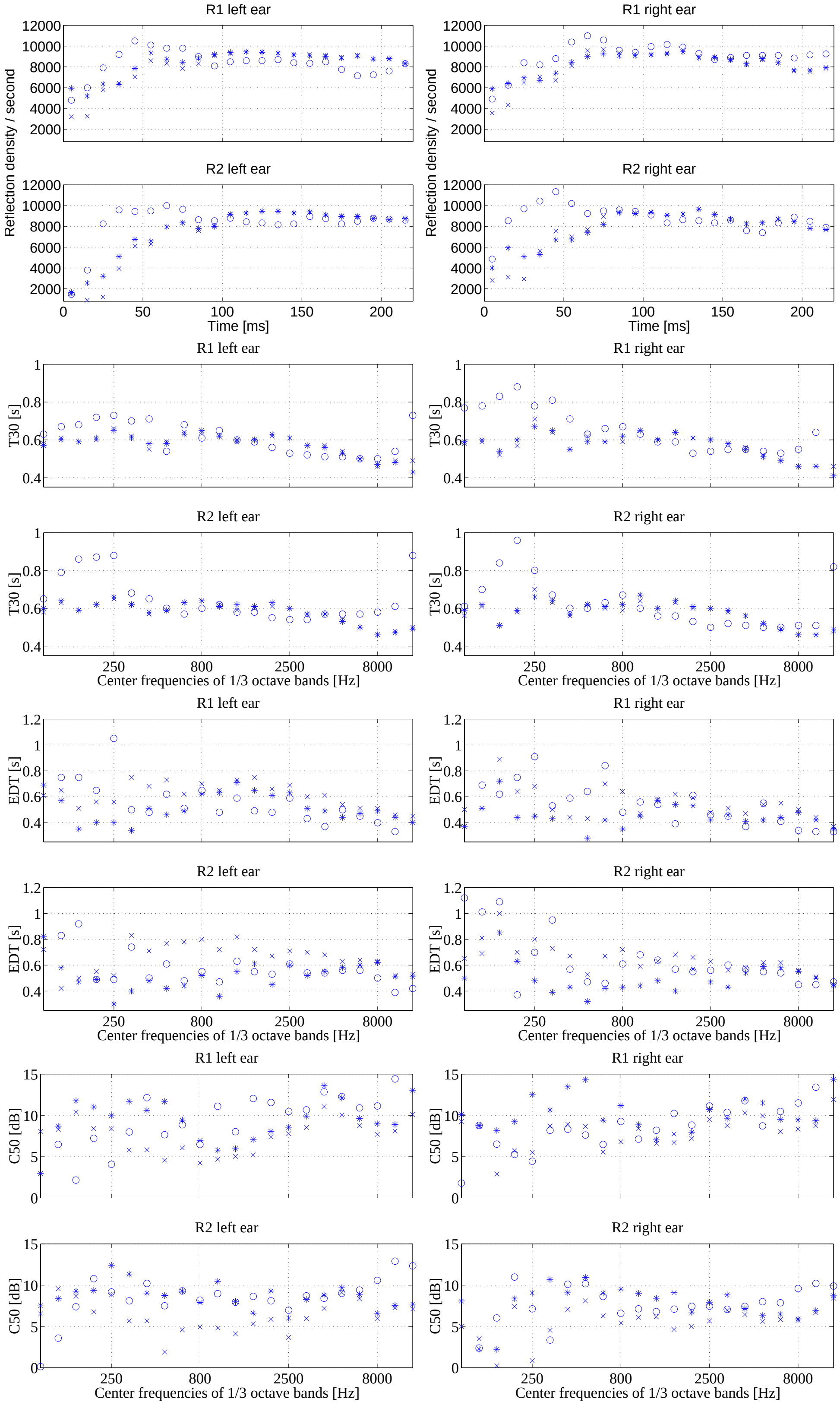


Figure 4.6: Objective parameters calculated from impulse responses depicted in Figures 4.4 and 4.5. From top: reflection densities, reverberation times, early decay times, and clarity indexes. Markers are the following: o = measured, x = auralized without diffraction, and ⋆ = auralized with diffraction.

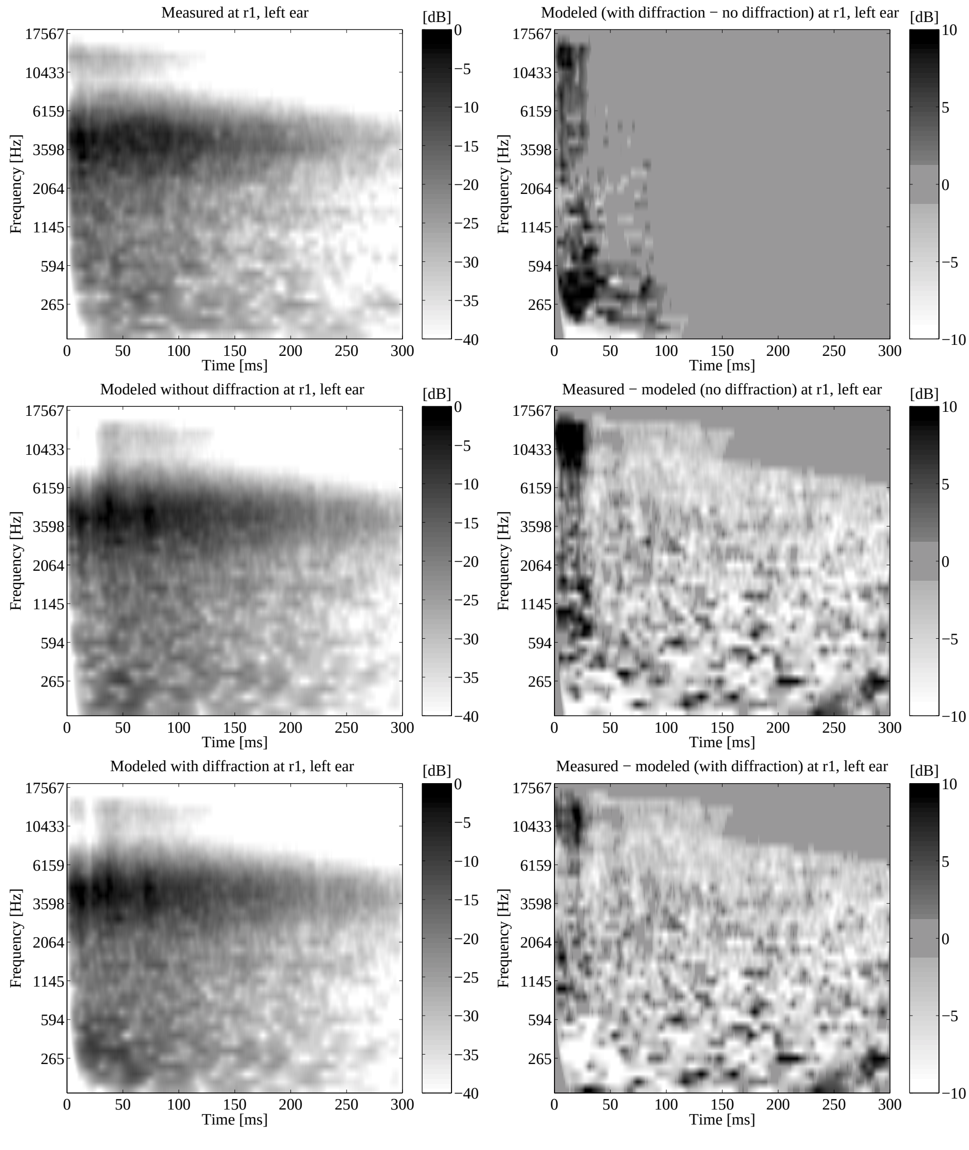


Figure 4.7: Visualization, motivated by auditory perception, of the impulse responses in the left column of Figure 4.4. Before calculating the difference plots the data values less than -40 dB of the maximum values were cut to highlight the most audible differences.

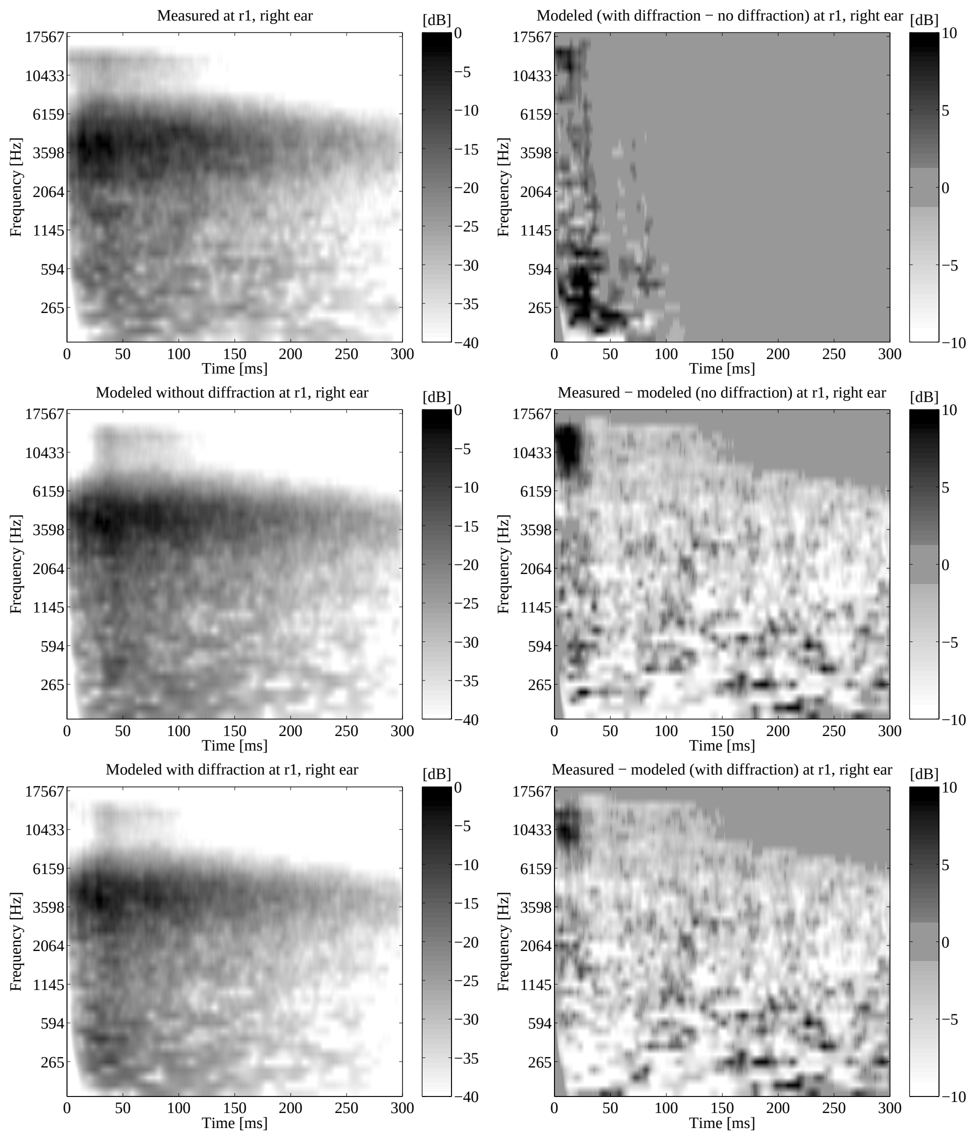


Figure 4.8: Visualization, motivated by auditory perception, of the impulse responses in the right column of Figure 4.4. Before calculating the difference plots the data values less than -40 dB of the maximum values were cut to highlight the most audible differences.

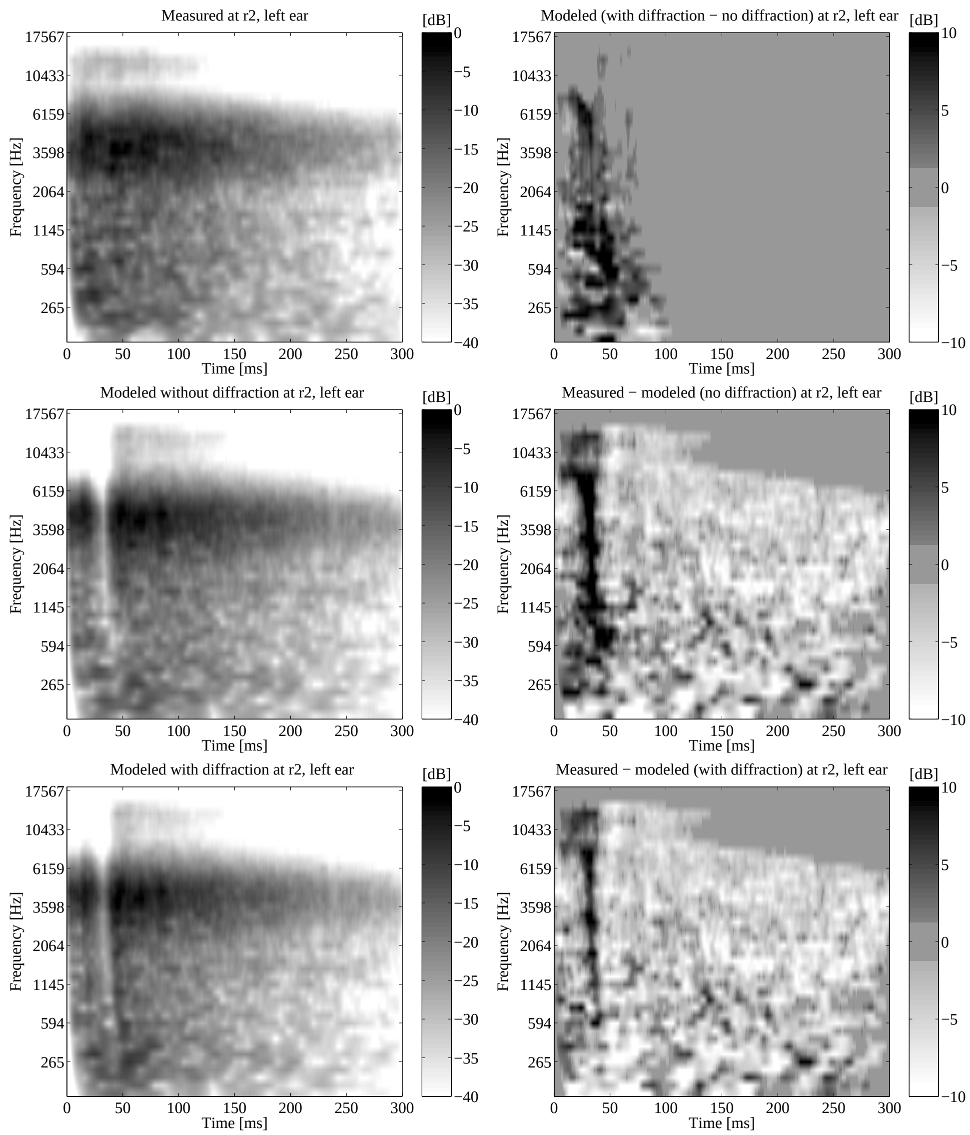


Figure 4.9: Visualization, motivated by auditory perception, of the impulse responses in the left column of Figure 4.5. Before calculating the difference plots the data values less than -40 dB of the maximum values were cut to highlight the most audible differences.

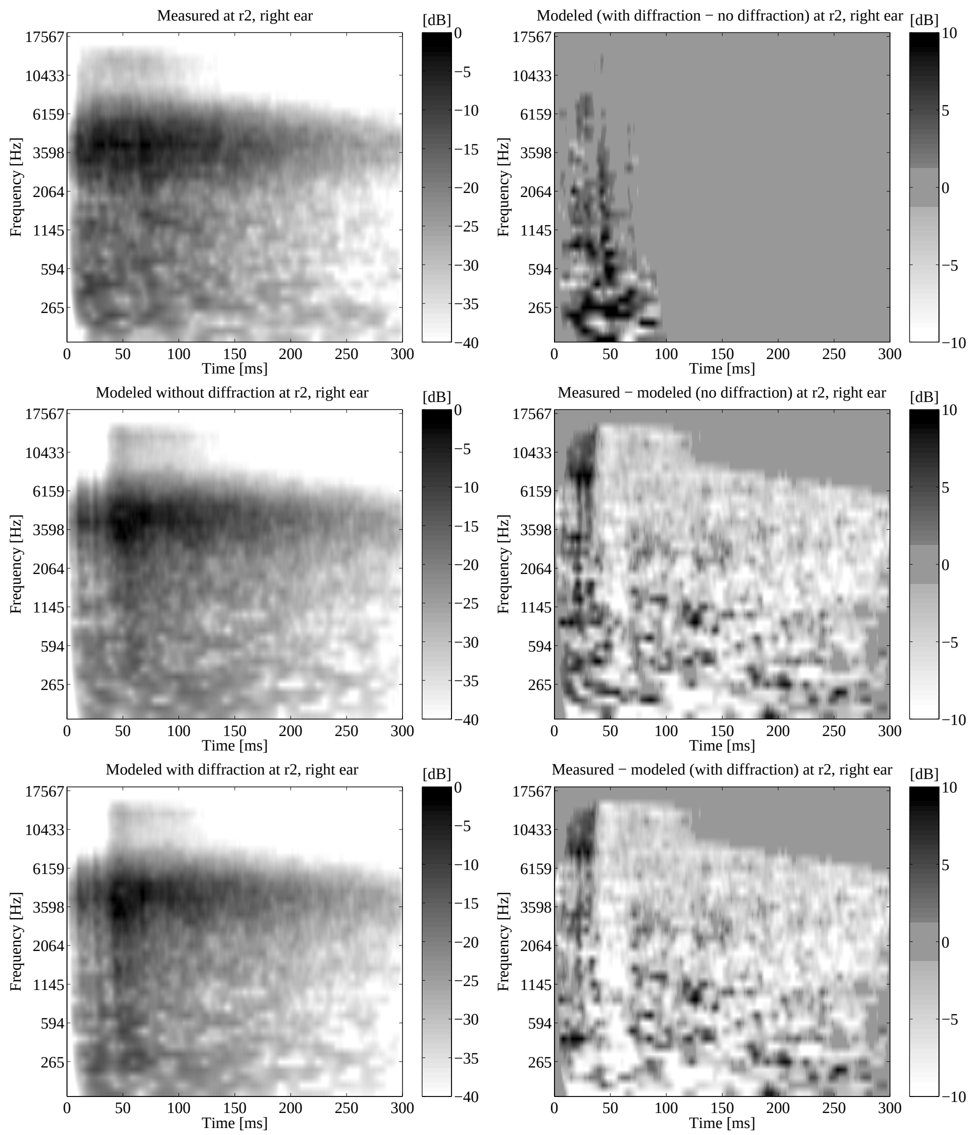


Figure 4.10: Visualization, motivated by auditory perception, of the impulse responses in the right column of Figure 4.5. Before calculating the difference plots the data values less than -40 dB of the maximum values were cut to highlight the most audible differences.

4.4 Results of Subjective Evaluation

The preliminary results of the last listening test have been presented earlier [P7], but here the final results with the data of all 20 subjects are presented. The data was analyzed with the SPSS V7.5 software. First, the correctness of data unrolling was assessed by tabulating the data by independent variables. Data was unrolled correctly, but one minor mistake in the test procedure was detected. Thirteen subjects out of twenty had judged one task twice and the corresponding pair was judged only by seven subjects. However, this defect in test procedure did not affect noticeably the results.

ITU-R BS.1116-1 [68] recommends the use of difference grades in the analysis and they were calculated by subtracting the test signal grades from the hidden reference grades for both the spatial difference and the timbral difference. Positive difference grades show directly if a subject had not found the hidden reference. The different grades allowed as well the analysis of the reliability of subjects with one-sided t-test as recommended in the ITU-R BS.1116-1. These t-tests were computed and no unreliable subjects were detected. Actually, only four subjects had made one error when selecting the hidden reference. The recommendation also suggests that the statistical analysis of the data should be performed with the analysis of variance (ANOVA). However, the data of the test did not fulfill completely the ANOVA assumptions, e.g., homogeneity of variances. This might be due to several facts, but perhaps there were too few subjects who completed the tasks. Anyhow, the ANOVA could not be applied to present data and instead the non-parametric equivalent of one-way ANOVA, namely the Kruskal-Wallis test, has been applied to raw data to study statistical significance of differences in the following analysis.

The first hypothesis was that the binaural soundtracks differed noticeably at frequencies above 5 kHz. This was suspected because the recordings sounded unnaturally bright and the objective analysis (Figures 4.7-4.10) suggested that clearly audible differences at high frequencies might occur. The hypothesis was studied as a function of bandwidth and recording method and the results are shown in Figure 4.11. It is clearly seen that binaural full bandwidth cases had gained lower grades than lowpass filtered binaural and full bandwidth monophonic cases. The differences were also statistically significant (Kruskal-Wallis test gave for DIFFSPAT $\chi^2 = 30.929, p = 0.000$ and for DIFFTIMB $\chi^2 = 92.336, p = 0.000$).

The results in Figure 4.11 indicate that the binaural recordings had some recording artifacts above 5 kHz. Recently, the utilized microphones and the pre-

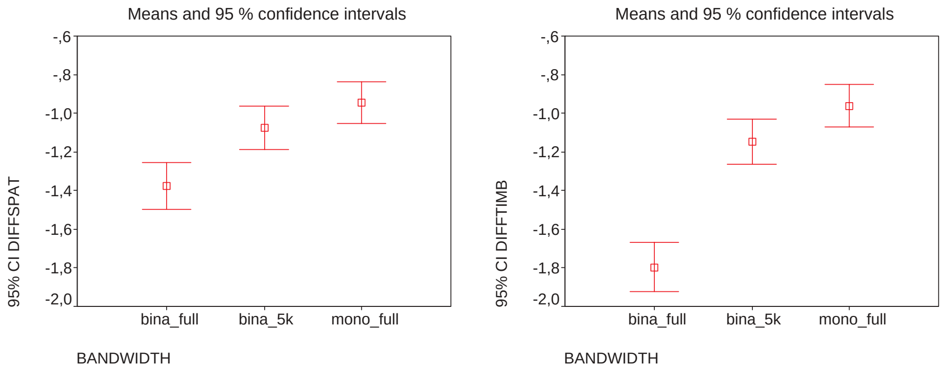


Figure 4.11: The results of the listening test as a function of bandwidth and recording method. DIFFSPAT is the difference grade for spatial differences and DIFFTIMB for timbral differences correspondingly.

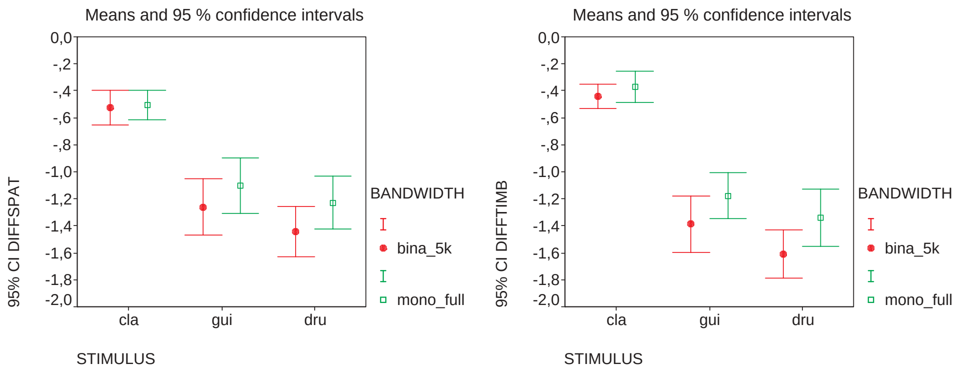


Figure 4.12: The results of the listening test as a function of stimulus. DIFFSPAT is the difference grade for spatial differences and DIFFTIMB for timbral differences correspondingly.

amplifier have been found unreliable in other recording sessions and might have affected the recordings for this work. Another possible reason is the critical microphone placement when performing the binaural recordings (or HRTF measurements) [138]. Nevertheless, it can be summarized that these recording artifacts prevented the subjects from quantifying the auralization quality properly. Thus, all binaural full bandwidth results were unreliable and they have been excluded for the rest of the analysis.

The second hypothesis was that auralized soundtracks were so good that they could hardly be distinguished from the recorded ones. This issue was studied as a function of stimulus and the results are depicted in Figure 4.12. The auralizations with the clarinet stimulus had been graded so well that the mean of difference grade was almost zero corresponding to “imperceptible” both for spatial and timbral differences. The results for more critical stimuli (guitar and snare drum) can be considered to be fairly good, because the applied double-blind triple-stimulus hidden-reference test detects all possible, even negligible, differences. However, while the means of both guitar and drum were almost “perceptible, but annoying” the variances were quite large. The statistical significance was investigated with Kruskal-Wallis test and the results were for binaural DIFFSPAT $\chi^2 = 38.805, p = 0.000$ and DIFFTIMB $\chi^2 = 81.805, p = 0.000$, and for monophonic DIFFSPAT $\chi^2 = 58.303, p = 0.000$ and DIFFTIMB $\chi^2 = 88.187, p = 0.000$.

The last interesting question was whether the diffraction modeling was needed in this geometry as there were no occluders between the sound source and the receiver positions. The objective evaluation suggested that diffraction modeling was important as it increased the reflection density and in this way complemented the image-source method [165, 170]. However, the results of the listening test in Figure 4.13 suggest that diffraction modeling degraded the auralization quality in the receiver position r1 but enhanced the quality in the receiver position r2, when spatial properties were judged. Unfortunately, with non-parametric statistical tests the statistical significance of this possible interaction could not be explored.

4.5 Discussion

The auralization quality is multidimensional in nature and its evaluation is laborious since no commonly accepted evaluation methods or quality metrics exist. Descriptive language for spatial sound characteristics have recently been proposed by Berg [16] and by Zacharov [193]. Such studies will help in the definition of the optimal quality. However, good assessment methods are still to be explored.

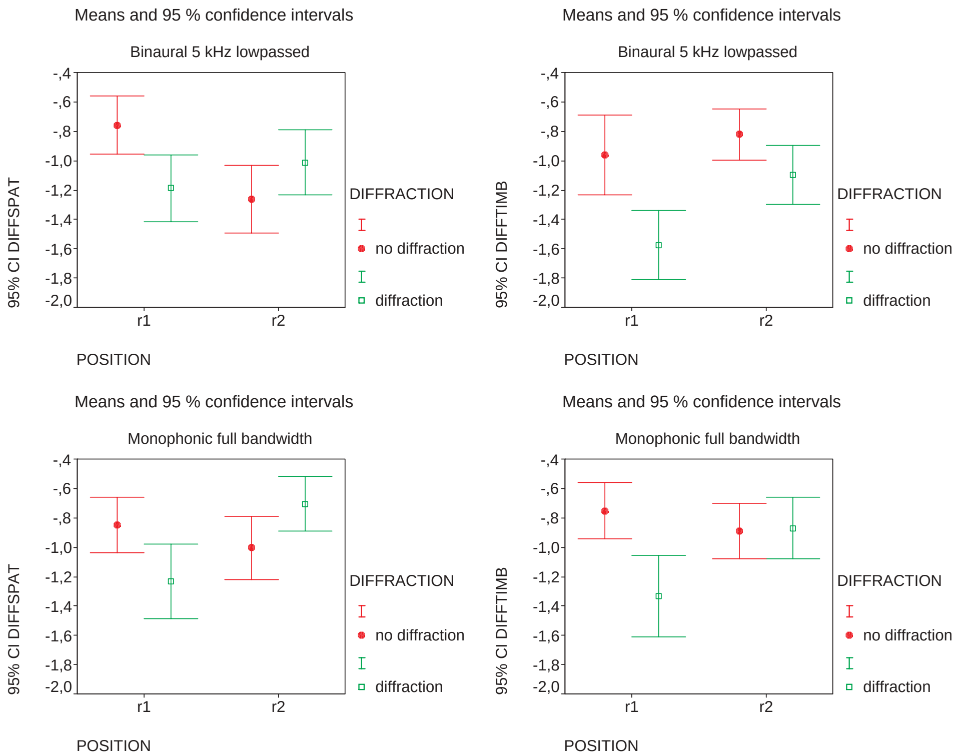


Figure 4.13: The results of the listening test as a function of both diffraction modeling and receiver position. DIFFSPAT is the difference grade for spatial differences and DIFFTIMB for timbral differences correspondingly. Top: The results with binaural lowpassed soundtracks. Bottom: The results with monophonic full bandwidth soundtracks.

	Pearson correlation	2-tailed sig.
Binaural 5 kHz lowpassed	0.498	0.000
Monophonic full bandwidth	0.581	0.000

Table 4.2: Pearson correlation between spatial and timbral differences.

In this thesis the comparison were made between the auralizations and real recordings of a space employing both monophonic and binaural recording techniques. The overall result of the evaluation was very good since only slight differences between the recordings and the auralizations were reported, even with snare drum hits being wideband transient-like signals. With the applied listening test method all the smallest possible differences could be found and in this context the result can be considered excellent. Some of the subjects even claimed that the auralizations sounded more natural than the recordings.

The spatial properties of the auralizations were little bit better graded than timbral characteristics, although the difference between these two attributes was small. The correlation between these two dependent variables was tested with Pearson correlation for all the data and the results are to be found in Table 4.2. The moderate correlations indicate that subjects were able to grade spatial and timbral differences separately. However, many subjects reported that it was hard to grade spatial and timbral properties independently. Perhaps, such uncertainty was due to the lack of proper training session before the test.

The diffraction modeling did not enhance the auralization quality significantly. This can be understood since the direct sound was in all rendering cases visible. Actually, in some rendering cases diffraction modeling even slightly degraded the quality. This might be due to some modeling errors, although the proposed filter design method is robust and quite accurate. Based on the presented evaluation results it can be concluded that diffraction modeling, being quite laborious, is not needed in geometries where no big obstacles causing significant diffraction components are present.

Although natural sounding auralization was realized with the presented techniques, the ultimate goal "authentic auralization" was not achieved in all aspects. Strictly speaking, according to Blauert [17, 128] the authenticity means an exact copy of the binaural signals. Naturally, such a copy is impossible to realize in practice. For example, in this case the starting point for auralization was not perfect since the 3D model was not an exact copy of the real room, as illustrated in Figure 4.14. However, plausible, perceptually authentic, auralizations were almost achieved since the auralizations were reported to sound very natural and close to the recorded sounds.

4.6 Future Directions

While the quality of auralizations was reported to be good it should still be improved. One essential issue, which was totally neglected in this thesis, is the modeling of diffuse reflections. Their inclusion in room acoustic modeling methods has been shown to be important [29, 30, 162, 183]. The next step in further development of the DIVA auralization system should be the implementation of diffuse reflections.

Another important enhancement to the presented auralization system can be obtained by applying more accurate late reverberation model. Current algorithm produces diffuse late reverberation and it can handle closed rooms with a diffuse sound field. However, it does not work properly in some cases, e.g., with flutter echos and non-exponential decays. Other minor ameliorations to the presented

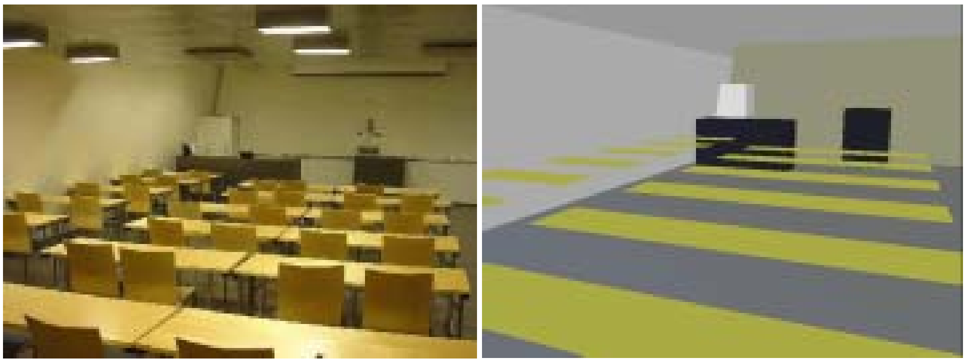


Figure 4.14: A photograph and the 3D model applied in auralization of the studied lecture room.

auralization can be obtained by adjusting signal processing parameters, for example, the level of late reverberation. However, such subtle improvements might not be worth implementation since the quality has been evaluated only with one geometry. In the future, more complex geometries containing larger variation of surfaces materials, should be considered to really find out if the presented auralization method can provide natural sounding and plausible auralizations with any 3D models.

For real-time applications the signal processing applied should be computationally optimized without degrading the quality. Such optimization requires a vast amount of subjective testing since the perceptual importance of all designed filters and the number of secondary sources should be found out. In addition, efficient algorithms and elegant implementation is required. Such optimization requires high-level programming skills and understanding of human perception of spatial sound which is a complex and multidimensional phenomenon.

The quality of the recorded binaural reference signals was not perfect. Indeed, to make high-quality binaural recordings across the whole audible frequency range was found to be very hard. In the future, when performing similar comparisons as presented in this thesis, recordings have to be made with a great deal of accuracy. Such problems as critical microphone placement in the entrance of ear canals [138] and individual differences between the shapes of human head and pinnae [157] have to be solved. In addition, other methods to capture spatial sound can be considered, although the real-head recording technique is the best practical way to capture spatial information also with a moving listener.

5 CONCLUSIONS

The work presented in this thesis can be divided into the design and implementation of a physically-based auralization system, and the evaluation of quality of the realized system with case studies.

The design and implementation of the auralization system which renders a 3D model of a space audible have been presented in publications [P1-P3]. The novelties in these publications include:

- Procedures for incorporation such acoustic phenomena as sound source directivity, air and material absorption, and the characteristics of spatial hearing into dynamic rendering. The corresponding digital filters were fitted in a robust and automatic manner to analytical or measured responses by applying several perceptually optimized filter design techniques.

- The implementation of diffraction modeling to the presented auralization system in a computationally efficient way.

- The realization of dynamic sound rendering employing various interpolation techniques.

A novel objective analysis method for analyzing room acoustics, motivated by auditory perception, has been presented in publication [P4]. The method enables visualization of room impulse responses utilizing similar time and frequency resolution as human hearing. By using directional microphones, the directional characteristics of a decaying sound field can be studied with described method.

The evaluation of the presented auralization system has been reported in publications [P5-P7] including the following novelties:

- The analysis method, motivated by auditory perception, was utilized in the auralization quality evaluation.

- The quality of auralization has been validated with listening tests and it has been demonstrated that the auralizations could hardly be distinguished from the recordings made in the studied lecture room.

6 SUMMARY OF PUBLICATIONS AND AUTHOR'S CONTRIBUTION

This chapter summarizes the publications incorporated in this thesis and describes the contributions of the author. It should be pointed out that many of the ideas and some of the work presented was performed in cooperation with other researchers. All ideas behind the DIVA auralization system are the results of collaboration with Prof. Lauri Savioja, Dr. Jyri Huopaniemi, Prof. Tapio Takala, Mr. Jarmo Hiipakka, Ms. Riitta Väänänen, and Dr. Ville Pulkki. The author has implemented a major part of the auralization system described in this thesis. The idea of an efficient auralization of edge diffraction is a result of cooperation with Prof. Peter Svensson. Ms. Hanna Järveläinen has helped the author with the listening test design and in statistical analysis. The original idea of applying auditory modeling features in analysis of impulse responses was suggested by Prof. Matti Karjalainen, but the author developed it further. Recently, the author has shared ideas with Dr. Ville Pulkki on the evaluation and quality issues of the auralization.

The author of this thesis is the primary author of all the publications [P1]-[P7], with the exception of [P1]. Publication [P1] has previously formed a part of one thesis. The scientific contribution of the articles is as follows.

Publication [P1]
This article discusses the theory and techniques for room acoustic modeling and rendering. It also describes an auralization system that has been implemented. The proposed auralization structure is flexible, it can be used for both static and dynamic as well as for real-time and non-real-time auralizations. Novelties of the system include, e.g., time-variant processing of image sources. The auralization system has been applied to make a state-of-the-art demonstration video of the concert hall Marienkirche.

In this article the author has implemented the DIVA auralization software with Prof. Savioja and the author has written Section 6 and one half of Sections 4 and 5.

Publication [P2]
A novel framework for evaluating both static and dynamic virtual acoustic environments is introduced. The framework is based on the comparison of real-head recordings and auralizations. In addition, the signal processing structure for auralization is discussed in more detail than in article [P1] and a novel implementation of sound source directivity filters is presented.

The author has performed all filter designs and simulations presented in this publication and he has written 90 % of the text.

Publication [P3]
A new efficient way to include edge diffraction to parametric auralization is introduced. In the presented solution, diffraction is modeled with a low-order warped IIR filters fitted to the analytically calculated responses. Thus, an efficient implementation of the diffraction is proposed.

In this publication the author performed all simulations and filter designs and written 90 % of the text.

Publication [P4]
A novel objective analysis method for room responses is proposed. The method is motivated by auditory perception and it utilizes time and frequency resolution

similar to the human hearing. The analysis method includes the use of directional microphones yielding cues about the diffuseness and the directional characteristics of sound fields in the time-frequency domain. This approach is particularly interesting in the visualization of concert hall acoustics. It is also applicable in the analysis of artificial reverberation and related audio effects. In this thesis it has been used to study the quality of auralization.

The author has chosen the suitable models of human hearing for the method and he has accomplished all simulations. The author has written 90 % of the text of the article.

Publication [P5]

Objective evaluation of the measured and modeled binaural room impulse responses is presented, calculating traditional room acoustic criteria. The auditory motivated analysis is applied as a novel tool in comparison of impulse responses.

The author is the sole author of this publication.

Publication [P6]

First subjective evaluation of both static and dynamic auralizations is reported. The listening test was completed by twelve subjects whose assignment was to judge the different aspects of the quality between real-head recordings and auralizations. The first results indicate that minor modeling errors still occur. The results of the second test propose that the comparison is harder when subjects were aware of rendering cases.

The author has prepared all soundtracks utilized in the conducted listening tests and written 90 % of the text. The listening test design and statistical analysis has been performed in collaboration with Ms. Hanna Järveläinen.

Publication [P7]

The quality of auralization including edge diffraction, is evaluated with objective and subjective criteria. The evaluation results show that with sound of a clarinet the auralization is almost identical with the one obtained by recording the anechoic signal in a real space. However, with transient-like signals the auralization slightly differs from the recorded ones. Despite, the evaluation results proved that plausible and almost authentic natural sounding binaural auralization is possible with the presented auralization methods in rooms similar to the studied one.

The author has prepared all the soundtracks utilized in the listening tests and performed the analysis of the results. In addition, the objective evaluation is completed by the author and he has written 95 % of the text.

[1] ISO/IEC International Standard (IS) 14496. Information technology - coding of audio-visual objects (MPEG-4), 2001.

[2] ISO Standard 9613-1. *Acoustics — Attenuation of Sound During Propagation Outdoors — Part 1: Calculation of the Absorption of Sound by the Atmosphere*. 1993.

[3] W. Ahnert and R. Feistel. EARS auralization software. *J. Audio Eng. Soc.*, 41(11):894–904, Nov. 1993.

[4] W. Ahnert and R. Feistel. Comparison of auralized impulse responses acquires by both measurement and simulation. In *Proc. 15th Int. Congr. Acoust. (ICA'95)*, volume 1, pages 353–356, Trondheim, Norway, June 1995.

[5] W. Ahnert and R. Feistel. Use of different methods to calculate scattering effects. In *Proc. 15th Int. Congr. Acoust. (ICA'95)*, volume 1, pages 357–360, Trondheim, Norway, June 1995.

[6] W. Ahnert and R. Feistel. Real time auralisation with EASE 3.0. In *Proc. of the Joint 137th Meeting of ASA and 2nd Convention of EAA (Forum Acusticum)*, March 14-19 1999. paper no. 3pAAa4.

[7] J. B. Allen and D. A. Berkley. Image method for efficiently simulating small-room acoustics. *J. Acoust. Soc. Am.*, 65(4):943–950, 1979.

[8] P. Astheimer. What you see is what you hear – acoustics applied in virtual worlds. In *Proc. IEEE 1993 Symp. on Research Frontiers in Virtual Reality*, pages 100–107, San Jose, CA, 1993.

[9] P. Astheimer. Acoustic simulation for visualization and virtual reality. In R. Veltkamp, editor, *Eurographics State-of-the-Art reports*, pages 1–23, Maastricht, Netherlands, Sept. 1995.

[10] M. Barron. *Auditorium Acoustics and Architectural Design*. E & FN Spon, 1993.

[11] S. Bech. Spatial aspects of reproduced sound in small rooms. *J. Acoust. Soc. Am.*, 103(1):434–445, Jan. 1998.

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

[13] D.R. Begault. Virtual acoustics: evaluation of psychoacoustic parameters for auralization. In *Proceedings of IWHIT'97, Third International Workshop on Human Interface Technology*, Aizu-Wakamatsu, Japan, Mar. 12-14 1997.

[14] D.R. Begault, E.M. Wenzel, and M.R. Anderson. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *J. Audio Eng. Soc.*, 49(10):904–916, 2001.

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

[16] J. Berg and F. Rumsay. In search of the spatial dimensions of reproduced sound: Verbal protocol analysis and cluster analysis of scaled verbal descriptors. In *the 108th Audio Engineering Society (AES) Convention*, Paris, France, Feb. 19-22 2000. preprint no. 5129.

[17] J. Blauert. *Spatial Hearing. The psychophysics of human sound localization*. MIT Press, Cambridge, MA, 2nd edition, 1997.

[18] J. Blauert and H. Lehnert. Binaural technology & virtual reality. In *Proc. 2nd Int. Conf. on Acoustics and Musical Research (CIARM'95)*, pages 3–10, Ferrara, Italy, May 19-21 1995.

[19] J. Blauert, H. Lehnert, J. Sahrhage, and H. Strauss. An interactive virtual-environment generator for psychoacoustic research. I: Architecture and implementation. *ACUSTICA united with acta acustica*, 86(1):94–102, 2000.

[20] J. Borish. Electronic simulation of auditorium acoustics. Technical Report STAN-M-18, Stanford University, Dept. of Music, CCRMA, Stanford, California, May 1984.

[21] J. Borish. Extension of the image model to arbitrary polyhedra. *J. Acoust. Soc. Am.*, 75(6):1827–1836, 1984.

[22] M.A. Casey, W.G. Gardner, and S. Basu. Vision steered beam-forming and transaural rendering for the artificial life interactive video environment (ALIVE). In *the 99th the Audio Engineering Society (AES) Convention*, New York, USA, 1995. preprint no. 4052.

[23] F. Cruz-Barney and O. Warusfel. Prediction of the spatial informations for the control of room acoustics auralization. In *the 103rd Audio Engineering Society (AES) Convention*, New York, Sept. 26-29 1997. preprint no. 4628.

[24] B.-I. Dalenbäck. *A New Model for Room Acoustic Prediction and Auralization*. PhD thesis, Chalmers University of Technology, Gothenburg, Sweden, 1995.

[25] B.-I. Dalenbäck. Room acoustic prediction and auralization based on a unified treatment of diffuse and specular reflection. In *Proc. 15th Int. Congr. Acoust. (ICA'95)*, volume 1, pages 425–428, Trondheim, Norway, June 1995.

[26] B.-I. Dalenbäck. Room acoustic prediction based on a unified treatment of diffuse and specular reflection. *J. Acoust. Soc. Am.*, 100(2):899–909, Aug. 1996.

[27] B.-I. Dalenbäck. Verification of prediction based on randomized tail-corrected cone-tracing and array modeling. In *the CDROM of the ASA/EAA/DEGA Joint Meeting*, Berlin, Germany, March 14-19 1999. paper n. 3pAAa1.

[28] B.-I. Dalenbäck, M. Kleiner, and U.P. Svensson. Audibility of changes in geometric shape, source directivity, and absorptive treatment – experiments in auralization. *J. Audio Eng. Soc.*, 41(11):905–913, Nov. 1993.

[29] B.-I. Dalenbäck, M. Kleiner, and U.P. Svensson. A macroscopic view of diffuse reflection. *J. Audio Eng. Soc.*, 42(10):793–806, 1994.

[30] B.-I. Dalenbäck, M. Kleiner, and U.P. Svensson. Auralization, virtually everywhere. In *the 100th Audio Engineering Society (AES) Convention*, Copenhagen, Denmark, May 11-14 1996. preprint no. 4228.

[31] B.-I. Dalenbäck and D. McGrath. Narrowing the gap between virtual reality and auralization. In *Proc. 15th Int. Congr. Acoust. (ICA'95)*, volume 2, pages 429–432, Trondheim, Norway, June 1995.

[32] B.-I. Dalenbäck, U.P. Svensson, and M. Kleiner. Prediction and auralization based on a combined image source/ray-model. In *Proc. 14th Int. Congr. Acoust. (ICA'92)*, Beijing, China, July 1992. paper F2-7.

[33] T. Djelani, C. Pörchmann, J. Sahrhage, and J. Blauert. An interactive virtual-environment generator for psychoacoustic research II: Collection of head-related impulse responses and evaluation of auditory localization. *ACUSTICA united with acta acustica*, 86(6):1046–1053, 2000.

[34] N.I. Durlach, A. Rigopulos, X.D. Pang, W.S. Woods, A. Kulkarni, H.S. Colburn, and E.M. Wenzel. On the externalization of auditory images. *Presence: Teleoperators and Virtual Environments*, 1(2):251–257, 1992.

[35] G. Eckel. Applications of the cyberstage sound server. In *Proc. AES 16th Int. Conf. on Spatial Sound Reproduction*, pages 478–484, Rovaniemi, Finland, April 10-12 1999.

[36] A. Farina. Auralization software for the evaluation of the results obtained by a pyramid tracing code: Results of subjective listening tests. In *Proc. 15th Int. Congr. Acoust. (ICA'95)*, pages 441–444, Trondheim, Norway, June 1995.

[37] A. Farina. Verification of the accuracy of the pyramid tracing algorithm by comparison with experimental measurements of objective acoustic parameters. In *Proc. 15th Int. Congr. Acoust. (ICA'95)*, pages 445–448, Trondheim, Norway, June 1995.

[38] A. Farina. Introducing the surface diffusion and edge scattering in a pyramid-tracing numerical model for room acoustics. In *the 108th Audio Engineering Society (AES) Convention*, Paris, France, Feb. 19-22 2000. preprint no. 5145.

[39] A. Farina and E. Ugolotti. Subjective comparison of different car audio systems by the auralization technique. In *the 103rd Audio Engineering Society (AES) Convention*, New York, Sept. 26-29 1997. preprint no. 4587.

[40] A. Farina and E. Ugolotti. Numerical model of the sound field inside cars for the creation of virtual audible reconstructions. In *Proc. Digital Audio Effects Workshop (DAFX'98)*, pages 57–61, Barcelona, Spain, Nov. 19-21 1998.

[41] S. Foster, E.M. 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.

[42] T.A. Funkhouser, I. Carlbom, G. Elko, G. Pingali, M. Sondhi, and J. West. A beam tracing approach to acoustic modeling for interactive virtual environments. In *SIGGRAPH 98 Conference Proceedings*, pages 21–32. ACM SIGGRAPH, July 1998.

[43] T.A. Funkhouser, P. Min, and I. Carlbom. Real-time acoustic modeling for distributed virtual environments. In *SIGGRAPH 99 Conference Proceedings*. ACM SIGGRAPH, July 1999.

[44] W.G. Gardner. Virtual acoustic room. Master's thesis, MIT, 1992.

[45] W.G. Gardner. Efficient convolution without input-output delay. *J. Audio Eng. Soc.*, 43(3):127–136, 1995.

[46] W.G. Gardner. *3-D Audio Using Loudspeakers*. PhD thesis, MIT Media Lab, Sept. 1997.

[47] W.G. Gardner. Reverberation algorithms. In M. Kahrs and K. Brandenburg, editors, *Applications of Digital Signal Processing to Audio and Acoustics*, pages 85–131. Kluwer Academic Publishers, 1997.

[48] W.G. Gardner and D. Griesinger. Reverberation level matching experiments. In *Proc. of the Wallace C. Sabine Centennial Symposium*, pages 263–266, June 5-7 1994.

[49] W.G. Gardner and K. Martin. HRTF measurements of a KEMAR dummy-head microphone. Technical Report 280, MIT Media Lab Perceptual Computing, 1994.

[50] M. Gerzon. Periphony: With-height sound reproduction. *J. Audio Eng. Soc.*, 21(1/2):2–10, 1973.

[51] D. Griesinger. Practical processors and programs for digital reverberation. In *Proc. AES 7th Int. Conf.*, pages 187–195, Toronto, Ont., Canada, May 1989.

[52] 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. *Applied Acoustics*, 38(2-4, Special Issue on Computer Modelling and Auralisation of Sound Fields in Rooms):145–159, 1993.

[53] 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 Proceedings of SIGGRAPH'97*, page 81, Los Angeles, 1997. ACM SIGGRAPH.

[54] P.A. Hill, P.A. Nelson, and O. Kirkeby. Resolution of front-back confusion in virtual acoustic imaging system. *J. Acoust. Soc. Am.*, 108(6):2901–2910, Dec. 2000.

[55] U. Horbach, A. Karamustafaoglu, E. Corteel, and R. Pellegrini. Implementation of an auralization scheme in a digital mixing console using perceptual parameters. In *the 108th Audio Engineering Society (AES) Convention*, Paris, France, Feb. 19-22 2000. preprint no. 5099.

[56] U. Horbach, A. Karamustafaoglu, R. Pellegrini, P. Mackensen, and G. Theile. Design and applications of a data-based auralization system for surround sound. In *the 106th Audio Engineering Society (AES) Convention*, Munich, Germany, May 8-11 1999. preprint no. 4976.

[57] J. Huopaniemi. *Virtual acoustics and 3-D sound in multimedia signal processing*. PhD thesis, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, report 53, 1999.

[58] J. Huopaniemi and M. Karjalainen. HRTF filter design based on auditory criteria. In *Proc. Nordic Acoustical Meeting (NAM'96)*, pages 323–330, Helsinki, Finland, June 12-14 1996.

[59] J. Huopaniemi and M. Karjalainen. Review of digital filter design and implementation methods for 3-D sound. In *the 102nd Audio Engineering Society (AES) Convention*, Munich, Germany, Mar. 22-25 1997. preprint no. 4461.

[60] 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)*, pages 183–186, Mohonk Mountain House, New Paltz, New York, Oct. 17-20 1999.

[61] 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)*, Mohonk, New Paltz, New York, Oct. 19-22 1997.

[62] J. Huopaniemi, L. Savioja, T. Lokki, and R. Väänänen. Virtual acoustics - applications and technology trends. In *Proc. of the 10th European Signal Processing Conference (Eusipco2000)*, volume 4, pages 2201–2208, Tampere, Finland, Sept. 5-8 2000.

[63] J. Huopaniemi, L. Savioja, and T. Takala. DIVA virtual audio reality system. In *Proc. Int. Conf. Auditory Display (ICAD'96)*, pages pp. 111–116, Palo Alto, California, Nov. 8-11 1996.

[64] J. Hynninen. A software-based system for listening tests. Master's thesis, Helsinki University of Technology, Faculty of Electrical Engineering, Laboratory of Acoustics and Audio Signal Processing, 2001. 90 p.

[65] J. Hynninen and N. Zacharov. Guineapig - a generic subjective test system for multichannel audio. In *the 106th Audio Engineering Society (AES) Convention*, Munich, Germany, May 8-11 1999. preprint no. 4871.

[66] ISO Standard 226. *Acoustics – Normal equal-loudness level contours*. International Standards Organization, 1987.

[67] ISO Standard 3382. *Acoustics – Measurement of the reverberation time of rooms with reference to other acoustical parameters*. International Standards Organization, 1997.

[68] ITU-R. Recommendation BS.1116-1, Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. International Telecommunication Union Radiocomminication Assembly, 1997.

[69] J.-M. Jot. An analysis/synthesis approach to real-time artifical reverberation. In *Proc. 1992 IEEE Int. Conf. Acoust., Speech, and Signal Processing*, volume 2, pages 221–224, San Francisco, CA, Mar. 1992.

[70] J.-M. Jot. *Etude et réalisation d'un spatialisateur de sons par modèles physique et perceptifs*. PhD thesis, l'Ecole Nationale Superieure des Telecommunications, Télécom Paris 92 E 019, Sept. 1992.

[71] J.-M. Jot. Synthesizing three-dimensional sound scenes in audio or multimedia production and interactive human-computer interfaces. In *Proc. 5th Int. Conf. Interface to real & virtual worlds*, Montpellier, France, 1996.

[72] J.-M. Jot. Efficient models for reverberation and distance rendering in computer music and virtual audio reality. In *Proc. Int. Computer Music Conf. (ICMC'97)*, pages 236–243, Thessaloniki, Greece, Sept. 1997.

[73] J.-M. Jot. Real-time spatial processing of sounds for music, multimedia and interactive human-computer interfaces. *Multimedia Systems, Special Issue on Audio and Multimedia*, 7(1):55–69, 1999.

[74] J.-M. Jot, L. Cerveau, and O. Warusfel. Analysis and synthesis of room reverberation based on a statistical time-frequency model. In *the 103rd Audio Engineering Society (AES) Convention*, New York, Sept. 26-29 1997. preprint no. 4629.

[75] J.-M. Jot and A. Chaigne. Digital delay networks for designing artificial reverberators. In *The 90th Audio Engineering Society (AES) Convention*, Paris, France, 1991. preprint no. 3030.

[76] J.-M. Jot, V. Larcher, and O. Warusfel. Digital signal processing issues in the context of binaural and transaural stereophony. In *the 98th Audio Engineering Society (AES) Convention*, Paris, France, 1995. preprint no. 3980.

[77] J.-M. Jot and O. Warusfel. Le spatialisateur. In *Le son & l'espace. Rencontres Musicales Pluridisciplinaires Informatique et Musique. GRAME'95*, Lyon, France, Mar. 31 - Apr. 1 1995.

[78] J.-M. Jot and O. Warusfel. A real-time spatial sound processor for music and virtual reality applications. In *Proc. Int. Computer Music Conf. (ICMC'95)*, Banff, Alta, Canada, Sept. 1995.

[79] J.-M. Jot, O. Warusfel, E. Kahle, and M. Mein. Binaural concert hall simulation in real time. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'93)*, Mohonk, New Paltz, New York, Oct. 17-20 1993.

[80] Y. Kahana, P.A. Nelson, O. Kirkeby, and H. Hamada. Objective and subjective assessment of systems for the production of virtual acoustic images for multiple listeners. In *the 103rd Audio Engineering Society (AES) Convention*, New York, NY, Sept. 26-29 1997. preprint no. 4573.

[81] Y. Kahana, P.A. Nelson, O. Kirkeby, and H. Hamada. A multiple microphone recording technique for the generation of virtual acoustic images. *J. Acoust. Soc. Am.*, 105(5):1503–1516, Mar. 1999.

[82] M. Karjalainen, J. Huopaniemi, and V. Välimäki. Direction-dependent physical modeling of musical instruments. In *Proc. 15th Int. Congr. Acoust. (ICA'95)*, pages 451–454, Trondheim, Norway, June 1995.

[83] O. Kirkeby, P.A. Nelson, F. Orduna-Bustamante, and H. Hamada. Local sound field reproduction using digital signal processing. *J. Acoust. Soc. Am.*, 100(3):1584–1593, Sept. 1996.

[84] O. Kirkeby, P. Rubak, L. G. Johansen, and P. A. Nelson. Imlementation of cross-talk cancellation networks using warped FIR filters. In *Proc. AES 16th Int. Conf. on Spatial Sound Reproduction*, pages 358–365, Rovaniemi, Finland, April 1999.

[85] 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.*, 91(3):1637–1647, Mar. 1992.

[86] M. Kleiner, B.-I. Dalenbäck, and U.P. Svensson. Auralization – an overview. *J. Audio Eng. Soc.*, 41(11):861–875, Nov. 1993.

[87] M. Kleiner, B.-I. Dalenbäck, and U.P. Svensson. Auralization: Acoustical rendering. In *Proc. 2nd Int. Conf. on Acoustics and Musical Research (CIARM'95)*, pages 23–30, Ferrara, Italy, May 19-21 1995.

[88] A. Krokstad, S. Strom, and S. Sorsdal. Calculating the acoustical room response by the use of a ray tracing technique. *J. Sound Vib.*, 8(1):118–125, 1968.

[89] H. Kuttruff. Digital simulation of concert hall acoustics and its applications. *Acoustics Bulletin*, pages 5–8, Sept./Oct. 1991.

[90] H. Kuttruff. *Room Acoustics*. Elsevier Science Publishers, Essex, England, 3rd edition, 1991.

[91] H. Kuttruff. Auralization of impulse responses modeled on the basis of ray-tracing results. *J. Audio Eng. Soc.*, 41(11):876–880, Nov. 1993.

[92] T. I. Laakso, V. Välimäki, M. Karjalainen, and U. K. Laine. Splitting the unit delay – tools for fractional delay filter design. *IEEE Signal Processing Magazine*, 13(1):30–60, Jan. 1996.

[93] V. Larcher, J.-M. Jot, J. Guyard, and O. Warusfel. Study and comparison of efficient methods for 3D audio spatialization based on linear decomposition on HRTF data. In *the 108th Audio Engineering Society (AES) Convention*, Paris, France, Feb. 19-22 2000. preprint no. 5097.

[94] V. Larcher, J.-M. Jot, and G. Vandernoot. Equalization methods in binaural technology. In *the 105th Audio Engineering Society (AES) Convention*, San Francisco, USA, Sept. 26-29 1998. preprint no. 4858.

[95] H. Lehnert and J. Blauert. Virtual auditory environment. In *Proc. 5th Conference on Advanced Robotics*, volume 1, pages 211–216, 1991.

[96] H. Lehnert and J. Blauert. Principles of binaural room simulation. *Applied Acoustics*, 36(3-4):259–291, 1992.

[97] T. Lokki, J. Hiipakka, R. Hänninen, T. Ilmonen, L. Savioja, and T. Takala. Realtime audiovisual rendering and contemporary audiovisual art. *Organised Sound*, 3(3):219–233, 1998.

[98] T. Lokki, L. Savioja, J. Huopaniemi, R. Väänänen, and T. Takala. Creating interactive virtual auditory environments. *IEEE Computer Graphics and Applications*, 22(4):2–10, July/Aug. 2002.

[99] C. Lynge, J.H. Rindel, and G. Naylor. Modelling sound distribution systems in a room acoustical computer model. In *Proc. 15th Int. Congr. Acoust. (ICA'95)*, volume 2, pages 557–560, Trondheim, Norway, June 1995.

[100] P. Mackensen, M. Fruhmann, M. Thanner, G. Thiele, U. Horbach, and A. Karamustafaoglu. Head-tracker based auralization systems: Additional consideration of vertical head movements. In *the 108th Audio Engineering Society (AES) Convention*, Paris, France, Feb. 19-22 2000. preprint no. 5135.

[101] P. Maijala. Better binaural recordings using the real human head. In *Proc. Int. Congr. Noise Control Engineering (Inter-Noise 1997)*, volume 2, pages 1135–1138, Budapest, Hungary, Aug. 25-27 1997.

[102] D. Malham and A. Myatt. 3-D sound spatialization using ambisonic techniques. *Computer Music J.*, 19(4):58–70, 1995.

[103] W. Martens. Principal components analysis and resynthesis of spectral cues to perceived direction. In *Proc. Int. Computer Music Conf. (ICMC'87)*, pages 274–281, 1987.

[104] 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.*, 94(6):3255–3264, Dec. 1993.

[105] J. Martin and J.-P. Vian. Binaural sound simulation of concert halls by a beam tracing method. In *Proc. 13th Int. Congr. Acoust. (ICA'89)*, Belgrade, Yugoslavia, July 1989.

[106] MATLAB. Signal processing toolbox, version 4.2, 1998. MathWorks Inc.

[107] D. McGrath. High resolution simulation of acoustic environments. In *Proc. IVR'94*, 1994.

[108] J.D. Miller. SLAB: a software-based real-time virtual acoustic environment rendering system. In *Proc. Int. Conf. Auditory Display (ICAD'2001)*, pages 279–280, Espoo, Finland, July 29 - Aug. 1 2001.

[109] P. Minnaar, S.K. Olesen, F Christensen, and H. Møller. Localization with binaural recordings from artificial and human heads. *J. Audio Eng. Soc.*, 49(5):323–336, 2001.

[110] H. Møller. Fundamentals of binaural technology. *Applied Acoustics*, 36(3-4):171–218, 1992.

[111] H. Møller. Interfacing room simulation programs and auralisation systems. *Applied Acoustics*, 38(2-4, Special Issue on Computer Modelling and Auralisation of Sound Fields In Rooms):333–347, 1993.

[112] H. Møller and D. Hammershøi. HRTFs for binaural synthesis. In *Proc. IEEE Nordic Signal Processing Symposium (NORSIG'98)*, pages 37–44, Vigsø, Denmark, 8-11 June 1998.

[113] H. Møller, D. Hammershøi, C.B. Jensen, and M.F. Sørensen. Evaluation of artificial heads in listening tests. *J. Audio Eng. Soc.*, 47(3):83–100, 1999.

[114] H. Møller, M.F. Sørensen D. Hammershøi, and C.B. Jensen. Head-related transfer functions of human subjects. *J. Audio Eng. Soc.*, 43(5):300–321, 1995.

[115] H. Møller, C.B. Jensen, D. Hammershøi, and M.F. Sørensen. Using a typical human subject for binaural recording. In *the 100th Audio Engineering Society (AES) Convention*, Copenhagen, Denmark, May 11-14 1996. preprint no. 4157.

[116] H. Møller, C.B. Jensen, D. Hammershøi, and M.F. Sørensen. Evaluation of artificial heads in listening tests. In *the 102nd Audio Engineering Society (AES) Convention*, Munich, Germany, Mar. 22-25 1997. preprint no. 4404.

[117] B.C.J. Moore and B.R. Glasberg. A revision of Zwicker's loudness model. *ACUSTICA united with acta acustica*, 82:335–345, 1996.

[118] B.C.J. Moore, R.W. Peters, and B.R. Glasberg. Auditory filter shapes at low center frequencies. *J. Acoust. Soc. Am.*, 88(1):132–140, July 1990.

[119] F. R. Moore. A general model for spatial processing of sounds. *Computer Music J.*, 7(3):6–15, 1983.

[120] J. A. Moorer. About this reverberation business. *Computer Music J.*, 3(2):13–28, 1979.

[121] G. Naylor. Treatment of early and late reflections in a hybrid computer model for room acoustics. In *the 124th ASA meeting*, New Orleans, USA, Nov. 1992. paper 3aAA2.

[122] G. Naylor. ODEON – another hybrid room acoustical model. *Applied Acoustics*, 38(2-4, Special Issue on Computer Modelling and Auralisation of Sound Fields in Rooms):131–143, 1993.

[123] P.A. Nelson, H. Hamada, and S.J. Elliott. Adaptive inverse filters for stereophonic sound reproduction. *IEEE Transactions on Signal Processing*, 40(7):1621 –1632, July 1992.

[124] P.A. Nelson, F. Orduna-Bustamante, D. Engler, and H. Hamada. Experiments on a system for the synthesis of virtual acoustic sources. *J. Audio Eng. Soc.*, 44:990–1007, May 1996.

[125] P.A. Nelson, F. Orduna-Bustamante, and H. Hamada. Inverse filter design and equalization zones in multichannel sound reproduction. *IEEE Transactions on Speech and Audio Processing*, 3(3):185 –192, May 1995.

[126] 3D Working Group of the Interactive Audio Special Interest Group. 3D audio rendering and evaluation guidelines. Technical report, MIDI Manufacturers Association, Los Angeles, CA, USA, June 1998. [Cited 6.6.2002] Available from World Wide Web: <http://www.iasig.org/>.

[127] 3D Working Group of the Interactive Audio Special Interest Group. Interactive 3D audio rendering guidelines, Level 2.0. Technical report, MIDI Manufacturers Association, Los Angeles, CA, USA, Sept. 1999. [Cited 6.6.2002] Available from World Wide Web: <http://www.iasig.org/>.

[128] R.S. Pellegrini. *A virtual reference listening room as an application of auditory virtual environments*. PhD thesis, Institut für Kommunikationsakustik, Ruhr-Universität, Bochum, Germany, 2001.

[129] A.D. Pierce. *Acoustics - An Introduction to Its Physical Principles and Applications*. The Acoustical Society of America, 2 edition, 1994.

[130] C.J. Plack and A.J. Oxenham. Basilar-membrane nonlinearity and the growth of forward masking. *J. Acoust. Soc. Am.*, 103(3):1598–1608, Mar. 1998.

[131] W. Pompetzki. *Psychoakustische Verifikation von Computermodellen zur binauralen Raumsimulation*. PhD thesis, Ruhr-Universität Bochum, Verlag Shaker, Aachen, 1993.

[132] W. Pompetzki and J. Blauert. A study on the perceptual authenticity of binaural room simulation. In *Proc. of the Wallace C. Sabine Centennial Symposium*, pages 81–84, June 5-7 1994.

[133] V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *J. Audio Eng. Soc.*, 45(6):456–466, June 1997.

[134] V. Pulkki. *Spatial sound generation and perception by amplitude panning techniques*. PhD thesis, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, report 62, 2001.

[135] V. Pulkki and T. Lokki. Creating auditory displays with multiple loudspeakers using VBAP: A case study with DIVA project. In *Proceedings of the International Confonference on Auditory Display (ICAD'98)*, Glasgow, UK, 1-4 November 1998.

[136] V. Pulkki, T. Lokki, and L. Savioja. Implementation and visualization of edge diffraction with image-source method. In *the 112nd Audio Engineering Society (AES) Convention*, Munich, Germany, May 10-13 2002. preprint no. 5603.

[137] A. Reilly and D. McGrath. Convolution processing for realistic reverberation. In *the 98th Audio Engineering Society (AES) Convention*, Paris, France, 1995. preprint no. 3977.

[138] K. Riederer. Repeatability analysis of HRTF measurements. In *the 105th Audio Engineering Society (AES) Convention*, San Francisco, USA, Sept. 26-29 1998. preprint no. 4846.

[139] J.H. Rindel. Computer simulation techniques for the acoustical design of rooms - how to treat reflections in sound field simulation. In *Proc. ASVA'97*, pages 201–208, Tokyo, Japan, Apr. 1997.

[140] J.H. Rindel. The use of computer modeling in room acoustics. *Journal of Vibroengineering*, 3(4):219–224, 2000.

[141] J.H. Rindel and C.L. Christensen. An auralization tool for studying the interaction between room and music. In *Proc. 2nd Int. Conf. on Acoustics and Musical Research (CIARM'95)*, pages 201–206, Ferrara, Italy, May 19-21 1995.

[142] J.H. Rindel, C. Lynge, G. Naylor, and K. Rishøj. The use of a digital audio mainframe for room acoustical auralization. In *the 96th Audio Engineering Society (AES) Convention*, Amsterdam, The Netherlands, 1994. preprint no. 3860.

[143] J.H. Rindel and G. Naylor. Computer modelling of sound fields in rooms – the state of the art and outlook to the future. In *Proc. 14th Int. Congr. Acoust. (ICA'92)*, Beijing, China, July 1992. paper F2-1.

[144] D. Rocchesso and J. O. Smith. Circulant and elliptic feedback delay networks for artificial reverberation. *IEEE Trans. Speech and Audio Processing*, 5(1):51–63, Jan. 1997.

[145] J. Sandvad. Dynamic aspects of auditory virtual environments. In *the 100th Audio Engineering Society (AES) Convention*, Copenhagen, Denmark, May 11-14 1996. preprint no. 4226.

[146] J. Sandvad and D. Hammershøi. Binaural auralization: Comparison of FIR and IIR filter representation of HIRs. In *the 96th Audio Engineering Society (AES) Convention*, Amsterdam, The Netherlands, 1994. preprint no. 3862.

[147] L. Savioja. *Modeling Techniques for Virtual Acoustics*. PhD thesis, Helsinki University of Technology, Telecommunications Software and Multimedia Laboratory, report TML-A3, 1999.

[148] L. Savioja, J. Backman, A. Järvinen, and T. Takala. Waveguide mesh method for low-frequency simulation of room acoustics. In *Proc. 15th Int. Congr. Acoust. (ICA'95)*, volume 2, pages 637–640, Trondheim, Norway, June 1995.

[149] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen. Virtual environment simulation - advances in the DIVA project. In *Proc. Int. Conf. Auditory Display (ICAD'97)*, pages 43–46, Palo Alto, California, Nov. 3-5 1997.

[150] L. Savioja, T. Rinne, and T. Takala. Simulation of room acoustics with a 3-D finite difference mesh. In *Proc. Int. Computer Music Conf. (ICMC'94)*, pages 463–466, Aarhus, Denmark, 12-17 Sept. 1994.

[151] L. Savioja and V. Välimäki. Reducing the dispersion error in the digital waveguide mesh using interpolation and frequency-warping techniques. *IEEE Transactions on Speech and Audio Processing*, 8(2):184–194, Mar. 2000.

[152] L. Savioja and V. Välimäki. Interpolated 3-D digital waveguide mesh with frequency warping. In *Proc. 2001 IEEE Int. Conf. Acoust. Speech. and Signal Process. (ICASSP'01)*, volume 5, pages 3345–3348, Salt Lake City, Utah, USA, 15-19 May 2001.

[153] E.D. Scheirer, R. Väänänen, and J. Huopaniemi. AudioBIFS: describing audio scenes with the MPEG-4 multimedia standard. *IEEE Transactions on Multimedia*, 1(3):237–250, Sept. 1999.

[154] M.R. Schroeder. Natural-sounding artificial reverberation. *J. Audio Eng. Soc.*, 10(3):219–223, 1962.

[155] M.R. Schroeder. Digital simulation of sound transmission in reverberant spaces. *J. Acoust. Soc. Am.*, 47(2 (Part 1)):424–431, 1970.

[156] M.R. Schroeder and B. Atal. Computer simulation of sound transmission in rooms. *IEEE Conv. Record, pt. 7*, pages 150–155, 1963.

[157] E.A.G. Shaw. Acoustical features of the human external ear. In R.H. Gilkey and T.R. Anderson, editors, *Binaural and spatial hearing in real and virtual environments*, pages 25–47. Lawrence Erlbaum Associates Inc., New Jersey, 1997.

[158] B. Shinn-Cunningham, H. Lehnert, G. Kramer, E.M. Wenzel, and N. Durlach. Auditory displays. In R.H. Gilkey and T.R. Anderson, editors, *Binaural and spatial hearing in real and virtual environments*, pages 611–663. Lawrence Erlbaum, New Jersey, 1997.

[159] M. Slaney. An efficient implementation of the Patterson – Holdsworth auditory filter bank. Technical Report 35, Apple Computer, Inc., 1993. [Cited 6.6.2002] Available from World Wide Web: <http://www.slaney.org/malcolm/apple/tr35/PattersonsEar.pdf>.

[160] J.O. Smith and J.S. Abel. Bark and ERB bilinear transforms. *IEEE Transactions on Speech and Audio Processing*, 7(6):697–708, Nov. 1999.

[161] J. Stautner and M. Puckette. Designing multi-channel reverberators. *Computer Music J.*, 6(1):569–579, 1982.

[162] J.S. Suh and P.A. Nelson. Measurement of transient response of rooms and comparison with geometrical acoustic models. *J. Acoust. Soc. Am.*, 105(4):2304–2317, Apr. 1999.

[163] U.P. Svensson, R.I. Fred, and J. Vanderkooy. Analytic secondary source model of edge diffraction impulse responses. *J. Acoust. Soc. Am.*, 106(5):2331–2344, 1999.

[164] U.P. Svensson and U.R. Kristiansen. Computational modeling and simulation of acoustic spaces. In *AES 22nd Int. Conf. on Virtual, Synthetic and Entertainment Audio*, pages 11–30, Espoo, Finland, June 15-17 2002.

[165] U.P. Svensson, L. Savioja, T. Lokki, and U. Kristiansen. Low-frequency models for room acoustic prediction. In *Proc. 17th International Congress on Acoustics*, Rome, Italy, Sept. 2-7 2001. paper 7B.09.03.

[166] T. Takala and J. Hahn. Sound rendering. *Computer Graphics (SIGGRAPH 92)*, 26(2):211–220, 1992.

[167] T. Takala, R. Hänninen, V. Välimäki, L. Savioja, J. Huopaniemi, T. Huotilainen, and M. Karjalainen. An integrated system for virtual audio reality. In *the 100th Audio Engineering Society (AES) Convention*, Copenhagen, Denmark, May 11-14 1996. preprint no. 4229.

[168] 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)*, page 149, Orlando, Florida, July 19-24 1998. Presented in SIGGRAPH'98 Computer Animation Festival (Electronic Theater).

[169] T. Takeuchi and P.A. Nelson. Robustness to head misalignment of virtual acoustic imaging systems. *J. Acoust. Soc. Am.*, 109(3):958–971, Mar. 2001.

[170] R.R. Torres, U.P. Svensson, and M. Kleiner. Edge diffraction in room acoustics computations. In *Proc. EAA Symposium on Architectural Acoustics*, Madrid, Spain, Oct. 16-20 2000. paper AAQ13.

[171] R.R. Torres, U.P. Svensson, and M. Kleiner. Computation of edge diffraction for more accurate room acoustics auralization. *J. Acoust. Soc. Am.*, 109(2):600–610, 2001.

[172] N. Tsingos. *Simulation de champs sonores de haute qualité pour des applications graphiques interactives.* PhD thesis, Université Joseph Fourier, Grenoble I, Dec. 1998.

[173] N. Tsingos, T. Funkhouser, A. Ngan, and I. Carlbom. Modeling acoustics in virtual environments using the uniform theory of diffraction. In *SIGGRAPH 2001 Conference Proceedings*, pages 545–552. ACM SIGGRAPH, 12-17 Aug. 2001.

[174] N. Tsingos and J.-D. Gascuel. Soundtracks for computer animation: Sound rendering in dynamic environments with occlusions. In *Proc. Graphic Interface'97*, pages 9–16, Kelowna B.C., Canada, May 1997.

[175] N. Tsingos and J.-D. Gascuel. Fast rendering of sound occlusion and diffraction effects for virtual acoustic environments. In *the 104th Audio Engineering Society (AES) Convention*, Amsterdam, the Netherlands, May 16-19 1998. preprint no. 4699.

[176] R. Väänänen and J. Huopaniemi. Spatial processing of sound in MPEG-4 virtual worlds. In *Proc. of the 10th European Signal Processing Conference (Eusipco2000)*, volume 4, pages 2209–2212, Tampere, Finland, Sept. 5-8 2000.

[177] R. Väänänen, V. Välimäki, J. Huopaniemi, and M. Karjalainen. Efficient and parametric reverberator for room acoustics modeling. In *Proc. Int. Computer Music Conf. (ICMC'97)*, pages 200–203, Thessaloniki, Greece, Sept. 1997.

[178] D. van Maercke. Simulation of sound fields in time and frequency domain using a geometrical model. In *Proc. 12th Int. Congr. Acoust. (ICA'86)*, volume 2, paper E11-7, Toronto, Ont., Canada, July 1986.

[179] D. van Maercke and J. Martin. The prediction of echograms and impulse responses within the Epidaure software. *Applied Acoustics*, 38(2-4, Special Issue on Computer Modelling and Auralisation of Sound Fields In Rooms):93–114, 1993.

[180] J.-P. Vian. Different computer modelling methods – their merits and their applications. In *Proc. 12th Int. Congr. Acoust. (ICA'86)*, Toronto, Ont., Canada, July 1986. paper E4-10.

[181] J.-P. Vian and J. Martin. Binaural room acoustics simulation: Practical uses and applications. *Applied Acoustics*, 36(3-4):293–305, 1992.

[182] 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.*, 86(1):172–178, 1989.

[183] M. Vorländer. International round robin on room acoustical computer simulations. In *Proc. 15th Int. Congr. Acoust. (ICA'95)*, volume 2, pages 689–692, Trondheim, Norway, June 1995.

[184] M. Vorländer, L. Kortchmar, and J. Slama. Influence of the spatial sound distribution on sound quality at workplaces. In *Proceedings of Internoise 2000*, volume 6, pages 3471–3476, Nice, France, Aug. 27-30 2000.

[185] M. Vorländer and R. Thaden. Auralisation of airborne sound insulation in buildings. *ACUSTICA united with acta acustica*, 86(1):70–76, 2000.

[186] M. Vorländer and R. Thaden. Investigation of speech privacy in buildings by means of auralisation. In *Proc. EAA Symposium on Architectural Acoustics*, Madrid, Spain, Oct. 16-20 2000. paper AAQ10.

[187] E.M. Wenzel. Spatial sound and sonification. In G. Kramer, editor, *Auditory display: Sonification, Audification, and Auditory Interface, SFI Studies in the Sciences of Complexity*. Addison-Wesley, 1994. Presented in the Int. Conf. on Auditory Display (ICAD'92).

[188] E.M. Wenzel. What perception implies about implementation of interactive virtual acoustic environments. In *the 101st Audio Engineering Society (AES) Convention*, Los Angeles, Nov. 8-11 1996. preprint no. 4353.

[189] E.M. Wenzel. Analysis of the role of update rate and system latency in interactive virtual acoustic environments. In *the 103rd Audio Engineering Society (AES) Convention*, New York, Sept. 26-29 1997. preprint no. 4633.

[190] E.M. Wenzel, J.D. Miller, and J.S. Abel. A software-based system for interactive spatial sound synthesis. In *Proc. Int. Conf. Auditory Display (ICAD'2000)*, pages 151–156, Atlanta, GA, April 2-5 2000.

[191] E.M. Wenzel, J.D. Miller, and J.S. Abel. Sound lab: A real-time, software-based system for the study of spatial hearing. In *the 108th Audio Engineering Society (AES) Convention*, Paris, France, Feb. 19-22 2000. preprint no. 5140.

[192] E.M. Wenzel, P.K. Stone, S.S. Fisher, and S.H. Foster. A system for three-dimensional acoustic 'visualization' in a virtual environment workstation. In *Proceedings of the First IEEE Conference on Visualization*, pages 329 –337, 1990.

[193] N. Zacharov and K. Koivuniemi. Audio descriptive analysis & mapping of spatial sound displays. In *Proc. Int. Conf. Auditory Display (ICAD'2001)*, pages 95–104, Espoo, Finland, July 29 - Aug. 1 2001.

ERRATA

Publication [P2]

Page 4, column 2, paragraph 5: "... IIR filter (order 30) ..." should be "... IIR filter (order 15) ...".

Publication [P3]

Page 3, in Equation (2): "$\beta_{\pm}$" should be "$\beta_{\pm\pm}$".

Publication [P4]

Page 165, column 1, paragraph 1: "... loudness 5 curve ..." should be "... loudness curve ...".