NIKLAS RÖBER

INTERACTION WITH SOUND

⠝⠊⠅⠇⠁⠎ ⠗⠕⠑⠃⠑⠗

⠊⠝⠞⠑⠗⠁⠉⠞⠬ ⠾ ⠎⠕⠥⠝⠙

INTERACTION WITH SOUND

EXPLORATIONS BEYOND THE FRONTIERS OF 3D VIRTUAL AUDITORY ENVIRONMENTS

Dissertation

zur Erlangung des akademischen Grades

Doktoringenieur
(Dr.-Ing.)

vorgelegt der Fakultät für Informatik
der Otto-von-Guericke Universität Magdeburg.

vorgelegt von: Dipl.Ing. CV Niklas Röber

Eingereicht am: September 16, 2008

LOCATION:
Otto-von-Guericke Universität
Fakultät für Informatik
Institut für Simulation und Graphik
Universitätsplatz 2
39106 Magdeburg

TIME FRAME:
June 2003 – September 2008

THESIS ORGANIZATION

The thesis is accompanied by a DVD that contains additional examples that can directly be accessed through the border icons in this pdf. Four different icons are used, which are linked to either sound or video examples, applications, or additional documents, see here also Appendix C. Additionally, color is used to highlight Bibliographical References, and references to certain Figures, Chapters and Sections. These references act as hyperlink and directly connect to the sections referenced. Turquoise References provide additional links to data folders and examples on the DVD.

Sound Example.

With this reading manual in place, I now wish all readers plenty of joy reading and fun with the additional examples and applications!

Some of the best things emanate from sound,
— a secret, a vow, a universe

— PRINCE (Musicology) (Prince, 2004)

Dedicated to my family, my friends
and my students.

ABSTRACT

Sound and acoustics both play an important role and are an integral part of our daily life. Most people, however, fail to appreciate how good they really are in interpreting sounds and noises, and are unaware of how much information is really perceived though the auditory channel alone.

This research therefore concentrates on the study of 3D auditory display systems and develops spatial sonification and interaction techniques to support an intuitive exploration of 3D virtual auditory environments. With this goal, techniques from the visual domain are adopted and existing concepts of information visualization are transferred to improve the design and sonification of 3D auditory environments. The focus lies here especially on an audio-centered design that concentrates on the benefits of an auditory display of information. Thereby aspects for an intuitive and natural exploration and interaction are discussed, leading to improvements that also include efficient graphics-based 3D sound rendering and simulation techniques. The thesis explores in this respect several areas of application. These range from interactive 3D audio-only computer games, over the design of augmented audio reality scenarios to the introduction of interactive audiobooks, which integrate interactive elements into audiobooks and radio plays. These applications are prototypically implemented, as well as examined and evaluated in detail through various user evaluations.

Additionally, promising areas of further research are discussed throughout to develop a firm basis for future development.

KURZFASSUNG

Sound und Akustik spielen beide eine wichtige Rolle und nehmen in unserem täglichen Leben einen wesentlichen Platz ein. Die meisten Menschen haben jedoch keine genaue Vorstellung davon, wie gut sie Geräusche und Töne hören und interpretieren können, und wissen nicht wieviele Informationen allein durch den Hörsinn aufgenommen werden.

Diese Arbeit konzentriert sich daher auf die Erforschung von 3D akustischen Anzeigen und entwickelt räumliche Sonifikations- und Interaktionsmethoden, welche eine intuitive Erfahrung von 3D virtuellen auditiven Umgebungen ermöglicht. Mit diesem Ziel werden Techniken aus dem visuellen Bereich adaptiert und bestehende Konzepte der Informationsvisualisierung auf die Sonifikation von 3D auditiven Umgebungen übertragen. Einer der Schwerpunkte liegt dabei auf einem akustisch bezogenen Design, welches die Vorteile einer auditiven Darstellung von Informationen hervorhebt. Hierbei werden sowohl die Aspekte für eine natürliche und intuitive Erkundung und Interaktion diskutiert, als auch Techniken für eine verbesserte, Graphikbasierte 3D Soundberechnung und Simulation vorgestellt. Unter diesen Gesichtspunkten untersucht die Arbeit auch verschiedene Anwendungsfelder. Diese reichen von interaktiven 3D Audiospielen, über die Entwicklung von erweiterten akustischen Realitäten, zu der Einführung von Interaktiven Hörbüchern, welche interaktive Elemente in Hörbücher und Hörspiele integriert. All diese Anwendungen sind prototypisch implementiert und mit Hilfe von Nutzerstudien näher untersucht worden.

Weiterhin werden Bereiche die für eine weiterführende Erforschung interessant erscheinen in der gesamten Arbeit diskutiert, um eine starke Basis für zukünftige Entwicklungen zu legen.

Among the maxims on Lord Naoshige's wall was this one:
"Matters of great concern should be treated lightly."
— Master Ittei commented:
"Matters of small concern should be treated seriously."
— Tsunetomo Yamamoto (Yamamoto, 1710-1717/2002)

ACKNOWLEDGMENTS

This thesis is the result of a five year journey that would not have been possible in this manner without the helping hands of many others. Therefore, I would like to use this space to thank everyone who helped and supported me during this time.

I am more than grateful to my family and my friends for their invaluable support and help through often difficult times. Without the support of my parents and my brother, I would clearly not have come this far. I am also indebted to my friends, with which I could not only discuss parts and open questions of my research, but who also often reminded me that there is something aside the research in a university. Here I am especially thankful to Steven and Robert Bergner, Joachim Böttger, Franziska Krüger, Dennis Kurz, Jenny Papenbrock, Schnix, Martin Spindler, Lars Stockmann, Daniel Walz and the "Lehmann Community". I am also grateful to all the other PhD students and friends I have at the Department of Simulation and Graphics, and to everyone who proof-read my thesis once it was written. Here I would like to thank Axel Berndt, Karin Engel, Tobias Germer, Dr. Timo Götzelmann, Prof. Dr. Knut Hartmann, Jan Kemmler, Alexander Kuhn, Lennart Nacke, Karsten Rink, Maik Schulze, Dr. Henry Sonnet and Maizatul Mohamad Yatim, as well as André Miede for his exceptional classic thesis LaTeX style.

During the research on this thesis, I had the opportunity to supervise a large number of students in their own study-related projects. Many of them contributed to this research through the implementation of small parts and through discussions and assistance in research questions. Here I am deeply grateful to the contributions of Sven Andres, Sven Baum, Daniel Bergner, Axel Berndt, Eva Deutschmann, Jörg Futterlieb, Michael Hentze, Sebastian Hoffmann, Hanno Hugenberg, Arvid Malyszczyk, Maik Mewes, André Miede, Dennis Sasse, and Matthias Weiss. I am especially grateful and indebted to a number of exceptional students that I had the pleasure to supervise and work with, and whose enthusiasm and diligence not only aided my research, but with whom I could also discuss the tiniest pieces in the greatest possible detail: Cornelius Huber, Ulrich Kaminski, Mathias Otto, Lars Stockmann and Daniel Walz.

Finally, I would like to thank my supervisors and reviewers, Prof. Dr. Maic Masuch, Jun. Prof. Dr. Raimund Dachselt and Dr. Ian Pitt. I am deeply grateful for their invaluable help and assistance during the times of research, and especially for their support in the rough and often difficult times towards the end. I would also like to express my deepest gratitude to our three technicians, Heiko Dorwart, Dr. Volkmar Hinz and Thomas Rosenburg, as well as to Stephanie Quade, Petra Schumann and Petra Specht for their enormous help. I am also grateful for the technical assistance provided by the group of Prof. Dr. Edgar Nett and the Fraunhofer VDTC, as well as thank the Blindenverband Sachsen Anhalt and the community of the Magdeburg Cathedral for their assistance and participation in the evaluation of the developed augmented audio reality system. I would like to thank Prof. Dr. Claus Rautenstrauch, who provided me with a scholarship that allowed me to finish this thesis in a way that would otherwise not have been possible.

THANK YOU!

CONTENTS

NOTATION

The following list defines the notation used in this thesis:

α	Absorption Coefficient
β	Diffraction Angle
ϵ	Threshold
υ_{f_j}	Weighting Factor for Frequency Band f_j
μ	Surface Roughness
λ	Wavelength
Φ	Radiant Power
ρ	Media Density
τ	Transmission Coefficient
A	Area
c	Velocity
cm	Centimeter
C	Interaction Data
D	Display Styles
dB	Decibel
$\mathcal{E}$	Enhanced 3D Environment
E_G	Geometrical Data
E_S	Structural Information
E_{kin}	Kinetic Energy
E_{pot}	Potential Energy
E_i	Incident Energy
E_e	Exitant Energy
f	Frequency
f_j	Frequency Band j
f_{update}	Update Frequency
Hz	Hertz, cycles per second
I	Intensity
k_d	Dispersion Error
kHz	Kilo Hertz (1000 Hz)
L	Radiance
$\mathcal{M}$	Enhanced 3D Model

M	Mass
m	Meter
ms	Millisecond
nm	Nanometer
O_S	Symbolic Information
p	Pressure
r	Radius
s	Second
t	Time
V	Volume
ν_J	Acoustic Pressure

PUBLICATIONS

Some ideas and figures from this thesis already appeared previously in the following publications:

[P1] Niklas Röber and Maic Masuch. *"Interacting with Sound: An interaction Paradigm for virtual auditory Worlds."*, In Proceedings of 10th ICAD Conference, Sidney, Australia, July 2004.

[P2] Niklas Röber and Maic Masuch. *"Auditory Game Authoring: From virtual Worlds to auditory Environments."*, In Proceedings of CGAIDE Conference, London, England, November 2004.

[P3] Niklas Röber and Maic Masuch. *"Playing Audio-only Games: A compendium of interacting with virtual, auditory Worlds."*, In Proceedings of the 2nd DIGRA Gamesconference 2005, Vancouver, Canada, June 2005.

[P4] Niklas Röber and Maic Masuch. *"Leaving the Screen: New Perspectives in Audio-only Gaming."*, In Proceedings of 11th ICAD Conference, Limerick, Ireland, July 2005.

[P5] Niklas Röber, Sven Andres and Maic Masuch. *"HRTF Simulations through acoustic Raytracing."*, Technischer Report Nr.4, Fakultät für Informatik, Otto-von-Guericke Universität Magdeburg, January 2006.

[P6] Niklas Röber and Maic Masuch. *"Soundpipes: A new way of Path Sonification."*, Technischer Report Nr.5, Fakultät für Informatik, Otto-von-Guericke Universität Magdeburg, January 2006.

[P7] Niklas Röber, Eva C. Deutschmann and Maic Masuch. *"Authoring of 3D virtual auditory Environments."*, In Proceedings of Audio Mostly 2006 Conference, Piteå, Sweden, October 2006.

[P8] Axel Berndt, Knut Hartmann, Niklas Röber and Maic Masuch. *"Composition and Arrangement Techniques for Music in Interactive Immersive Environments"*, In Proceedings of Audio Mostly 2006 Conference, Piteå, Sweden, October 2006.

[P9] Niklas Röber, Martin Spindler and Maic Masuch. *"Waveguide-based Room Acoustics through Graphics Hardware."*, In Proceedings of ICMC 2006, New Orleans, USA, November 2006.

[P10] Niklas Röber, Cornelius Huber, Knut Hartmann, Matthias Feustel, and Maic Masuch. *"Interactive Audiobooks: Combining Narratives with Game Elements."*, In Proceedings of 3rd TIDSE Conference 2006, Darmstadt, Germany, December 2006.

[P11] Niklas Röber, Ulrich Kaminski and Maic Masuch. *"Ray Acoustics using Computer Graphics Technology."*, In Proceedings of 10th DAFx Conference, Bordeaux, France, September 2007.

[P12] Cornelius Huber, Niklas Röber, Knut Hartmann and Maic Masuch. *"Evolution of Interactive Audiobooks."*, In Proceedings of Audio Mostly 2007 Conference, Ilmenau, Germany, September 2007.

[P13] Niklas Röber and Maic Masuch. *"Interaction with Sound in auditory Computer Games."*, In Andy Hunt, "Applications of Gestural Control of Sound: Aiding Movement", 2007.

[P14] Lars Stockmann, Axel Berndt and Niklas Röber. *"A Musical Instrument based on 3D Data and Volume Sonification Techniques."*, In Proceedings of Audio Mostly 2008 Conference, Piteå, Sweden, September 2008.

[P15] Niklas Röber, Ulrich Kaminski, Martin Spindler and Maic Masuch. *"Graphics-based Acoustic Simulations."*, In IEEE Transactions on Visualization and Computer Graphics, (submitted).

1 INTRODUCTION

Imagination is more important than knowledge. — Albert Einstein

WE live in a sensual environment; a world that is filled and stimulated through objects that can be seen, heard, touched, smelled and some of them even tasted. Out of these five senses, vision is our strongest and most prominent, but we often fail to appreciate how good we really are at interpreting sounds and noises. Sound is everywhere around us and perceived through the sense of hearing. It enriches our visual environment and accompanies us through our daily routines, informs us about current events and occurrences, and assists us in performing tasks and duties by providing auditory feedbacks. Behind the ability to derive abstract information from auditory signals lies another, more emotional layer. In the form of music and auditory reminisces, it affects and touches us in a very deep and emotional way, and has therefore – in certain cases – a much stronger influence, superior to that of vision.

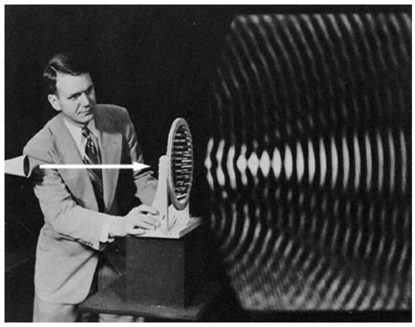
Sound and acoustics have both been practiced in the form of music and singing throughout human history, but had not been studied in greater detail until Pythagoras and Aristotle. Both looked at sound and acoustics from the viewpoint of science and discovered the very basic fundamentals of harmonics and the propagation of sound waves. Several of these effects, including the interactions (reflections) of sound waves with different materials and objects were already known at this time to Greek and Roman architects, who designed their theaters and halls with *good acoustics* in mind; the early beginnings of architectural acoustics. Today's physical definition of sound describes it as a disturbance of mechanical energy that is propagating through matter in longitudinal waves. The foundations for this understanding were laid during the time of Scientific Revolution by Galilei, Mersenne and Newton, who studied not only the physics of sound wave propagation, but also its psychological effects in perception. Newton later derived the relationship for wave velocity in solid objects, and thereby marked the beginning of the physical understanding of acoustics as we still have it today (Newton, 1687). In the 19th century, acoustics and sound propagation were studied by Helmholtz in Germany and Lord Rayleigh in England, who later also compiled the first monograph in this still very young field of research: *The Theory of Sound* (3rd Baron Rayleigh, 1877/1878).

In today's technological and information-driven environment, sound and acoustics have many applications, yet some areas are still unexplored. To shed some light – or sound waves – on some of these areas is the scope of this thesis. Sound and auditory perception have some very unique features, different to seeing and the perception of light, which shall be explored, examined and exploited for new applications and new possibilities. As characterized by the quote at the beginning of this chapter, imagining the unexplored possibilities is always the first step in the beginning of new research. In this thesis, it starts right in the next section, focussing on an INTERACTION WITH SOUND.

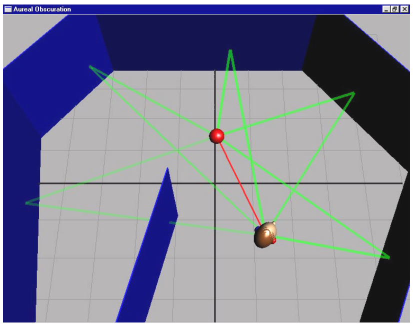
1.1 INTERACTION WITH SOUND

Envisioning an INTERACTION WITH SOUND exhibits an interactive, user-centered dialog that is located within an auditory environment. The conveyance of information in a so called *Auditory Display* focusses primarily on sound and acoustics, that means, abstract information is represented and *displayed* using auditory means and sound signals (Kramer, 1994). Although this seems difficult and applicable to a few cases only, these

(a) Léon Theremin is playing the *Theremin*, one of the first electronic music instruments. (Theremin, 1924)



(b) An acoustic lens visualizing sound waves and acoustic amplification patterns. (Knight, 1960)



(c) Wavetracing: A visualization of *Aureal's* realtime ray tracing system for room acoustic simulations. (Ramelet, 2000a)

(d) The ReacTABLE*: A tabletop designed musical instrument (synthesizer). (Kaltenbrunner et al., 2006)

Figure 1: Several examples for the *Interaction with Sound*.

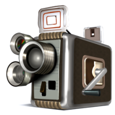
techniques are employed within a variety of applications. Most people are unaware of how important auditory information can be, and how well humans really are in the interpretation of acoustics and sound signals:

> *"Sound plays an integral role in our everyday encounters with the world, a role that is complementary to that of vision."* (Gaver, 1989)

An short example for such an auditory display is the acoustic notification for a new received email. A similar *Auditory Icon* (sound) is employed in all major mailing applications, and if we hear this sound, we immediately know that new email has arrived. However, an INTERACTION WITH SOUND describes a dialog with a broader application that can be employed for an acoustic display of various information and tasks.

The use of sound for an abstract information representation has several advantages. Most notably the possibilities for a subconscious presentation and reception, in which the information is presented in a way that the user can perform other – possibly visual – tasks with full performance, but is notified acoustically about occurring events. Existing graphics-based visualization systems can be enhanced by an auditory display, in a way that a multi-modal presentation is employed to increase the understanding of the data and the performance of the user. Although the required sound signal processing is in certain cases relatively complex, the overall hardware requirements are less demanding. This makes an application of an INTERACTION WITH SOUND very suitable for portable hardware and location-aware tasks. The design of the interactive components and the way an interaction is performed with these auditory environments is required to be audio-centered. This means that certain listening and interaction techniques from the real world should to be mimicked and integrated into the system, while focussing on an audio-only display of this information. Here the application of speech recognition and synthesis plays a large role, as speech permits for a very abstract, but also very direct presentation of information.

The design of an interactive auditory dialog system requires a semantic/semiotic component for the content, and the actual part of interaction. The fundamentals for an INTERACTION WITH SOUND are found in the area of *Auditory Displays*, but also in the visually dominant field of *Visualization* (Kramer, 1994). Schumann and Müller define three basic principles that characterize a *good* visualization by describing the expressiveness, efficiency/effectiveness and adequacy of the display (Schumann and Müller, 2000). These principles can directly be transferred to an auditory presentation of information and the design of auditory displays as well. To find and identify these similarities and shared aspects with other areas is one goal of this research. The results are used to define rules and guidelines for the design, authoring and application of 3D VIRTUAL/AUGMENTED AUDITORY ENVIRONMENTS.



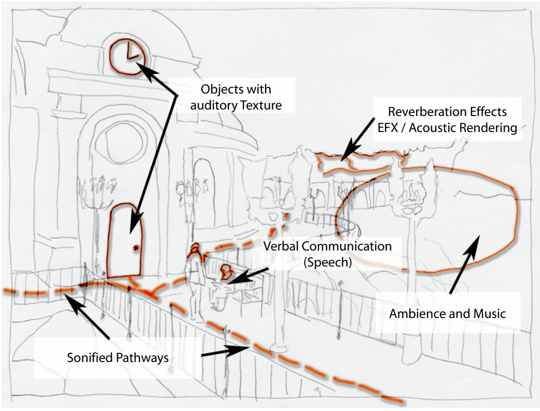
Playing the Theremin.

Examples of an INTERACTION WITH SOUND, fundamental to this research, can be seen in Figure 1. The first example in Figure 1a shows the Russian inventor Léon Theremin playing an electronic instrument that he created (Theremin, 1924). It is played by using hand gestures, with two metallic antennas sensing the hands position to control an oscillator and a volume effect. The resulting sound is very unique and the instrument is currently experiencing a renaissance. The second example in Figure 1b visualizes sound waves using an acoustic lens: *"This new technique of studying sound demonstrates the focusing effect of an acoustical lens on sound waves issuing from the horn at extreme left."* (Knight, 1960). The lens was developed at the *Bell Telephone Laboratories*, who experimented at this time with technologies for creating personal listening spaces. The study of human auditory perception and sound wave propagation accounts for a broad discussion in this research due to its high significance. Figure 1c displays the principle of *Aureal's* revolutionary wavetracing technology (Schneider and Muschett, 1998; Aureal, 2000; Ramelet, 2000a). In 1998/1999 Aureal developed the first PC sound hardware capable of performing simple, yet very realistic room acoustic simulations. This hardware employed a realtime technique called *Wavetracing* that used actual geometry to perform a more realistic – physically-sound – acoustic simulation. The ReacTABLE* system, as shown in Figure 1d, is an intuitive virtual instrument that simulates an analog synthesizer (Kaltenbrunner et al., 2006). It is played by hand using a tabletop design approach and employs computer vision techniques to interact with the instrument. Although these four examples seem very diverse, they all represent different areas of this research that are combined in an *Interaction with Sound*. Several of these ideas are referred to in later chapters, in which they are applied for an interaction and exploration of 3D virtual auditory environments.

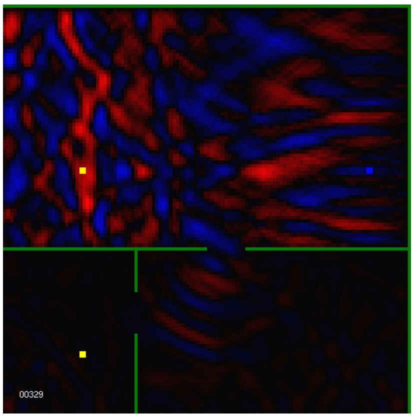
A Demonstration of Aureal's Wave-tracing Technology.

A Demonstration of the ReacTABLE System.*

The above examples show that sound and an auditory display of information has many advantages, but also several drawbacks. Some of the larger issues – compared to vision and graphics – are a limited signal bandwidth and a serial perception of information. Sound is only audible at the particular moment in time in which it is created. The ease with which humans are able to derive information from figures and drawings, but also how we perceive and listen to our environment, was learned and trained over centuries. This knowledge is very important, not only for the design of auditory user interfaces, but also for the encoding of information and its presentation. Although the visual senses have always dominated the perception of information, a proper training of the auditory skills permits an increase in performance to a level that is equal to the visual system (Gaver,



(a) A Definition of 3D virtual auditory Environments that visualizes the Concept and the Requirements to design and sonify 3D auditory Spaces, as well as denotes Techniques for an adequate Interaction.



(b) An Animation time frame that shows an efficient GPU-based Implementation of a physically correct Sound Propagation Model based on differential Equations.

Figure 2: Two Contributions of this Thesis.

1989). In several experiments this observation could be confirmed with the evaluation of own prototypes and example implementations.

Focussing on the advantages and possibilities of auditory display systems, many interesting and *cool* applications emerge capable to provide solutions for a variety of problems. The willingness to explore new, and sometimes unconventional user interfaces, is strongest among the communities of entertainment and edutainment software users (Yatim and Masuch, 2007). Especially computer games often experiment with new interaction paradigms, which later – if successful – are often applied to other, more conservative, software applications. The possibilities for using sound are not restricted to the sonification of abstract scientific data sets alone, but can, with the areas of enter- and edutainment in mind, be applied to highly immersive interactive auditory games and learning systems. These techniques can also be used to design augmented audio reality applications for the task of guiding tourists, or to aid the visually impaired in finding pathways and for the performance of daily routines. The same techniques can furthermore be used in ubiquitous and mobile computing scenarios to enhance our local environment by an *ambient intelligence*. The possibilities are not limited, but in order to realize these applications, more research in the directions of auditory perception and information presentation, especially for 3D virtual (augmented) environments is required. This shall be the scope of this research.

1.2 OBJECTIVES AND CONTRIBUTIONS

The preceding section discussed several interesting questions regarding the research for an *Interaction with Sound*. Current and existing research has so far not studied 3D virtual auditory environments in greater detail, especially not from a unifying perspective and with an audio-centered design in mind. The majority of related work employs auditory environments and 3D auditory displays to enhance the depiction of visual information, whereas this research explicitly focusses on audio-only representations to exploit the benefits of such a display. One of the objectives of this research is to explore possibilities, and to answer questions around the design and application of 3D auditory display systems and for 3D virtual auditory environments. The research thereby provides an extensive overview of the area and performs detailed studies on selected applications. A goal is to provide metrics and software construction kits, using which a user would be able to select appropriate techniques of information sonification and interaction that are both consistent with the task of application. The main objectives can be summarized as:

- An analysis of audio used in entertainment computing and the study of 3D virtual (augmented) auditory environments,
- The design of suitable sonification and interaction techniques for an exploration of and interaction with 3D auditory environments,
- A study of authoring and design guidelines for the construction of 3D virtual (augmented) auditory environments, and
- An exploration and evaluation of several example scenarios.

During the years of research, all of these areas have been studied in detail and many example applications were developed and evaluated. The main focus was thereby centered around the definition, analysis and classification of 3D virtual auditory environments along with several related application scenarios. Contributions have been made on several – partially very diverse – areas and fields of research:

- A refinement of existing 2D/3D data, image and volume sonification techniques.
- A finer definition of 3D virtual auditory environments using a *non-realistic* auditory scene design, as well as advancements regarding the sonification of 3D scene information and the development of 3D spatial interaction techniques.
- An advancement of augmented audio reality in terms of 3D spatial interactions and user-orientation/positioning techniques, as well as an evaluation of example scenarios using an efficient, self-developed low-cost system.
- A derivation of 3D scene authoring and design guidelines which are integrated into a 3D auditory scene authoring environment.
- A development of graphics-based 3D scene auralization and sound rendering techniques that improve existing solutions in terms of quality and efficiency.

Graphics-based Sound Simulations.

Figure 2 shows exemplarily two contributions of this research. Figure 2a visualizes a description of a 3D virtual auditory environment in which objects of importance are highlighted. These objects have been assigned with interactive auditory textures, which allow an intuitive exploration and sonification of this 3D auditory scene. Figure 2b displays another, more technical contribution of this thesis, and shows an animation frame of an interactive wave-based sound propagation system that exploits computer graphics hardware for an efficient and more accurate sound simulation. Although the two results seem to have – at a first glance – not much in common, yet both rely on each others principles. As the interaction with audio-only environments is difficult and requires accurate techniques for sound rendering and simulation that currently available APIs can only partially fulfill, additional research was required to design an efficient, yet highly realistic, sound rendering and simulation system.

1.3 PERSPECTIVES

From the perspectives described in this section, several directions of research are opening up. The one chosen in this thesis will continue towards 3D auditory displays and the design of 3D virtual/augmented auditory environments. Along this way not only the theoretical foundations are explained and discussed in detail, but also several fundamental techniques and prototypic applications are developed and explored. Chapter 2 starts with a theoretical and at the same time very global perspective of an INTERACTION WITH SOUND, and classifies the thesis' topic within its larger areas of research in computer science. Chapter 2 also presents the major scientific goals of this research and discusses issues of methodology and special requirements for an effective implementation. At the end of Chapter 2, an overview of the thesis' organization is provided, which describes the content and the contribution of each chapter.

The way of the Samurai is one of immediacy, and it is best to dash in headlong. — Tashiro Tsuramoto

2 ANALYSIS

THE goal of this research is to explore and foster 3D virtual/augmented auditory environments – especially within the domains of entertainment and edutainment applications – by using tools and techniques from computer games, 3D auditory displays and scientific data sonification. This chapter is used as a starting point for this research and introduces the most important concepts and terminologies, as well as classifies the thesis topic and analyzes the major scientific goals. Furthermore, this chapter provides an overview of the methodologies employed and discusses some of the requirements and settings necessary to obtain the identified objectives. As the research employs techniques from several different areas – not all from within computer science – this chapter aims at providing a broader perspective in which to present the topic as a whole and to unify the domains involved. In reference to the quote at the beginning of the chapter, it is often favorable to not ponder too long with a task at hand, but to *dash in headlong*.

2.1 ANALYSIS AND CLASSIFICATION

The topic of *Interaction with Sound* overlaps with several domains in computer science and its associated fields. Although the research is centered around sound and acoustics, methods of Human-Computer Interaction (HCI) play a major role in evaluating the quality and usability of the designed systems and applications. The heart of this thesis are 3D VIRTUAL AUDITORY ENVIRONMENTS, around which all research and discussions are arranged. With the means of perceiving information consciously reduced to hearing alone, the interaction and sonification techniques that are required have to be designed to compensate for some of the missing visual cues. These differences in perception, although resulting from missing pieces of information, yield to a new way of information presentation that offers alternative ways to describe 3D virtual environments. Depending on the task and area of application, the requirements therefore vary and are expressed through the different sonification/interaction techniques and the alternative auditory presentations used. Some of the basic techniques can also be applied and evaluated using the sonification of 2D and 3D scientific data sets. The focus, however, lies on 3D auditory spaces and a combination with concepts and techniques from 3D computer games to define interactive 3D virtual/augmented auditory environments.

Based on this short analysis, the fundamental areas of this research can be identified as the following five domains:

- Human-Computer Interaction (HCI),
- Auditory Displays,
- Audio and Acoustics,
- 3D Virtual and Augmented Environments, and
- Entertainment and Edutainment Applications (eg. Computer Games).

Each of these topics represents an individual research domain or area of application, but combined, they shape the foundation of this research with 3D virtual auditory environments at its center. Overlapping areas have to be identified and analyzed in order to successfully combine all topics and unify their strengths and advantages. So far, a clear definition of 3D virtual auditory environments that integrates both virtual reality (VR) and 3D auditory display systems is still not available, especially not from the perspectives of entertainment and edutainment applications. Although the term is used and known, it is employed in an ambiguous manner and defined from varying positions. Therefore, a clear definition with respect to the above mentioned research domains is required. This definition must also include 3D scene sonification and interaction techniques to convey information of the virtual environment to the user, and to input user information into the system. The goal of this thesis is to develop and establish 3D auditory environments as an *equal* to visual environments. In the following, each of these five research domains is shortly introduced and discussed in their own context, as well as in respect to their possible contributions and potential for 3D virtual auditory environments.

HUMAN-COMPUTER-INTERACTION defines the upper layer of abstraction of this research. It is the interdisciplinary study of interfaces and the interaction between humans and machines, and is concerned with their design, implementation and evaluation. With the focus on an auditory perception and communication, it is used to study the acoustic conveyance and presentation of data and information. In addition, HCI is also responsible for the design and development of efficient and effective techniques for an intuitive interaction with 3D auditory spaces. The goal will be the definition of building sets and guidelines, which allow a task-dependent selection of suitable sonification and interaction techniques. Using software evaluation techniques, these selections, as well as the prototypes developed, can later be evaluated and their functionality be assessed. Further contributions are concerned with the development of authoring and design guidelines to provide an adequate balance between an applications function and its aesthetics in presentation.

AUDITORY DISPLAYS represent the closest area of related work and at the same time provide the basic methods and technologies necessary to design and implement the ideas of this research. Auditory displays are similar to visual displays in the respect of *displaying* information, except that auditory displays are based on auditory means and primitives to convey abstract information and data. Many of the thesis' concepts and ideas are directly related to existing approaches emanating from this area. The approaches developed within this thesis can therefore be very well compared with related and existing concepts and techniques.

AUDIO AND ACOUSTICS are used in this research as the primary means of communication, and are therefore the focal point of all techniques, methods and applications developed. As the acoustic perception of information is quite different to visual seeing, a large portion of this research concentrates on an efficient presentation using a non-realistic auditory design, ie. an auditory representation that deliberately departs from a physically correct display and concentrates on a perceptual presentation. A direct requirement for this task are efficient techniques for for a high-quality acoustic rendering and 3D sound synthesis.

3D VIRTUAL AND AUGMENTED ENVIRONMENTS contribute both the platform and the stage to this research. Virtual reality and 3D virtual/augmented environments are very common in computer graphics, and therefore many of the visual display and interaction designs available can be transferred into their auditory counterparts. The challenge of this research is to map these concepts towards auditory presentations and onto auditory primitives respectively. This requires rules and guidelines for the mapping of abstract information, as well as suitable 3D spatial interaction techniques for the design and authoring of 3D virtual auditory environments.

entertainment and edutainment applications serve as basis and test ground for the majority of examples and implemented prototypes. Advantages of using computer games are the great number of possible scenarios and a high suitability to evaluate new and unconventional interface designs and interaction techniques. Furthermore, computer games and edutainment applications can be used to assess the quality of additional attributes, such as storytelling, narration and immersion. In a future setting, the here developed concepts and techniques for an interaction with auditory environments should be applicable to scenarios beyond entertainment and edutainment as well.

This domain analysis and the classifications of the topic allow now a formulation of the most important research questions and scientific objectives. With the focus on 3D virtual auditory environments and the goal to provide techniques for an intuitive 3D scene sonification and interaction, the two primary research objectives are:

- The analysis, survey and classification of 3D virtual auditory environments in terms of presentation, realism, interaction and area of application, as well as
- The development of metrics and techniques to define and select suitable and task-dependent 3D sonification and interaction techniques for the tasks of 3D scene exploration, the conveyance of abstract object/scene information, as well as to perform 3D spatial object/scene interactions.

A major challenge is the detailed analysis and comparison of visual and auditory environments regarding the conveyance and representation of information, as well as the development of methods for an audio-centered selection, object emphasis and scene/object interaction. As both environments represent and display abstract information using either visual and/or auditory means, several concepts from information visualization might be applicable and transferable to the auditory realm as well. Problems will not only arise due to the differences in perception, but also due to the different environments that both senses describe. With this in mind, the secondary objectives for this research are identified as:

- The analysis of audio in entertainment computing and in auditory displays.
- A detailed comparison of visual and auditory 3D environments with a focus on presentation and interaction.
- The proposition of techniques and guidelines for the design of enhanced non-realistic 3D virtual auditory environments.
- The design of 3D scene sonification and spatial interaction techniques, suitable for an exploration and interaction within 3D virtual auditory environments.
- The conception of authoring techniques and guidelines for the design of 3D virtual auditory environments.
- An evaluation regarding an applicability and implementation of 3D augmented audio reality.
- The design and implementation of an interactive audio framework, applicable to several different example scenarios and areas.

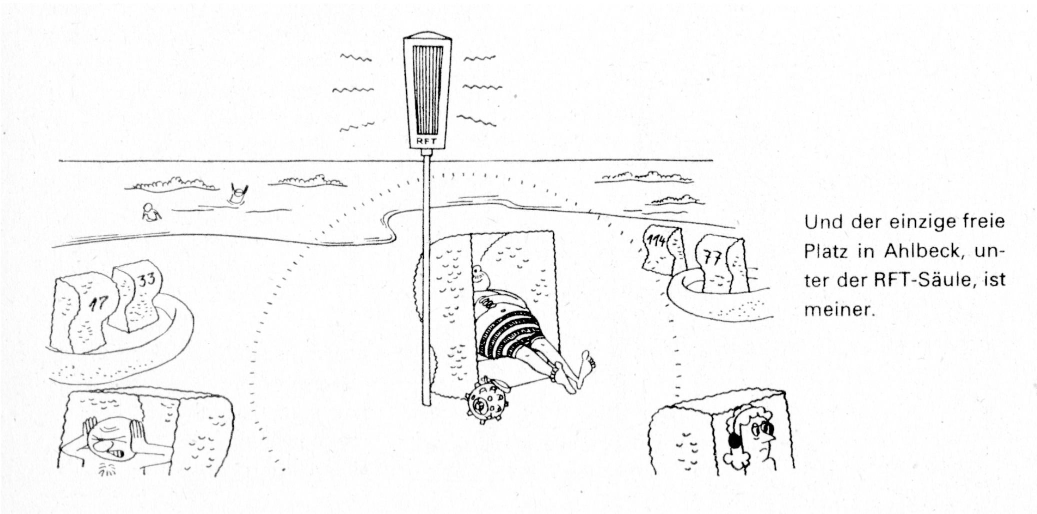


Figure 3: The auditory hedgehog: *"And the only free place in Ahlbeck, under the RFT speaker, was mine."* (Schmitt, 1968).

2.2 OBSERVATIONS

A major difference between an auditory and a visual perception of information is that auditory information is constantly audible during the actual display. Unlike eyes, which can be closed, every acoustic signal in the local environment that is audible *will be* audible. This problem makes a close proximity installation of several auditory displays difficult if the information is presented over a speaker array and not through headphones. Therefore, the research in this thesis employs headphones as actual *physical* auditory display in almost all examples.

The ever growing flow of noise and auditory output in our modern western civilization serves as an unpleasant experience to more and more people. Schmitt, a cartoonist born in the GDR, devised here several interesting ideas in his book *"Cartoons in the service of science"*, which are not only applicable to sound and acoustics (Schmitt, 1968). His auditory hedgehog *bends* sound waves into straight lines and is thereby creating a silent spot for his own personal enjoyment, compare with Figure 3. Using modern noise canceling headphones[1] and efficient acoustic insulations, this experience can be partially recreated.

Aside this interesting direction of research, the following two sections focus more on the objectives that were devised at the beginning of this chapter. A preliminary concept for the research along with necessary procedures is formulated and later extended by an initial requirements analysis that is concerned with an implementation of the above objectives.

2.2.1 *Procedures*

With the analysis completed and the scientific goals identified, a first concept can be drawn that summarizes the research and a realization of the envisioned objectives. But first, the major research questions that are to be answered by this thesis are defined in the form of hypotheses. The underlying theories will be examined in greater detail by

1 http://www.noisefreeheadphones.com/

the individual chapters and are later evaluated and either confirmed or rejected at the final discussions in Chapter 10.

The initial questions for the development of a first concept and approach towards 3D virtual auditory environments are:

- What can be described with 3D virtual auditory environments, and what are the possible applications?
- What level of realism is required and achievable?
- How important is 3D sound spatialization, and what are the individual perceivable differences?
- What is the most intuitive way to acoustically display information, and how can one interact with such an auditory environment?
- How can a real-world environment be combined with an artificial auditory space?
- What are the requirements for the design of 3D auditory environments, and where are the limitations?
- How accurate is the human perception of 3D sound and environmental acoustics, and what are the requirements for a realistic acoustic rendering and sound simulation?
- How suitable are 3D auditory environments for gameplay, and what level of immersion can be achieved?

Several steps are required to answer these questions. The first one is an extensive analysis of existing and related work along with a detailed study of auditory perception and acoustical effects. These results will allow a first approximation in the direction of what can be achieved and how 3D virtual auditory environments have to be designed and constructed for an effective and expressive auditory display of information.

The second step will involve a closer observation of 3D auditory environments in order to derive suitable task-related sonification and interaction techniques. A possible approach is here to employ a collection of generally suitable techniques and to use metrics to decide upon and select appropriate methods of interaction that are consistent with the sonification goal specified. Potential difficulties may arise from auditory perception, insufficient sound rendering and synthesis techniques, as well as from problems of mapping abstract information onto acoustic primitives for an auditory display. Depending on their severity, more research has to be conducted in either of these areas to improve the perception, sonification and interaction with 3D auditory environments.

In a third step, the idea of combining a virtual auditory environment with real surroundings moves into focus. The question is here whether or not it is possible to project a convincing artificial auditory environment onto an existing real one: enhancing and augmenting a natural auditory environment through artificial sound signals. Several ideas of promising applications are emerging and range from guiding systems for the visually impaired to interactive systems for an auditory storytelling. As the direction of research loosely points towards entertainment and edutainment applications, questions regarding an auditory gameplay and the achievable degree of immersion arise. Due to the missing visual cues, which are now substituted by the players own imagination, it can be assumed that an audio-only presentation achieves a higher level of immersion. This would make auditory environments very applicable for narration and adventure-based computer games.

Besides the development of prototypes and applications, it is also important to know the limitations of 3D auditory environments and auditory displays. Therefore, an additional emphasis lies on devising rules and guidelines for an authoring and design of interactive 3D virtual/augmented auditory environments.

2.2.2 *Requirements*

In order to attain the above objectives and to answer the research questions described, several requirements have to be fulfilled. The initial comparison of visual and auditory presentations includes a review of related work and similar approaches, but also the conception of small prototypes for a more detailed study of auditory perception. These prototypes are integrated into a more sophisticated audio framework, which will be utilized throughout the research for the evaluation of the proposed concepts and ideas. This framework might also be used for creating prototypic applications, but its main task is to evaluate specific sonification and interaction methods for the design of a task-dependent construction kit. Such a framework can be designed analogously to the layout of computer game engines, which support a visual and auditory display of information, as well as an interaction with 3D virtual environments. With the focus on 3D virtual auditory environments, a first estimate of the requirements is:

- A 3D polygon-based scene management system which supports collision detection,
- Highly efficient and accurate 3D sound rendering and room acoustic simulation techniques, as well as
- Several possibilities to implement 3D interaction paradigms, such as exploration, orientation and navigation, and 3D object/scene interactions.

An initial design approach that can be employed for the first evaluations and test scenarios can be developed using standard APIs and libraries, such as OpenGL, OpenSG, OpenAL and DirectX. This prototype can later be extended to support more advanced 3D spatial interaction techniques using 3D user tracking hardware. Another interesting possibility emerges with the replacement of the standard sound rendering API (eg. OpenAL), which can be substituted through a dedicated system that supports a higher accuracy and efficiency. This discussion of a requirements analysis is continued in later chapters, which explicitly concentrate on system design and implementation.

All of the research questions and objectives discussed are further examined and studied in their respective chapters. The next section provides an overview of the structure of the thesis, as well as a summary of each chapter.

2.3 THESIS OUTLINE

The thesis is organized and structured into 10 chapters, of which the first chapters are of more introductory nature and lay a foundation for the coming design of 3D virtual auditory environments. Code examples and implementation details can be found in Appendix A, but are also integrated within the algorithmic discussions of the respective chapters. Specific examples and results are discussed throughout the thesis, but are additionally summarized in Chapter 9 and Appendix C. Chapter 9 provides additional details and examines several prototypic implementations and case studies from a chapter-overlapping perspective.

The thesis is organized as follows:

FUNDAMENTALS — Chapter 3 and Chapter 4 illustrate and discuss the fundamentals of this thesis, as well as examine related work and similar systems. Chapter 3 concentrates on the fundamentals by introducing important concepts, terminologies and applications, along with the necessary details of auditory perception and sound signal processing. The goal is to familiarize the reader not only with physiological and psychological, but also with important technical aspects of sound perception and rendering. The chapter concludes with a discussion of audio in entertainment computing and examines several existing applications.

AUDITORY DISPLAY — After the discussion of basic fundamentals, Chapter 4 concentrates on the application's side and analyzes 2D and 3D auditory display systems and sonification techniques. The chapter examines existing concepts and implementations, with a focus on an intuitive and efficient sonification of information. The chapter also discusses issues and open problems of an auditory user interface (AUI) design, as well as provides a general overview on the subject of auditory presentation and display using several examples.

AUDITORY ENVIRONMENTS — The following Chapter 5 through Chapter 8 contain the main contributions of this research. Chapter 5 starts here with a discussion of 3D virtual auditory environments. After an abstract definition of virtual reality and 3D auditory environments, the chapter concentrates on an intuitive auditory scene design and develops the concept of a non-realistic auditory scene presentation. Using this approach, the remaining sections discuss and implement techniques for 2D/3D data and 3D scene sonification, as well as develop intuitive methods for 3D spatial interaction. The chapter concludes with a discussion and the design of an audio framework that allows an implementation and evaluation of the described concepts.

AUGMENTED AUDIO — Chapter 6 directly connects to Chapter 5 and extends the framework developed towards an augmented and mixed reality design. In this chapter, a low-cost augmented audio reality system is described and developed, and applications that combine a real-world environment with an artificial auditory scene are discussed. Although augmented audio reality possesses many advantages and possibilities, it also exhibits several difficulties that are addressed within this chapter. This includes an extension of the previously devised sonification and interaction techniques, as well as the development of efficient methods for user tracking and positioning. In alliance with Chapter 5, the last section extends the existing audio framework and discusses possible applications for augmented audio reality.

AUTHORING AND DESIGN — While the two previous Chapter 5 and Chapter 6 were more concerned with an implementation of 3D virtual/augmented auditory environments, Chapter 7 provides a closer look regarding issues of authoring and design. As the auditory channel has a small bandwidth and information can only be perceived in a serial manner, these issues are of high importance. The chapter therefore concentrates on the development of rules and guidelines for the various authoring tasks, and also devises an authoring environment that implements these concepts exemplarily.

ACOUSTIC RENDERING — Chapter 8 is motivated through the special requirements for the design of auditory environments as discussed in Chapter 5 and Chapter 6. Chapter 8 discusses and implements efficient and accurate techniques for an acoustic rendering of 3D spatial auditory environments. After a short introduction of the subject, several techniques for the spatialization of monaural sounds, as well as for the simulation of room acoustics are explained. Due to several qualitative and quantitative requirements, an efficient implementation of these techniques using computer graphics and graphics hardware is provided, as well as compared with state-of-the-art implementations. The chapter concludes with the discussion of a sound engine that not only exploits computer graphics hardware, but also transfers graphics-based designs towards 3D acoustics and 3D sound simulations.

CASE STUDIES — Several examples and results are discussed throughout their respective thesis chapters. Chapter 9 summarizes the majority of these implementations and presents applications and case studies from a broader perspective. The majority of applications and examples that are discussed within this chapter were examined and evaluated through user studies. Their results are discussed individually for each area of application. The chapter focusses especially on the evaluation of sonification and interaction techniques for 3D auditory scenes, but also discusses 2D/3D data sonification, audio-only computer games, augmented audio reality applications and interactive audiobooks. The chapter concludes with an additional analysis of the sound rendering techniques developed in Chapter 8, and summarizes the results from all applications.

CONCLUDING REMARKS — Finally, the thesis is summarized and conclusions are drawn in Chapter 10. The initial goals are compared with the results achieved, and propositions are developed to describe the essence of this research. The chapter also discusses ideas and possibilities for future improvements that could not be addressed in this thesis.

APPENDICES — Succeeding the chapters of this thesis are three appendices, which provide additional code examples (Appendix A), discuss the user studies and evaluations in greater detail (Appendix B), as well as list and describe the examples contained on the accompanying DVD (Appendix C).

3 FUNDAMENTALS

Although all knowledge begins with experience, it does not all arise from experience. — Immanuel Kant

PERCEPTION describes the human process of acquisition, selection and interpretation of sensory information. The five human senses are *Sight, Hearing, Touch, Smell* and *Taste*, which are, although not of equal, but all of high importance (Goldstein, 2007). The knowledge we possess, which enables us to interact with our environment is entirely based on personal sensory experiences. However, one will never know how well the human perception corresponds to reality itself.

The main focus in this thesis lies on auditory perception and the display of information using primarily sound and acoustics. Therefore, this chapter serves as an introduction to this topic and discusses necessary and important concepts and terminologies. It starts with an examination of common data and information visualization techniques with the goal of deriving essential paradigms that can be adopted and employed in an auditory information display as well. Following this, the chapter discusses fundamentals of auditory perception and psychoacoustics, as well as provides a short overview of sound rendering and the principles of physical sound propagation. The chapter concludes with an analysis of sound employed in entertainment computing and discusses audio-only computer games, edutainment examples, as well as audiobooks and radio plays.

3.1 INFORMATION ANALYSIS AND DISPLAY

Using our five senses we constantly analyze our local surroundings and evaluate the data we perceive. Most of this analysis is performed subconsciously, in which only *important* information attracts our interest and reaches our awareness (Goldstein, 2007; Matlin, 1987). Importance is here defined through the context in which the information is acquired, as well as through previous encounters of the same stimulus. The overlap of several modalities (senses) thereby increases the importance and allows a more detailed analysis of the occurring event. If an interesting stimulus reaches our awareness, for example we hear a car, other senses are also directed to the same stimulus to gather more, and possibly different, information (Goldstein, 2007; Begault, 1994). In this example we would turn our head to identify an approaching car and stop at the sidewalk. This perception of information has evolved over a long period of time, and can, if the processes of perception are well understood, also serve an analysis and exploration of abstract scientific data sets and information.

Information analysis and visualization are both an essential part of modern life. We encounter it in our daily routines by interpreting weather maps and stock market data, as well as by exploring our favorite news website or the local computer hard drive. In all cases, the underlying abstract information is transformed into pictures that engage our visual system for interpretation (Schroeder et al., 2004; Schumann and Müller, 2000). As a result, we perceive an understanding of the information and can act accordingly, eg. bring an umbrella if the weather map shows rain. The majority of information is thereby presented visually, although, and due to the raising flow of available information, the use of other modalities, such as hearing and haptics, is increasing as well.

With this research's focus on auditory techniques for the representation of data and abstract information, the following two sections assess the possibilities to adopt methods from scientific visualization and a graphics-based 3D interaction towards an acoustic sonification that is combined with a 3D spatial audio-centered interaction.

3.1.1 *Visualization and Information Presentation*

The visualization of data and information has various applications and is used in many areas with an increasing deployment. The goal for using visualizations is the analysis, understanding and communication of models, concepts and data (Globus and Raible, 1992; Schumann and Müller, 2000). The employed *visualization pipeline* is thereby always the same and independent from the data and the visualization goal, see Figure 4. At the beginning it starts with raw data, which is often transformed in a second step to extract the important – sometimes hidden – information. For this step, techniques from the areas of data mining and/or signal/image processing are often employed. The third step is of high importance and involves the mapping of abstract data values onto visual geometric primitives. These objects are rendered and displayed in the last and final step. Figure 5 shows a selection of common visualization techniques, in which the same information is displayed using different primitives. The pipeline that is shown in Figure 4 can – with adaptations – directly be transferred and applied to a sonification of data as well, see here also Chapter 4. The most important difference is the mapping of data onto acoustic primitives, such as loudness, pitch and frequency, and the final auditory display of this information.

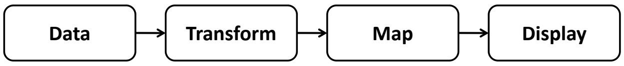


Figure 4: Visualization Pipeline.

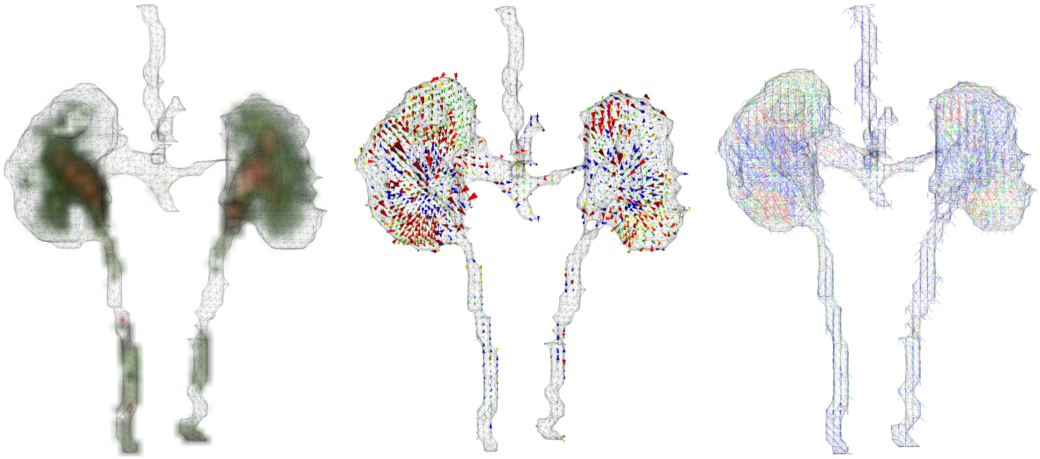
One of the critical and often most difficult steps in the visualization pipeline is the extraction of important information. Smith et al. describe it as to

> *"find signatures in the data, features or set of features that 'pop out' as a result of a precognitive processing." (Smith et al., 1992)*

This data extraction must be performed in a way that no ambiguities and no additional information is added. Schumann and Müller define here three basic rules that have to be obeyed: *expressiveness*, *efficiency/effectiveness* and *adequacy* (Schumann and Müller, 2000). Expressiveness refers to an unaltered representation of the underlying information — the content, while efficiency and effectiveness are a measure of how good the information is mapped through geometric primitives and how well the visual system is supported. Adequacy describes a cost-benefit ratio and is a measure of how intuitive a visualization is. As with the visualization pipeline, also these rules directly apply to sonification and to an acoustic presentation of data and information.

Figure 5 shows three examples of the most commonly used 3D data visualization techniques (Röber, 2000; Tory et al., 2001). It shows time frames of a dynamic SPECT data set of a human kidney and ureter, and displays the washout of previously injected radio-pharmaceuticals. The data is visualized using volume rendering (Figure 5a), 3D glyphs (Figure 5b) and as a 3D hedgehog display using line segments (Figure 5c) (Schroeder et al., 2004; Kitware Inc., 2008). While the first example only shows the concentration of the radiopharmaceuticals, the second and third visualization also show directional information of the flow and the speed of the washout. All examples additionally display contextual information by showing the kidneys/ureter structure and position (wireframe), an important concept that allows an easier perception and correlation of the data.

While the examples in Figure 5 clearly display the information contained, a multi-modal, or audio/visual representation and exploration of the data might improve the understanding and the efficiency of the perception. Rossiter and Ng developed a system



(a) Volume Rendering. (b) Flow Visualization (Glyphs). (c) Flow Visualization (Lines).

Figure 5: 3D Visualization Examples.

that traverses a volumetric data set and assigns, after a prior classification, all data values a certain musical instrument with an according fundamental frequency and amplitude (Rossiter and Ng, 1996). On this basis, the system can be used for an audio/visual exploration of a volume data set, in which sound is employed as an additional classification hint. Later, Amandusson extended the VTK by basic sonification techniques that allow such a multi-modal presentation of certain data sets using computer graphics and an accompanying auditory sonification (Amandusson, 2003; Schroeder et al., 2004; Kitware Inc., 2008). These examples display some of the possibilities of an acoustically enhanced visualization system, to which we will refer in more detail in Chapter 5.

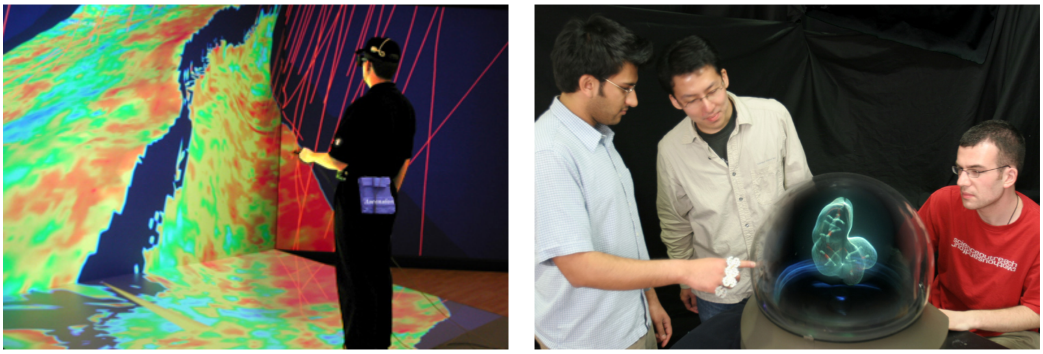
3.1.2 *Human-Computer Interface Design*

Without a proper interface and suitable techniques for interaction and exploration, the visualization methods discussed in the last section are only half as conclusive. *Human-Computer Interface (HCI)* is the study of the interface and the interaction between users and a computer system. As the applicability of HCI is very broad and an in-depth discussion beyond the scope of this thesis, this section introduces only basic concepts for 3D displays and presentations, as well as for spatial interactions with 3D data sets and within 3D virtual environments.

Raskin complains that both, HCI and interface design are too often reduced of being only the interaction with a window system, but:

> *"the way you accomplish tasks with a product, what you do and how it responds – that's the interface." (Raskin, 2000)*

Several guidelines for designing 3D user interfaces are available in the literature and shall here only be reviewed briefly (Shneiderman, 2004; Faulkner, 1998). Most systems and design philosophies focus on a user-centered design and identify the user as the main element after which everything else is arranged. Designers compile the needs, limitations and goals of the user and create an interface that addresses these elements. Several compilations with principles for an efficient user interface design exist, and are based on perceptual, mental model, attention and memory based guidelines (Tidwell,



(a) 3D CAVE Display with spatial Interaction[1] (Ascension Technology, 2007).

(b) Volumetric Display with Gesture Interaction (Grossman et al., 2004, 2007).

Figure 6: 3D Display and Interaction Examples

1999; Wickens et al., 2003). Specifically important for the design of a user interface are: *Tolerance, Simplicity, Visibility/Hearability, Affordance, Consistency, Structure* and *Feedback* (Tidwell, 1999).

Besides these seven elements for an efficient and intuitive interface design are additional requirements based on application and interaction-dependent needs. Figure 6 shows two examples of 3D displays that allow real 3D spatial interactions. Figure 6a shows a CAVE environment with a six degree-of-freedom (DOF) tracking equipment (Ascension Technology, 2007), while Figure 6b visualizes a gesture-based interaction system with a 3D volumetric display (Grossman et al., 2004, 2007). Both applications require a different design and different techniques for the user interaction. The user in the CAVE system (Figure 6a) is *inside* the visualization, while the user of the volumetric display controls the application from the *outside*. The interaction with an application is often one of the *key* elements that decide whether or not an application is successful. Crawford stated in terms of interaction that

> *"the term interactivity is overused and underunderstood." (Crawford, 2002)*

Crawford defines interaction as the *listen, speak, act* interaction loop, which centers the user in the focus of the application and the interface design (Crawford, 2002). Important for Crawford thereby is that the user receives enough feedback from the system at an interactive rate. This *update rate* varies depending on the type of application, and is for 3D environments and a graphics-based visualization specified as being equal or preferably above 15Hz (Shneiderman, 2004; Crawford, 2002)

The interaction in this research is centered around an analysis of 3D data sets and an exploration of 3D virtual environments. For the performance of these tasks, three viewpoint-metaphors have evolved which support different analysis approaches: *scene-in-hand, eyeball-in-hand,* and *flying-vehicle-control* (Ware and Osborne, 1990). In the form of *listening-metaphors*, eg. *ear-in-hand*, these techniques can directly be applied and used for an auditory sonification of data (Stockmann, 2008). For user interaction and the input of information often specialized 3D tracking equipment (Ascension Technology, 2007) and other 3D technology is used. These devices provide up to six degree-of-freedom and allow an implementation of real 3D interaction metaphors, including 3D gestures and the design of 3D user interfaces (Hand, 1997; Dachselt and Hinz, 2005). The design of such 3D user interfaces and spatial interactions is often based on 3D widgets that are

1 Image courtesy of Schlumberger and Norsk Hydro

centered *around* the user and which mimic real-world interactions and behavior (Hand, 1997; Dachselt and Hinz, 2005).

As evident from the last two sections, the basic principles and design rules that apply to 3D visualizations and 3D visual user interfaces, are both – nearly directly – applicable to data sonification and auditory displays as well. Following chapters will continue these discussions and develop, based on these rules, dedicated spatial sonification and interaction techniques for an exploration of 3D virtual/augmented auditory environments.

3.2 AUDITORY PERCEPTION AND PSYCHOACOUSTICS

The acoustic perception and classification of our environment differs – beyond doubt – from seeing and the information gathered through our visual senses. In certain cases, for instance a moving car, both environments overlap and display the same information, but with different impressions. In most cases, however, the visual and the auditory information perceived are disjunct. However, in the form of synesthesia, an acoustic sensation can also be experienced as color and vice versa (Goldstein, 2007). Auditory perception offers a very broad spectrum that ranges from the listening of sounds and noises to the cognition of speech and music. With this thesis research focussing on an auditory representation of abstract data and information, it is imperative to understand the processes of listening and auditory perception, as well as to identify the challenges in the design of an auditory display. This section provides the fundamental details that are required for the later discussions on 2D/3D data sonification and the design of 3D virtual/augmented auditory environments (Warren, 1999).

3.2.1 *The auditory System*

The physical definition of sound describes it as mechanical wave of pressure variations that propagates through matter and participating media. These pressure variations are sensed by the ear, which transmits the frequencies and amplitudes perceived for an interpretation in the brain (Goldstein, 2007; Matlin, 1987). A human listener is able to perceive frequencies ranging from about 20Hz to 20kHz with an intensity range of 120dB, eg. from the rustling of leaves to a starting aircraft. Loudness is thereby frequency-dependent, in which we perceive lower frequencies with a given amplitude *louder* than higher frequencies of the same amplification.

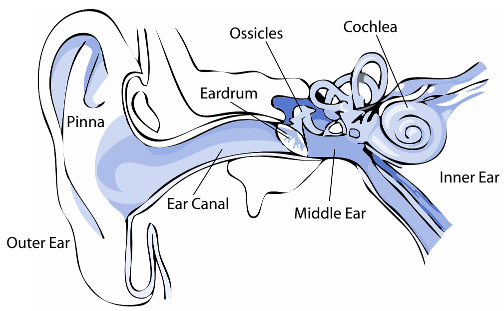
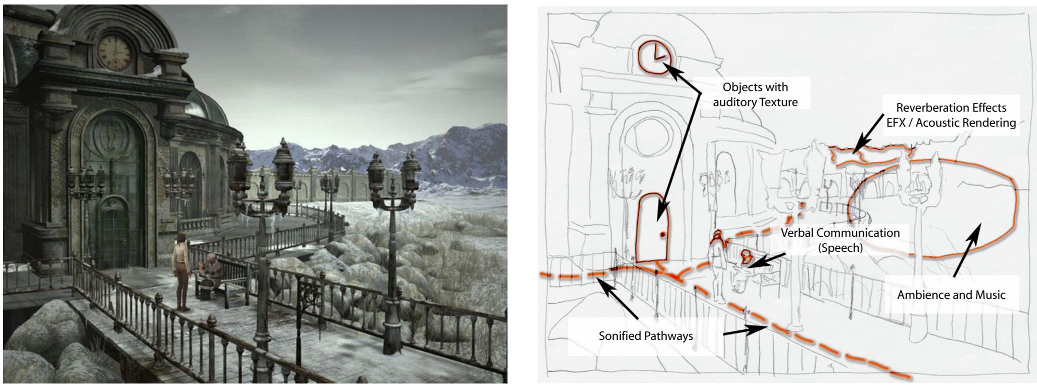


Figure 7: The Human Ear[2].

Figure 7 shows a cross section of a human ear with its outer, middle and inner parts. The outer ear consists of the pinna, the ear canal and the eardrum, which protects the fine and very sensitive parts of the middle and inner ear. The ear canal enhances frequencies between 2,000Hz and 4,000Hz due to resonance effects, and is, together with the shape of the pinna, responsible for 3D sound perception. The middle ear is the small room between the eardrum and the inner ear. It contains the auditory ossicles: *hammer, anvil* and *stirrup*, which transmit and further amplify the sound waves perceived. The inner ear is filled with a viscous fluid and transmits the vibrations from the occicles to the cochlea. The recepting organ that

2 http://www.health.state.ny.us/nysdoh/antibiotic/4815.htm



(a) Visual Scene Presentation.

(b) Auditory Scene Presentation.

Figure 8: Visual vs. auditory Scene Presentation (Syberia (Microïds, 2003))

is responsible for the actual *hearing* is the Organ of Corti. It runs through the center of the cochlea duct and consists of many thousand stereocilia (hair cells). As the ossicles transmit the vibrations to the fluid, the cochlea duct is set to vibrate at the same frequency, and with it the Organ of Corti. These vibrations excite the stereocilia, which itself convert and transmit them to the listening nerves. These impulses are transmitted to the brain and analyzed in the listening center. Additionally, the inner ear also contains the vestibular apparatus, the organ of balance, refer to Figure 7.

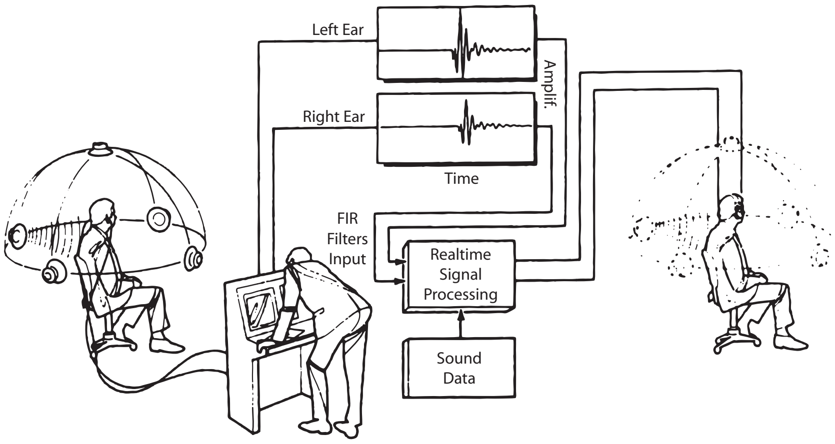
3.2.2 *Perception and Psychoacoustics*

Hearing and auditory perception is much more complex than just the interpretation of different frequencies and amplitudes. Through the analysis of raw sound data, the human brain is able to identify sounds and noises, as well as to understand speech and to appreciate the sound and rhythm of music. Additionally, all sounds are perceived in 3D with a respective position and distance relative to the listener. This allows the creation of a mental model of the local auditory environment. The process in which these auditory signals are divided and classified into perceptual components is called *auditory scene analysis* (Bregman, 1990). Bregman describes this analysis analog to seeing, in which distinct auditory events are identified and classified similar to visible objects. These acoustic events are grouped and positioned and represent elements of the auditory scene. This auditory scene usually differs from the perceived visual information, as we also receive auditory signals from objects behind ourselves and from objects that are hidden and not visually observable. Figure 8 displays a comparison of a visual and an auditory scene representation, in which Figure 8b labels the most important information and shows how this data can be communicated by using auditory means.

The human brain is thereby able to identify several psychoacoustic parameters, such as loudness, pitch and harmonics, with their respective physical correspondent, eg. loudness/amplitude, pitch/frequency. Interesting to note is that auditory signals are processed faster and are also easier to detect than visual signals (Wenzel, 1992).

Psychoacoustics

Psychoacoustics is an area of research that connects human listening experiences with physical parameters of sound wave propagation. Listening is defined as being either



(a) HRTF Measurement using Pinna Responses.
(b) FIR Measurement and Filtering/Convolution.
(c) Synthesized Cues by HRTF Filtering.

Figure 9: HRTF Measurement and 3D Sound Synthesis (Wenzel, 1992).

analytic or synthetic. Analytic listening concentrates on a certain part and can also be described as an active listening, while synthetic listening describes the perception of *the whole* and can be defined as a passive listening (Williams). An example is the well known *Cocktail Party Effect*, in which one notifies the own name in a different conversation, even though one was not paying direct attention. A similar example that is based on psychoacoustics is the MP3 music format. It compresses music using perceptual models to discard and reduce parts and pieces of information that are less or not audible by human listeners.

Scale Illusion (Headphones).

Phantom Words (Stereo Speakers).

Similar to optical illusions, also auditory illusions exist and have to be carefully studied in order to avoid, or eventually exploit these illusions within information and data sonification techniques. Analog to visual perception, this is also known as *Auditory Gestalt* and is further discussed in Section 4.2. Besides the more familiar Doppler effect, several other auditory illusions and effects exist (Deutsch, 1995/2003; New Scientist, 2008). Examples are an auditory masking effects, but also auditory grouping and scale illusions, listen to the sound examples on the right (Deutsch, 1995/2003). In all cases, the brain adds *missing information* and interprets the sounds differently. Therefore, these illusions have to be accounted for in the design of sonification techniques and in the development of auditory icons and earcons (Kramer, 1994).

Binaural Beats Demo (Headphones).

An interesting side note is the area of *brain entrainment*, in which binaural beats are used to deliberately induce auditory processing artifacts to create a physical stimuli within the brain (Lane et al., 1998). Brain entrainment is said to help *synchronize the two hemispheres of the brain* using binaural beats. Listen to the example on the right for a short relaxation. The effect was originally discovered in 1839 by Heinrich Dove, who described interference beats created by two tones played separately with slightly different frequencies applied to each ear (Dove et al., 1842).

3D Sound Localization and Environmental Perception

The acoustically most interesting point in the development of 3D virtual auditory environments is the human capability of deriving directional and environmental information from perceived sounds (Goldstein, 2007; Vorländer, 2007). Humans perceive relatively precisely the location and the distance of a sound source, as well as information about its local surroundings. The smallest audible angle is in the front around the median plane. Depending on the frequency range, this angle varies between one minute to one degree, but increases away from the median plane towards the back of the head (Warren, 1999). The perception of environmental information is facilitated through echo and reverberation effects, which not only allow a classification of the environment, but also to derive information about the materials and the location of obstacles and walls (Goldstein, 2007; Everest, 2001).

3D Sound Demonstration.

The perception of *3D sound* is based on three listening cues: *temporal*, *intensity* and *spectral differences* that are perceived differently in each ear (Goldstein, 2007; Everest, 1994; Vorländer, 2007). As sound propagates from the source to the listener's ear, it is changed and transformed by the lister's body, head and outer ear (pinna). These effects can be described and measured using a so called Head-related Transfer Function (HRTF). A measured HRTF enables one to synthesize a binaural signal from a monaural sound to recreate a *virtual* 3D sound source. Figure 9 visualizes the principles of HRTF measurement and the synthesis of 3D sound. HRTFs are a function of distance, direction and frequency, and, unfortunately, strongly listener dependent. Available generalized HRTFs often cause ambiguities in source localization, especially with elevation and front/back confusions (Hofman et al., 1998; Richardson and Kaiwi, 2002). The acoustic community therefore researches for methods and techniques to individualize HRTFs for a more precise and listener-dependent 3D sound synthesis. Personalized HRTFs would be a great help in many situations, especially in the areas of audio-only and augmented audio reality application (Röber and Masuch, 2005b; Röber et al., 2006a).

3D Acoustics Demonstration.

Besides directional information of sound source locations, the listener also receives environmental clues that allow him to evaluate the local environment and the material it is composed off. Figure 8 shows a comparison of a visual and an auditory scene representation of the 3D adventure game *Syberia* (Microïds, 2003). The auditory scene representation (Figure 8b) contains a description of what information must be audible in order to receive sufficient information to perform navigation and orientation tasks. The sonification of this scene contains additional (artificial) auditory elements that aid to the user's orientation, exploration and navigation, see also Chapter 5. The included environmental details provide the necessary references to determine the listener's position, as well as to localize the 3D sound sources.

The advantages of using a spatialized acoustic representation are an enhanced situational awareness in a 3D scene that not only improves the understanding of the environment, but also enhances the interaction with the objects therein. Together with other modalities, eg. in an audio/visual representation such as in a 3D computer game, 3D sound reinforces the perceived visual information and delivers additional clues about the the game:

> *"Immersive audio means better gameplay. (...) aural cues are just as important as visual ones." (CNET.COM, 2005)*

For the sonification of 2D and 3D data sets, spatialized sounds enhance the stream segregation and allow a finer separation of data points, eg. to increase the resolution of the auditory display (Kramer, 1994).

3.2.3 *Speech Perception*

Sonification and auditory displays operate from their definition on the basis of non-speech sounds. Speech is, on the other hand, a very direct way of communicating content and ideas, and in places difficult to substitute by other acoustic primitives. Some parts in

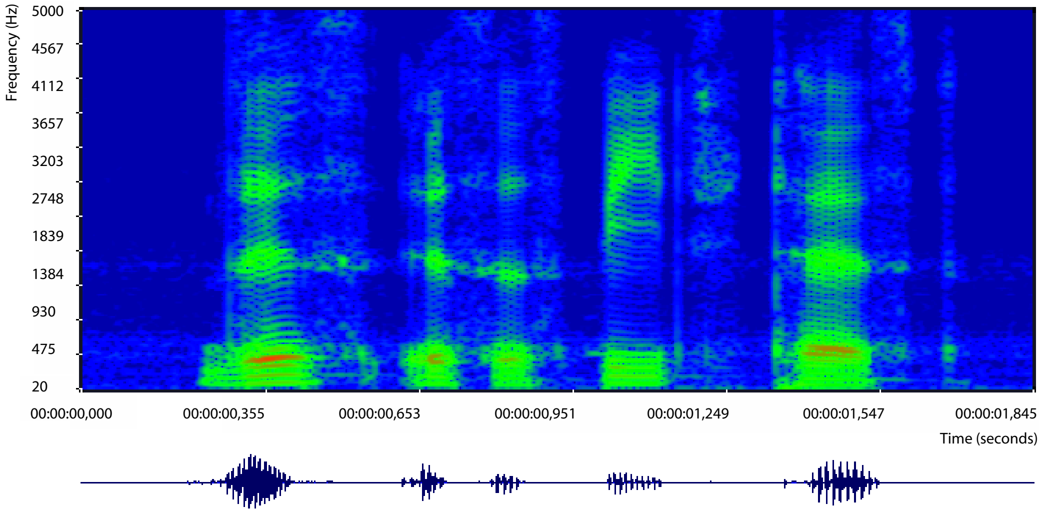


Figure 10: Speech Spectrogram *"This is a Speech Test!"*.

this research explore the possibilities of integrating speech recognition and synthesis into the display of virtual auditory environments. As both, speech recognition and synthesis, still have its difficulties, this section introduces the general concepts of human speech perception, as well as provides a short overview of the current technical possibilities.

The perception of speech is based on the perception of acoustic cues. These cues differentiate speech sounds and classify them into several phonetic categories. *Phonemes* are the smallest units of sound and language, and describe small speech fragments, such as /a/ or /t/ (Goldstein, 2007). The perception of these phonemes is based on formants, which can be described as peaks in the frequency spectrum and are caused by acoustic resonance effects. An example can be seen in Figure 10, which shows the spectrogram of the sentence *"This is a Speech Test!"*. Modern speech recognition and also speech synthesis applications are based on these formants and phonemes. As the voice of each speaker varies not only in pitch, but also in the pronunciation of certain words, these systems still have to be trained by the speakers voice which makes an easy and straightforward application still difficult (Wendemuth et al., 2004).

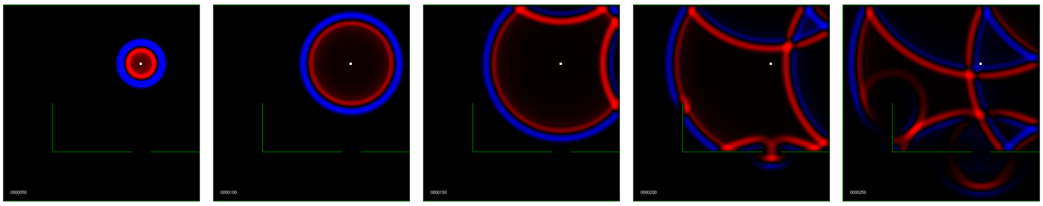
Speech Synthesis Examples.

Speech recognition and synthesis are currently employed in several scenarios, and as research progresses, future applications and interfaces may entirely be based on speech. Common applications that rely on a speech interface include computer games – not only – for the visually impaired (Malyszczyk and Mewes, 2005; Kehoe and Pitt, 2006; Atkinson and Gucukoglu, 2008) and professional dictation software[3].

3.3 Sound Rendering and Physical Acoustics

With the fundamentals of psychoacoustic perception discussed in the last section, the reproduction, synthesis and *rendering* of virtual 3D sound sources moves into focus. The techniques therefor are as important for auditory displays and 3D auditory environments as computer graphics and rendering are for 3D visualizations and visual displays. This section, however, only lists the basic and most fundamental concepts, but provides additional references for further discussions (Smith, 1997; Zölzer, 2002; Rocchesso, 2003).

3 Dragon Naturally Speaking http://www.nuance.com/naturallyspeaking/



(a) Wave at $t = 50$. (b) Wave at $t = 100$. (c) Wave at $t = 150$. (d) Wave at $t = 200$. (e) Wave at $t = 250$.

Figure 11: Propagation of Sound Waves.

Important for an effective display of 3D virtual auditory environments are techniques for 3D sound spatialization and for the simulation of environmental/room acoustics. As these techniques are – with limitations – already available in existing APIs, the following sections provide and discuss implementation details, but also reference additional approaches that permit a higher quality and more accurate 3D sound rendering and acoustic simulation, refer also to Chapter 8.

3.3.1 *Physical Acoustics*

Sound Wave Propagation.

Sound is the propagation of pressure variations that travel through a participating media, and characterized by physical parameters such as frequency, wavelength, period, amplitude, intensity, speed and the direction of propagation. Frequency f describes the cycles per second (Hz) and refers also to the wavelength $\lambda = \frac{1}{f}$. The wavelength of sounds that are perceivable by humans ranges from $\approx$ 17m at 20Hz down to $\approx$ 1.7cm at 20kHz. The speed of propagation is dependent on the participating media and averages in air at room temperature to $c \approx 343\frac{m}{s}$. As sound waves propagate through a room, they interact with the medium of transportation, as well as with obstacles and objects within. Sound energy is thereby partially absorbed, reflected, refracted, diffracted and scattered (Kuttruff, 2000; Zölzer, 2002). Figure 11 visualizes the propagation of sound waves in 2D (single sine pulse) and exhibits several wave-based propagation effects, such as reflections, interference and diffraction.

A physically correct simulation of sound wave propagation along with object interactions is a very difficult and computationally intensive task, and an area with a high priority of research (Zölzer, 2002). Although the sound spatialization and acoustic simulation techniques used for the display of 3D auditory environments are required to be of high accuracy, a physically correct model is not necessary. On the contrary, several effects have to be enhanced (for instance the Doppler), while others can be omitted. Section 5.2 continues this discussion in the design of display techniques that focus on a perceptual, rather a physically correct model.

3.3.2 *Sound Signal Processing*

Digital sound signal processing is concerned with the filtering and modification of digital sound signals and defines a fundamental basis for this research. It spans from the acquisition of sound data (sampling), over filtering and processing, towards the synthesis of digital sound signals and sound playback. Figure 12 visualizes this pipeline using a filter that simply reduces the input signals amplitude by $\frac{1}{2}$. The four diagrams at the bottom show the sampling of the data with a continuous representation for the analog

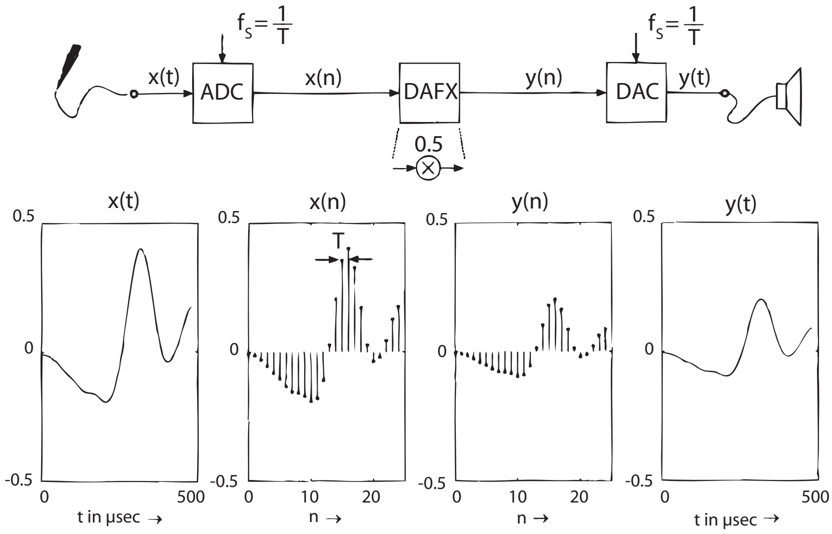


Figure 12: Sound Signal Processing (Zölzer, 2002).

parts (left and right), and a discrete representation through the digital signal processing (middle) (Smith, 1997; Zölzer, 2002; Rocchesso, 2003).

Most of the signal processing that is required for this research, eg. for 3D sound rendering and filtering, is already implemented and available within APIs such as OpenAL/EFX, Microsoft DirectX Sound, AM3D and FMOD (Boer, 2002b; Firelight Technologies Pty, Ltd, 2001-2008; AM3D A/S, 2008). However, during the research for this thesis and the evaluation of the developed prototypes, an insufficiency in the existing sound APIs has emerged. The implementations therein are partially based on very simple, if not crude, approximations of the real-world acoustics (Hiebert, 2006; Peacock et al., 2006).

3.3.3 *Sound Rendering and Audio Hardware*

The PC was originally designed as a machine for office work, and the demands for sound rendering and the playback of music were rather small. The computer was equipped with a small speaker that could only produce beeps of different length and frequency. But with the introduction of the first computer games, the requirements for quality audio increased. The first PC sound cards available were developed by AdLib Inc. in 1987 and by Creative Labs in 1989 (Zander, 1995). These addon cards dramatically extended the PCs acoustic capabilities and supported a development in the direction of a multimedia system that was used for computer gaming, as well as for sound editing and music production (Zander, 1995).

Current PC sound hardware is available in a large variety, and ranges from sound cards for computer games to professional audio hardware for music production. Some of this hardware is even equipped with a programmable DSP[4], but not to the extend of free programmable graphics hardware. Today's sound cards are capable of a multi-channel sound output, sound spatialization using on-board HRTF filters, MIDI output for music synthesis, as well as possess large capacities for sound signal processing. However, in respect to realistic 3D sound spatializations and simulations of environmental acoustics, large deficiencies still exist. Creative's flagship with the *XFi* processor at its core still

4 DSP – Digital Signal Processor

uses the same approximations and HRTF filters as its predecessor (Creative Labs, 2005; Ramelet, 2000b). In 1999/2000 Aureal Semiconductor designed the *SQ3500* and developed several ingenious concepts, including a technique called *Wavetracing* that permitted a much more realistic simulation of room acoustics using real room geometry (Aureal, 2000; Ramelet, 2000a). Unfortunately, Aureal did not endure a patent related lawsuit, leaving Creative Labs as the monopolist in this sector (Gasior, 1999).

Dedicated and partially programable DSP sound hardware already exists, but is, however, limited and only available to music artists and composers. Available PC sound hardware is still rather fixed in its functionality and signal processing pipeline. An onboard and free programmable DSP pipeline would clearly enhance the current possibilities for sound rendering and acoustic simulation. In a near future, a similar process as with the evolution of programmable graphics hardware might be observed with the development of PC sound hardware.

Implementation

Several libraries are available for an implementation of virtual sound and acoustics (Boer, 2002b; Firelight Technologies Pty, Ltd, 2001-2008; AM3D A/S, 2008; Hiebert, 2006; SDL, 2008). Most of these APIs support the rendering of 3D sound sources, while only a few are able to emulate room acoustics as well. An audio API that is employed in many com-

```
#include"al.h"
#include"alc.h"

Device = alcOpenDevice(NULL);                           // Initialization
if (Device) {
    Context = alcCreateContext(Device, NULL);
    alcMakeContextCurrent(Context);
}

alGenBuffers(NUM_BUFFERS, g_Buffers);                   // Generate Buffer
loadWAVFile("test.wav", &format,&data,&size,&freq,&loop);  // Load test.wav
alBufferData(g_Buffers[0], format, data, size, freq);   // Copy Data

alGenSources(1, source);                                // Generate Source
alSourcei(source[0], AL_BUFFER, g_Buffers[0]);          // Attach Buffer

ALfloat source0Position[]={ 2.0, 0.0,-2.0};             // Source Position
ALfloat source0Velocity[]={ 2.0, 0.0,-2.0};             // Source Velocity

alSourcef(source[0], AL_PITCH, 1.0f);                   // Adjust Parameters
alSourcef(source[0], AL_GAIN, 1.0f);
alSourcefv(source[0], AL_POSITION, source0Position);
alSourcefv(source[0], AL_VELOCITY, source0Velocity);
alSourcei(source[0], AL_LOOPING, AL_FALSE);

alListenerfv(AL_POSITION, listenerPosition);            // Listener Position
alListenerfv(AL_VELOCITY, listenerVelocity);            // Listener Velocity
alListenerfv(AL_ORIENTATION, listenerOrientation);      // Listener Orientation

alSourcePlay(source[0]);                                // Start Playback
```

Listing 3.1: Basic OpenAL Example.

puter games and multimedia applications is OpenAL (OpenAL, 2008). OpenAL, although a free library, is supported and developed by Creative Labs, and able to spatialize monaural sound sources, as well as to emulate room acoustics using low-pass filtering and mixing techniques (Hiebert, 2006; Peacock et al., 2006).

Listing 3.1 shows a minimalist example for creating 3D sound sources using OpenAL (Hiebert, 2006; Peacock et al., 2006). The example code of other libraries is very similar and shows a strong relation to OpenGL and other graphics APIs. This makes OpenAL not only very easy and convenient to use, but also allows the utilization of a visual scenegraph system for scene and contend management. The example in Listing 3.1 shows on line 4 the initialization of OpenAL, and on lines 10-15 the creation of a 3D sound source. The source is assigned a position along other parameters in lines 17-24 and the playback is started on line 30. Despite the discussed limitations, OpenAL was used in many examples and prototypes in this research. However, additional research was also conducted to improve the rendering of 3D sounds and the simulation of room acoustics.

3.4 AUDIO IN ENTERTAINMENT COMPUTING

Apart the abstract discussions on implementation details in the last section, this part describes the employment of audio in entertainment computing, and thereby directly concentrates on the application and use of sound and music. While the first computer games were played solely without sound, soon audio/visual computer games were developed for arcade game machines, consoles and the Amiga and Commodore home computers. Later, with the introduction of PC sound hardware, also the PC game market accelerated (Zander, 1995; Boer, 2002b). However, for several years the importance of sound and acoustics stood in the shadow of fast evolving computer graphics, until a few years ago sound hardware also emerged into the 3D realm (Ramelet, 2000a,b; Boer, 2002b). Today, game players and developer are both aware of the importance of sound. George Lukas, best known to the public as a movie director and creator of stunning visual effects, stated once that:

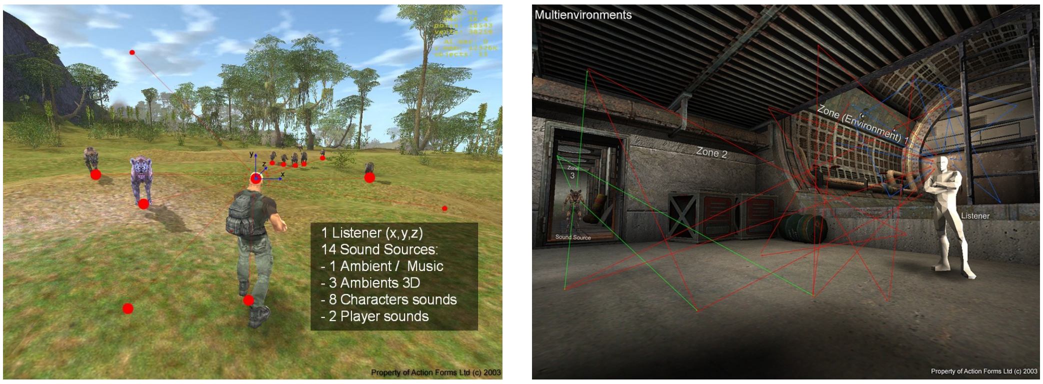
> *"...sound is 50% of a movie experience." (THX Consortium, 2000)*

The same can be applied to computer games and other forms of interactive narration. The music of several of today's games is specially created by well known artists and composers and released as additional soundtrack on CDs and DVDs.

Sound and music are both often used to create ambience and to express emotions, a quality that virtual environments, such as computer games, often lack. Sounds are employed to describe objects, the environment and actions, while music is used to trigger emotions and to influence the play of the game (Boer, 2002b; Kiegler and Moffat, 2006). Examples are *Driver76* (Sumo Digital, 2007), in which the music puts the player back into the seventies, or *Silent Hill Origins* (Konami, 2007) and *RidgeRacer2* (Namco, 2007), where music induces fear in a survival-horror game and daringness in a car racing simulation.

Silent Hill Origins Trailer.

Important for the perception of sound and music is also the presentation, whether it is perceived over headphones, a small speaker array or a multi-channel surround sound system. All systems have their respective advantages and drawbacks and depending on the listening task one might chose either one. However, for the perception of 3D sounds and 3D virtual auditory environments, headphones in combination with a binaural sound rendering are most favorable. This approach eliminates interferences of the listening environment (eg. the room's acoustics and noises), as well as enhances the perception of 3D sound sources and the virtual acoustics. The most realistic presentation is possible using a technique called wavefield synthesis, in which, according to Huygens principle,



(a) 3D Sound, Music and Sound Effects.   (b) Room Acoustic Simulations.

Figure 13: Sound and Effects in a 3D Computer Game (Menshikov, 2003).

several sound waves from large speaker arrays are superimposed to create spherical sound waves of virtual sound sources, within or outside this speaker array (Boone, 2001). Due to the large number of speakers that are required, this technique is not yet suitable for the mass market. However, new approaches that utilize this technique are already integrated into the OpenAL library, and employ it for a hyper-realistic auditory gameplay (Gräfe et al., 2007).

With the grown awareness of acoustics over the last years, also the acoustic design has changed and moved towards a *cinematic sound in computer games* (Hämäläinen, 2003). Figure 13a shows an overview of the utilization of sound and music in current 3D games, while Figure 13b visualizes the use of multi-environments for a more realistic room acoustics simulation using OpenAL's EFX engine (Menshikov, 2003; Hiebert, 2006; Peacock et al., 2006). The specification of distance models thereby allows the modeling of an acoustic depth cuing, while the reverberation settings for the various environments enable an acoustic simulation of different rooms, from free field to a small stone cave.

3.4.1 *Music-centered Computer Games*

Besides the classic 3D audio/visual computer games exists a large variety of niche products. Some of them almost entirely focus on music and an interaction design based on harmonies and rhythms. The concept of combining music with an interactive gameplay dates back into the year 1787 and was introduced by Mozart in his *Musikalisches Würfelspiel* (Mozart, 1787). The gameplay of today's music games often centers around the replay and/or the extension and enhancement of given beats and rhythms. The player thereby creates interactive music that instantaneously provides a feedback of the player's performance. The interaction and gameplay is often tied into puzzles, in which the user has to activate controls timed by the games music/soundtrack.

Music games either concentrate on the simulation of musical instruments, or allow the player as much freedom as possible during the *composition* process. Such games are often bundled with additional controllers that substitute an instrument, or the part that is simulated. Figure 14 shows several examples. *Guitar Hero* and *Samba de Amigo* employ a toy guitar and rattles, which both have to be played or shaken in rhythm with the music presented (Harmonix Music Systems, 2005; Sega, 2000). A highscore and the music generated provide the user with a performance feedback. *PaRappa the Rapper, Sing Star* and *Dance Dance Revolution* were initially developed as extensions for the Playstation and

(a) REZ (United Game Artists, 2001). (b) Guitar Hero (Harmonix Music Systems, 2005). (c) Metronome (Tarsier Studios, 2008).

Figure 14: Music-centered Computer Games

the Playstation2, but are nowadays also present on almost every other game platform available (NaNaOn-Sha, 1996; Sony Entertainment, 2004; Konami, 2001). These games simulate the player's performance as a singer and dancer, and therefore the ability to follow a given beat and to stay in sync with the games soundtrack.

REZ Demo Video.

Besides these simulations, other games focus more on the compositional part and the creation of music by the user. Such games often exhibit a high degree of immersion and dissolve the player in a state of trance. *REZ* is here the most well known implementation and classified as a rail shooter that generates hypnotic graphics and trance-like music (United Game Artists, 2001). Adaptations to this concept are *Lumines*, *Every Extend Extra*, *flOw* and *Electroplankton*, which all feature immersive graphics and sound effects, and let the player compose his personal piece of music while flying through a world of sound (Q Entertainment, 2004; Buena Vista Games, 2006; thatgamecompany, 2008; Indies Zero, 2005). The not yet released game *Metronome* has an alternative, yet very intriguing game approach. The game itself is an auditory action adventure, in which the player has to search a city for specific sounds that the player has to record and play back at other locations to fight enemies and to unlock hidden secrets (Tarsier Studios, 2008).

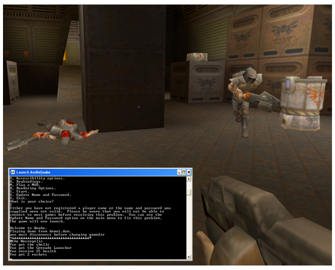
The City of Metronome.

3.4.2 *Audio-only Computer Games*

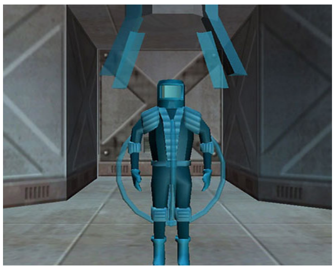
A genre of its own are so called *Audio-only Computer Games*, which are played and perceived primarily using sound and acoustics. The gameplay is in most cases very similar to audio/visual games, with the difference that all feedback information is conveyed through auditory icons and earcons. As in audio/visual games, one is looking for certain patterns to emerge, eg. listening for certain earcons and auditory icons, to which the player reacts accordingly.

The genre started with the development of accessible games for the visually impaired, in which game titles were created by small companies and blind programmers (Andresen, 2002; Mischke and Scardovelli, 2005; van Tol and Huiberts, 2006). The majority of these games is still developed for the Windows platform, although some titles are now also available on consoles and mobile platforms. Many genre from the audio/visual domain have been adopted, including adventures, simulations, racing and role-playing games (van Tol and Huiberts, 2006). Over the last years, the genre has enjoyed an increase in popularity and many new titles with interesting approaches have been released. Dedicated game engines that explicitly focus on the development of audiogames have been designed and further foster the evolution of this still very young genre (Bartiméus Accessibility Foundation, 2008).

Terraformers Demo.



(a) AudioQuake (Atkinson and Gucukoglu, 2008).



(b) Terraformer (Pin Interactive, 2003).

(c) Seuss Crane (Destiny Media, 1999).

Figure 15: Audio- and auditory Adventure Games

Shades of Doom Demo.

Seuss Crane Detective for Hire.

Even though some of the old PC text adventures in combination with a speech synthesis can also be considered as audiogame, they are in this discussion ignored because they were not initially designed with an auditory gameplay and an acoustic presentation in mind. One of the first audiogames developed was *Real Sound: Kaze no Regret* by the Japanese company WARP (Warp, 1999). WARP was founded by the Japanese musician Kenji Eno, who developed the game as a homage to his blind fans Milestones in the genre of audiogames are the 2001 released *Shades of Doom* (GMA Games, 2001) and the 2003 released *Terraformer* (Pin Interactive, 2003), see also Figure 15b. Terraformer also provides a visual interface and therefore represents a so called hybrid game, which allows a play of sighted and visually impaired users together. Other examples in Figure 15 show the 1999 released web-based adventure game *SeussCrane: Detective for Hire* and the 2008 released accessible version of *Quake* (Destiny Media, 1999; Atkinson and Gucukoglu, 2008). New trends for an audio-centered gameplay point in the direction of augmented audio reality, in which a real environment is enhanced through artificial sounds for play and entertainment (Cohen et al., 2004).

3.4.3 *Audiobooks and Radio Plays*

Jules Verne's "Journey to the Centre of the Earth".

Hans Christian Andersen's "The Nightingale".

Audiobooks and radio plays can not directly be compared with computer games, but possess several characteristics that makes them very interesting. They allow the creation of a very special form of interactive narration that is comparable to auditory adventures and action games. Section 9.6 will revert to this discussion with the introduction of *Interactive Audiobooks*.

Audiobooks are narrated books that are recited by a single person, while radio plays are based on several actors that are playing the story, and is further enhanced by music and additional sound effects. Over the last years, audiobooks and radio plays have enjoyed a constant increase in popularity, which is mainly due to their easy use and the high level of narrative immersion (Fey, 2003; audiobooks.com, 2008). A combination of the narrative advantages with interactive elements from adventure-based computer games allows the creation of highly immersive and possibly non-linear storylines. An early example can be seen in Figure 15c, which shows a web-based radio play in which the listener chooses the next location of the story, thereby solving a crime mystery (Destiny Media, 1999).

With the growing awareness of the potential of 3D sound spatialization, this technology was more recently applied to 3D audiobooks and radio plays as well, and allows the listener a stronger immersion and involvement (Verne, 2005; Andersen, 2005).

Computer Science is no more about computers than astronomy is about telescopes. — E. W. Dijkstra

4 AUDITORY DISPLAYS

AUDITORY DISPLAYS are the primary area of related work in this research. They can be considered of being the auditory analog to a visual monitor and are utilized as system to *display* abstract data using auditory means. Auditory displays are employed in a variety of applications, ranging from computers and medical workstations, to information displays in aircraft cockpits and control centers in nuclear power plants. The main difference to the more commonly employed visual display is that the information is perceived solely through the sense of hearing. This opens many advantages and possibilities for the design and application of this technology, but also inflicts difficulties caused by an audio-only presentation. The goal of this chapter is to provide an overview of the techniques developed in this area and to examine their strengths and weaknesses for a later application within 3D virtual auditory environments.

A large portion of this research is directly based on 3D spatial auditory displays, therefore the first section of this chapter serves as an entry point and introduces the most fundamental concepts and techniques. Available possibilities for an audio-centered data mapping are examined in the following section, which also discusses auditory Gestalt principles, the analogic/symbolic continuum, as well as spatial sonification and data representation techniques. The last section presents several examples and areas of applications, and focusses on issues of construction, authoring and design.

4.1 OVERVIEW AND DEFINITION

The research on auditory display systems and sonification techniques accelerated with the introduction of the ICAD conference, which was held for the first time in 1992 (Kramer, 1994). However, the first publications that were concerned with an application of sound and acoustics for the display of abstract data were already published in 1954 and 1961 (Pollack and Ficks, 1954; Speeth, 1961). The article by Pollack and Ficks described a quantitative display through the use of several auditory variables. Employing tone and noise bursts, this first auditory display was able to represent eight binary variables through the use of pitch, loudness, temporal ratio, duration and stereo location (Pollack and Ficks, 1954). A second article published by Speeth in 1961 employed a technique called *audification* to sonify seismic data (Speeth, 1961). An accelerated playback of the recorded seismic waves greatly improved their analysis and understanding.

The term *Auditory Display* has not yet been conclusively defined, but was coined and influenced by several ICAD conferences and the foundation of a research community:

> *"Auditory Display research applies the ways we use sound in everyday life to the human/machine interface and extends these via technology. The function of an auditory display is to help a user monitor and comprehend whatever it is that the sound output represents." (Kramer, 1994; ICAD Community, 1992 - 2008)*

A second and very similar definition was provided later by Hermann in 2006:

> *"Auditory Displays are systems where a human user makes sense of data using his/her listening skills, like for instance any data under analysis or data that represent states of the information processing system." (Hermann, 2006)*

Both definitions concentrate on the user's ability to identify and interpret auditory signals and the development of audio-based human/machine interfaces through the sonification of arbitrary (scientific) data. The research is founded in an interaction with sound in real-world scenarios and strives to design information processing systems that map abstract data values onto auditory variables, which in turn can be interpreted and analyzed by a human listener.

The research of auditory displays can also benefit from an analysis and observation of the interaction with visual systems. Although visual and auditory displays seem on a first glance to be very distinct, yet they share many concepts and principles. The primary difference is that an auditory perception is linear in time, ie. (3D) sounds exist in time but are perceived over space, while a visual perception is linear in space, ie. visual objects exist in space but are perceived over time. Despite these differences in perception, the interpretation and analysis of data and information is very similar. The basic principles inherent in a *good* visualization can directly be applied to an acoustic presentation of information using an auditory display as well (Schumann and Müller, 2000).

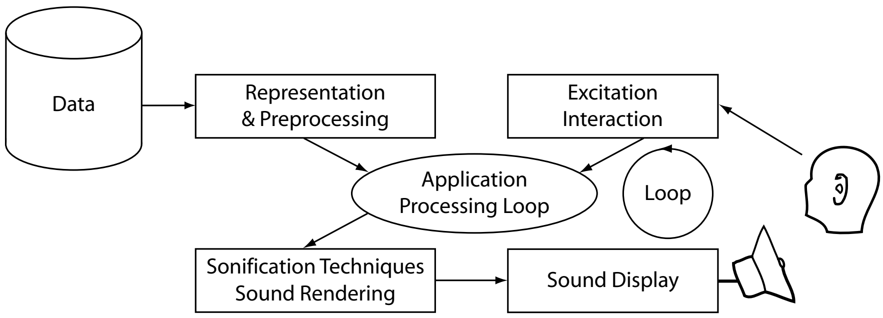


Figure 16: Auditory Display Interaction Loop (Hermann, 2006).

An overview of the principles characterizing an auditory display system can be seen in Figure 16 (Hermann, 2006). The overall layout is very similar to the visualization pipeline as presented in Figure 4. Based on the input of raw data, first a pre-processing is required to map acoustic primitives to certain aspects within the data. The user interacts with the system using an interaction/feedback loop to present specific parts of the data, or to focus on certain details. The information is then sonified and presented acoustically using the system's effectors. A single or stereo speaker setup is often sufficient, while 3D auditory displays require a presentation via headphones or the use of a large multi-channel speaker array, such as (5.1) surround sound systems (Boer, 2002b). These systems often employ cross-talk cancelation for a binaural display, as well as channel-blending techniques to improve the localization of virtual sound sources.

Key components of an auditory display framework are the sonification and interaction techniques used. Sonification can be thought of as being the auditory analog to graphics-based visualization and is defined by Kramer et al. and Hermann similarly as:

> *"Sonification is the use of sound - mainly non-speech audio signals - for representing or displaying data. Similar to scientific visualization, sonification aims at enabling human listeners to make use of their highly-developed perceptual skills (in this case listening skills) for making sense of the data. More specifically, sonification refers to the technique used to create a sound signal in a systematic, well defined way, that involves data under examination as an essential ingredient for sound computing." (Kramer et al., 1997; Hermann, 2006)*

Benefits of Auditory Displays:	Difficulties with Auditory Displays:
• Eyes free application	• Low resolution of variables
• Rapid detection	• Limited spatial precision
• Alerting	• Lack of absolute values
• Orienting	• Lack of orthogonality
• Backgrounding of tasks	• Annoyance
• Parallel listening	• Interference with speech
• High dimensionality	• Not bound by line-of-sight
• High temporal resolution	• Absence of persistence
• Affective response	• No printouts
• Auditory Gestalt formation	• User limitations

Table 1: Benefits and Difficulties for using Auditory Display Systems (Kramer, 1994).

Both acknowledge the connection between sonification and interaction to define an active sonification technique, in which the user explores specific parts of the data or information, which is in turn sonified – eg. acoustically represented (Hermann, 2002). A selection of adequate techniques for the sonification of data is one of the critical aspects in the design of an auditory display system. It includes the mapping of data attributes onto auditory primitives, which is highly dependent on the data used and the area of application. Generally, auditory displays are very applicable for monitoring tasks and for the exploration and analysis of linear data sets. Monitoring tasks are applicable, as one is listening for certain patterns to emerge that acoustically *pop out* from the auditory display (Smith et al., 1992). This exhibits one of the main advantages of auditory displays, as monitoring, or template matching tasks, can easily be performed in the background with other – possibly visually engaged – tasks at full attention. Data exploration and analysis, however, can not be backgrounded and requires the undivided attention of the listener. In these cases, one explores data sets and information, searching for certain characteristics. Although this approach requires a template matching approach as well, yet one can not precisely anticipate what will be heard.

Table 1 shows an overview of the advantages, as well as the difficulties for using auditory display systems (Kramer, 1994; Shilling and Shinn-Cunningham, 2002). One of the most important benefits is the non-necessity of paying direct attention, but also the parallel listening and the high dimensionality / temporal resolution that can be achieved are significant. Interesting to note is that the resources required for the development of an auditory display are far below the requirements for an equally designed graphics-based visualization system. Major difficulties for the use of auditory displays are a low precision, as well as the lack of absolute values. The majority of auditory parameters are not perceptually independent. This lack of orthogonality can influence the perception of a certain parameter if a second one is altered. Auditory displays are sometimes combined with a graphics-based visualization system. Such multivariate and acoustically enhanced displays generally improve the perception and understanding of the data presented through an intermodal correlation that results in a higher realism and increases the efficiency of the system.

4.2 FUNDAMENTALS AND PRINCIPLES

After this short introduction of auditory displays, the remaining sections in this chapter further explore the fundamentals in the design, as well as exemplarily discuss areas of applications. The following section continues the discussions of Section 3.2 and introduces important data and information sonification techniques. Although the section aims at a broad presentation, a slight focus is found in the discussion of spatial auditory displays and 3D sonification and interaction techniques.

4.2.1 *Auditory Gestalt and Presentation*

Examples for auditory Gestalt.

The concept of *Gestalt* refers to a unified interpretation of sensory information that is perceived in the form of patterns and interpreted in a way that the sum is greater than its parts. The original concept is based on the research of Mach and von Ehrenfels, and although it originates in the discussions of audio and music perception, the concept of Gestalt is most commonly known for its visual examples (Mach, 1886; von Ehrenfels, 1890; Goldstein, 2007). With the introduction of computer-aided sound synthesis and the development of auditory displays, Gestalt theory becomes applicable again to the perception of sound and music (Moore, 1989; Bregman, 1990; Williams). Through the perception of auditory (respective visual) information, the data perceived is analyzed, ordered and grouped regarding seven basic principles: *similarity, proximity, good continuation, habitat or familiarity, belongingness, common fate* and *closure* (Bregman, 1990; Purwins et al., 2000).

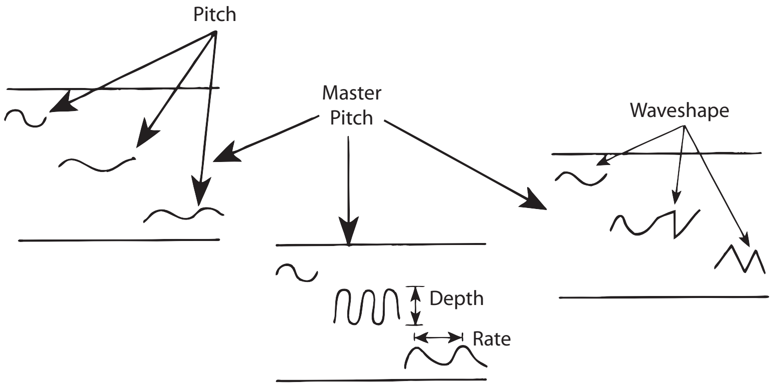
All seven principles are based on a certain grouping and classification of the information perceived. *Proximity*, for instance, refers to an auditory distance of features that groups objects with a smaller distance closer together and separates them from other elements. *Similarity* is very much alike proximity, but refers here to the inherent qualities of a sound object. The principle of *Continuation* extends proximity over time and classifies smoothly varying frequencies of a changing sound source. Sudden changes in the perception are identified as a new source. *Closure* completes perceived fragments features within the *Good Gestalt* principle, while *Common Fate* groups frequency components, that possess a similar sound or in which parameter changes occur synchronously.

The perception of Gestalt information allows one to discern overall relationships and trends within the presented data and enables a listener to pick out meaningful events within several data streams (Bregman, 1990; Purwins et al., 2000). This is a very important feature for the design of sonification techniques and important for the mapping of data onto auditory primitives.

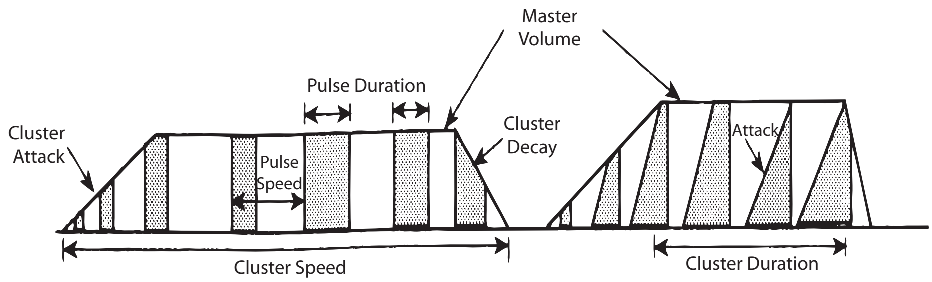
Principles of organizing sounds and the delivery of specific information through sound and sonification have also been discussed by Ballas and Kramer (Ballas, 1992; Kramer, 1992). Ballas provides an initial overview of how sound can be used to deliver information using a framework of linguistic analogies (Ballas, 1992). He identifies the main challenges for a sonification design as:

> *"to capitalize on the extensive understanding of sound production and perception in the fields of language, music and acoustics to invent new sounds that communicate from abstract sources and domains to a listener who has a complex but fixed receiving and interpretation capability."* (Ballas, 1992)

Kramer focusses more on the concepts of auditory Gestalt and devises sound organizing principles using the nesting of loudness and pitch parameters (Kramer, 1992), refer also to Figure 17. A nesting of parameters allows the control of a single auditory variable over several time scales simultaneously, thus enables the display of high-dimensional



(a) Pitch Nesting.



(b) Loudness Nesting.

Figure 17: Parameter Nesting (Kramer, 1992).

data sets. Figure 17 shows examples for the nesting of loudness and pitch, which are composed of sounds of different time scales that differ in loudness and pitch. This allows the creation of complex auditory earcons, similar to the design of 3D glyphs used in a graphics-based 3D visualization, to represent high-dimensional data sets, such as the sonification of multiple stock data sets (Kramer, 1992; Schumann and Müller, 2000).

Sonification of Stock Market Data.

Another topic in auditory perception is attention and the actual process of listening. This is described by the *Figure/Ground Problem* and refers to a selective listening process, in which one concentrates on specific parts of a possibly complex sound. Williams define this as an analytic vs. a synthetic listening:

> *"Synthetic perception takes place when the information presented is interpreted as generally as possibly, as for example, hearing a room full of voices, or listening to the overall effect of a piece of music.*
> *Analytic perception takes place when the information is used to identify the components of the scene to finer levels, for instance, listening to a particular utterance in the crowded room or tracking one instrument in an orchestral piece or identifying the components of a particular musical chord." (Williams)*

This analytic/synthetic perception is very important for the development of auditory displays and the design of sonification techniques. Depending on the sonification goal, either an analytic (concentrating on a part), or a synthetic (concentrating on the whole) presentation is required, which influences not only the overall presentation, but also the interaction with the system.

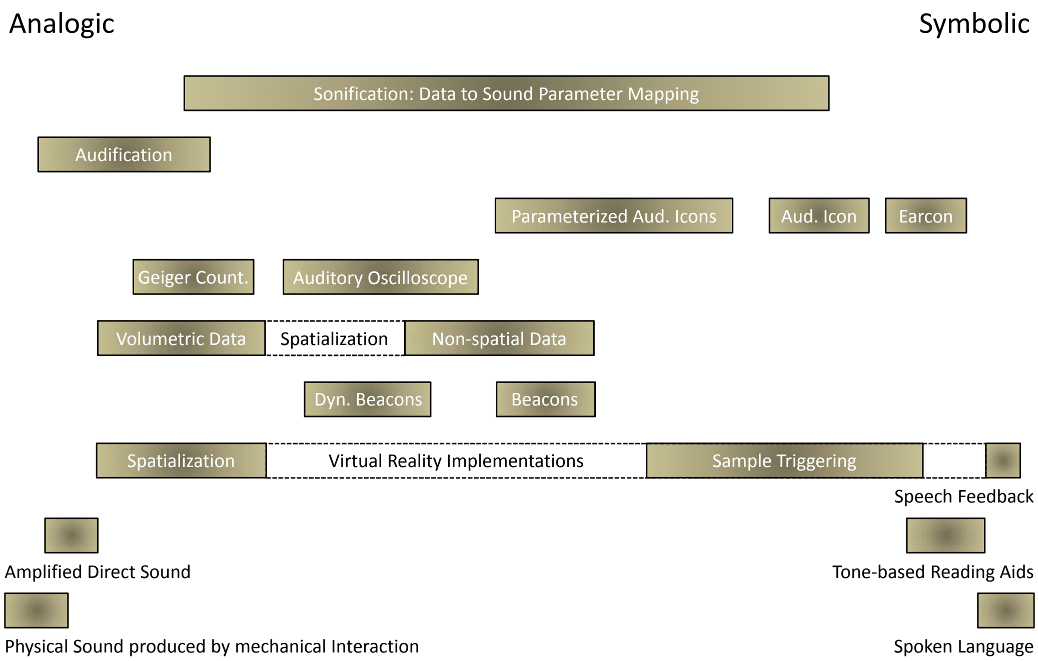


Figure 18: The analogic/symbolic Continuum (Kramer, 1994).

4.2.2 *The Analogic — Symbolic Continuum*

For a better understanding of auditory display systems and a classification of the sonification techniques available, Kramer devised the *Analogic — Symbolic Continuum*, along which all sonification techniques are arranged:

> *"A symbolic representation categorically denotes the thing being represented, while the analogic representation directly displays relationship. An analogic representation is one in which there is an immediate and intrinsic correspondence between the sort of structure being represented and the representation medium. By symbolic representation we refer to those display schemes in which the representation involves an amalgamation of the information represented into discrete elements and the establishment of a relationship between information conveying elements that does not reflect intrinsic relationships between elements of what is being represented." (Kramer, 1994)*

The concept was developed in continuation of the work of Sloman, who discussed analogic representations and their differences regarding a *Fregean*, or symbolic, information representation (Kramer, 1994; Sloman, 1971). Figure 18 provides on overview of the continuum and shows the classification of several sonification techniques. The common Geiger counter, which uses a pulse-based representation to identify radiation rates, is classified as an analogic technique, as it directly corresponds to what is being represented (Rutherford and Geiger, 1908). An acoustic alarm, on the other hand, denotes a possibly complex event or family of events, and is therefore a symbolic representation. The classification of sonification methods in terms of analogic and symbolic becomes later very important for the development of 3D scene sonification techniques to acoustically represent 3D scene and 3D object information.

Figure 18 identifies the position and the range of several important auditory display techniques. *Audification* is one of the most direct representations, in which the data is simply mapped to sound and played back. An example is the work by Heyward, who utilizes this technique for a highly analogic representation of recorded seismograms (Heyward, 1992), refer also to Section 4.3. *Auditory Icons*, which were initially described by Gaver, are very often employed and used in a large variety of sonification tasks. Auditory icons acoustically describe the thing being represented using an auditory caricature of the task (Gaver, 1989; Mynatt, 1992). An example of such an auditory icon is the sound of crumpling paper for the deletion of files and data in a computer trash can. A second example can be seen in Table 2. Cohern employs auditory symbols and familiar sounds to describe and characterize a computer's performance using discrete auditory events (Cohern, 1992).

Event	Message
Login	Knock-Knock
Connection Reminder	"Ahem..."
Low % CPU Time	Slow Walking
Low/Medium % CPU Time	Medium Walking
Medium % CPU Time	Fast Walking
Medium/High % CPU Time	Jogging
High % CPU Time	Running
Logout	Door Slam

Table 2: Event to Message Mapping (Cohern, 1992).

A further extension towards a symbolic representation is the *Earcons* approach that was introduced by Blattner et al.. Earcons possess a language-like characteristic, and are – when learned – very efficient and easy to use to identify very specific items or events (Blattner et al., 1989). Their complexity and design is parallel to visual icons and is composed of basic building blocks to describe complex structures.

An extension to this concept was later introduced by Bölke and Gorny, who established the term *Hearcons* for spatialized auditory icons and earcons. Hearcons are characterized through the four parameters: sound, loudness, position and extent, and were initially employed for an interaction of visually impaired users with a computer system (Bölke and Gorny, 1995).

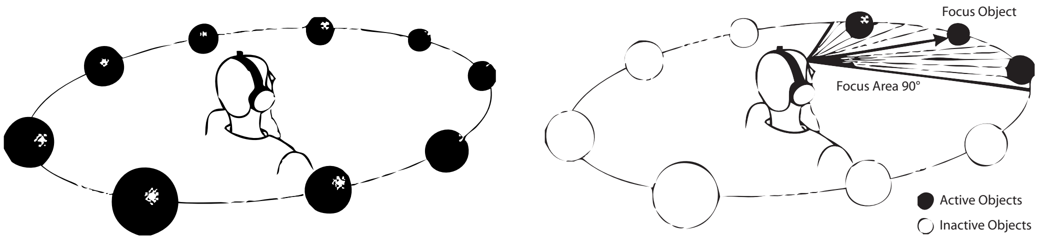
Beacons and dynamic Beacons.

Beacons, and especially dynamic beacons, characterize a more analogic representation of an analog-to-symbol conversion (Kramer, 1992). A beacon employs Gestalt information when a snapshot of the auditory data representation is performed. A dynamic beacon that is learned and intuitively understood, functions as a symbol of the event that it describes.

While some of these techniques directly display the underlying information, such as audification, others utilize auditory symbols that must be learned. Once these symbols are learned and directly understood, a link between the data and their sonic representation has been formed and the analogic display becomes symbolic.

4.2.3 *Spatial Auditory Displays*

With the assistance of sound spatialization, the advantages of a symbolic and an analogic representation can be combined. Similar to the analogic representation of visual objects



(a) The complete Menu. (b) A Selection.

Figure 19: Ring-based Auditory User Interface (Crispien and Fellbaum, 1996).

and symbols in 2D/3D space, sounds can be spatialized and assigned a certain location within a 3D environment, therefore:

> *"spatialized sound can, with limitations, be used to analogically represent three-dimensional volumetric data. The exploitation of this analogy is perhaps the greatest power of spatialized sound, in that it presents spatially indexed data in a way that is highly intuitive and, therefore, instantly understood." (Kramer, 1994)*

Sound spatialization is therefore highly applicable to represent spatial data, such as 3D volumetric data sets, but can also be employed to identify the position of objects in 3D virtual auditory environments. In here, 3D sound spatialization denotes the position of an object, while the sound itself describes its function. A second advantage of sound spatialization is an enhanced stream segregation for a parallel listening approach. This allows, through an analogic/synthetic listening process, the presentation and perception of multiple data streams as one, or a focussed listening to a single source.

Examples for spatial auditory displays include tele-robotic control, aeronautical displays and shuttle launch communications, as well as 3D user interfaces and 3D virtual auditory environments (Wenzel, 1992; Crispien and Fellbaum, 1996; Begault, 1994). An air traffic collision avoidance system (TCAS) discussed by Begault illustrates the symbolic/analogic nature of the display, as it directly correlates *virtual* 3D sound sources with their respective objects in the real environment (Begault, 1994). As of this spatial representation of information, 3D auditory displays require also spatial interactions and a 3D audio-centered user interface (Wenzel, 1992; Crispien and Fellbaum, 1996). Figure 19 shows exemplarily a 3D auditory ring-based menu system that is centered around the user (Crispien and Fellbaum, 1996). Through sound spatialization, the individual menu items are perceived as distinct sound sources, with which one can interact to make selections and changes, refer Figure 19b. Such menu systems can also be nested and layered, eg. extend over different layers in space and time, which allows the design of highly complex systems. Sound spatialization is also often employed in audio-only and audio/visual computer games to improve the user's orientation and to increase the degree of realism (Menshikov, 2003; van Tol and Huiberts, 2006), see also Section 3.4.

Demonstration of the TCAS System.

Sound spatialization is a key component in this research whose importance can not be underestimated. Combined with an intuitive and flexible 3D interaction design, it is employed in the majority of applications and example scenarios. Here Chapter 5, and especially Section 5.3, concentrate on the design of intuitive and comprehensive data and 3D scene sonification techniques using 3D sound spatialization.

After these reviews of the principles of auditory display systems and the fundamentals in auditory perception, the following part presents example implementations and discusses areas of application for auditory display systems.

4.3 AREAS OF APPLICATION

The above discussions already displayed the versatile nature and the many applications for 2D and 3D auditory displays. This concluding section summarizes the main principles and exemplarily evaluates case studies and areas of application. The majority of applications for auditory displays reside in the areas of task monitoring and auditory data analysis and exploration. However, the possibilities of representing abstract data and information solely using the means of sound and acoustics is also well known in the artistic community, which employs sonification techniques to acoustically describe stock market data, to sonify measurements of ocean buoys and to represent soundscapes of urban environments (Gaye et al., 2003; Janata and Childs, 2004; Polli, 2004; Polly, 2003). The example sonifies the top of the atmosphere at the northernmost point of a storm model and indicates strong changes in the winds. These projects are based on a sensory input and measured scientific data sets, which are mapped onto acoustic primitives and sonified for exploration and analysis, but also for pure enjoyment.

Atmospheric Weather/Works.

With the availability of efficient, yet inexpensive mobile platforms, several stationary auditory display systems migrate towards an augmented and mixed reality application. They are employed in edutainment and guiding applications, but also to develop an audio-delivered ambient intelligence (Eckel, 2001a; Sennheiser, 2008). In the Listen project, presented by Eckel, an interactive augmented auditory display is devised that provides users with intuitive access to personalized and situated audio information spaces while they naturally explore everyday environments (Eckel, 2001a,b). A direct classification of these displays is difficult, but they are most strongly related to spatial auditory user interface. Kobayashi and Schmandt and Ishii et al. have both developed a system that, in analogy to Figure 19, positions sounds and information in 3D space around the user's head (Kobayashi and Schmandt, 1997; Ishii et al., 1998). The user is thereby free in the interaction and decides what information is most important. A similar approach was later chosen by Holland and Morse, who designed an audio-based user interface for operating the Global Positioning System (GPS). The interface is designed in a way that it allows mobile computer users to perform localization tasks, while their eyes, hands and attention are otherwisely engaged (Holland and Morse, 2001). The often occurring front/back confusions of spatialized sound sources were overcome through an additional muffling and low-pass filtering of sounds positioned in the rear.

The following four sections briefly highlight some of the most important areas of application, of which some are directly related to the later introduced sonification and interaction techniques for the exploration of 3D virtual/augmented auditory environments. Audiogames resemble a special case of entertainment-related auditory display systems. An overview was provided and discussed in Section 3.4.

4.3.1 *Monitoring Applications*

The task of a monitoring application is to simply identify and display changes that are occurring within a system. Therefore, the application searches for a certain pattern in the raw data set, which, if found, is sonified accordingly. Common examples for monitoring applications are a medical heart monitor and a Geiger counter, but also financial trading and stock market data sonifications (Kramer, 1992; Fitch1992 and Kramer, 1992). All examples employ a very direct audification and display of the underlying data. The advantages for using auditory displays in a monitoring task is that the display of information is performed *eyes-free* and can easily be backgrounded, thus allowing to direct one's attention to other – possibly visual – tasks (Cohern, 1992).

4.3.2 *Data Exploration and Analysis*

An auditory data exploration and analysis requires sonification techniques for an analytic listening and presentation. Similar to monitoring applications discussed in the previous section, data exploration and analysis is based on a template matching, in which the user acoustically explores the data using a data-to-sound mapping approach. The main difference to monitoring is that one can not precisely anticipate what will be heard. Therefore, an interactive user interface is required that allows access to data scaling, routing, processing, selecting, analysis and other operations. To find significant signatures and features in a data set, exploratory techniques are employed, which might, in turn, result in the design of an optimal display system for a later monitoring task.

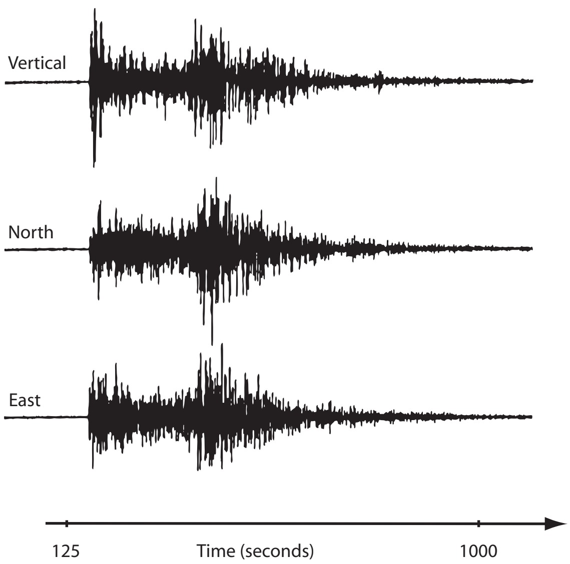


Figure 20: Three Component Seismogram of a nucular Explosion at NTS (Recorded at Lajitas, Texas, USA) (Heyward, 1992).

Audifications of Earthquakes and Explosions.

Applications for data exploration and analysis using auditory displays are in the areas of census, environmental, math, physics, geography, and many more (Kramer, 1994; Blattner et al., 1992; McCabe and Rangwalla, 1992). Heyward uses audification for the display of seismic data (Heyward, 1992), see also Figure 20, which shows three seismograms that can directly be audified. An analysis of these seismograms is often used in oil and coal explorations to find new resources. Auditory displays are not only applicable for statistical analysis of complex multi-dimensional data sets, such as for studying census and environmental data (Scaleety and Craig, 1991; Madhyastha and Reed, 1992), but also to explore physical and mathematical simulations (Rabenhorst et al., 1990; Kramer and Ellison, 1991; Mayer-Kress et al., 1992). Rabenhorst et al. present a system for the complementary visualization and sonification of multi-dimensional data sets, and describe the advantages of such a system (Rabenhorst et al., 1990).

4.3.3 *Auditory User Interfaces*

Important for all implementations is the integration of an auditory user interface (AUI) that allows the user to interact with the system and to perform changes and adjustments. An advantage of an audio-based user interface is that the area of interaction is not limited to the space in front of the user, but can be arranged 360° around the listener (Crispien and Fellbaum, 1996). This allows an interaction using a ring-based metaphor, as it was already discussed and is pictured in Figure 19. Along this ring menu system, several auditory widgets can be aligned, such as auditory sliders, buttons, checkboxes and more. Mynatt developed an auditory extension for an existing windows-based graphical user interface to evaluate the possibilities and advantages of an AUI (Mynatt, 1992). She experimented with several acoustic parameters to convey the content of the interface system and evaluated the results through an informal user study. The evaluation revealed that several of the parameters have to be enhanced and exaggerated, in order to be perceived properly (Mynatt, 1992; Brewster et al., 1992). Additionally, and in continuation of the work of Kramer, Mynatt is nesting symbols within symbols to acoustically represent different states of a button (Kramer, 1992; Mynatt, 1992). This approach can be described as a first implementation of a so called *Auditory Texture*, which will be discussed in more detail in Section 5.3.

Several Earons for a Paint Application.

3D Menu Example.

Other examples for auditory user interfaces include the sonic enhancements for two dimensional graphical displays by Blattner et al., as well as the 3D spatial interfaces for dynamic soundscapes from Kobayashi and Schmandt and the ambientRoom by Ishii et al. (Blattner et al., 1992; Gaver, 1992; Kobayashi and Schmandt, 1997; Ishii et al., 1998). The last two examples already employ spatialized sound sources to represent a real 3D user interface.

4.3.4 *Assistive Technologies*

Visually impaired people have developed certain skills to rely on auditory information alone for a navigation and orientation in the real world. This orientation is based on familiar sounds, slight vibrations and the personal experiences developed over the years. Many tools have been developed in the last decades to aid the navigation and orientation of the blind community. Some of them use the global positioning system (GPS) in combination with virtual maps and a speech processor to identify personal locations and to sonify interesting and important points nearby (Frauenberger and Noisternig, 2003a; Strothotte et al., 1995).

However, these techniques are not only applicable for guiding the visually impaired, but can also be employed to assist sighted users in visually demanding tasks, such as driving or the operation of complex machines. Several of the previous publications focus explicitly on aiding the visually impaired, while the majority of applications focus on both user groups. Nevertheless, several projects, especially in the domain of audio-only computer games, were only initialized due to requests of blind players (Warp, 1999). Much of the research in this thesis, although it was not focussing on the visually impaired, can be applied towards the development of assistive technologies. Some of the prototypes have already been evaluated and tested by several blind users, refer to Chapter 9.

4.4 Summary

After this short analysis of auditory display techniques and a more detailed discussion of the important fundamentals in auditory perception, the following Chapter 5 continues to explore 3D auditory display systems with the focus on designing 3D virtual auditory environments. Several of the here introduced sonification techniques will be studied in more detail in the following chapter, as well as extended and applied for the sonification of regular 2D/3D data sets and for 3D scene auralization. By looking at the examples discussed over the last two chapters, one easily recognizes a vast applicability and a high potential for sound and acoustics. Some of these examples are further studied in the following chapters and reevaluated for an application within 3D virtual auditory environments.

I am the primitive of the way I have discovered. — Paul Cézanne

5 AUDITORY ENVIRONMENTS

VIRTUAL auditory environments can be described as being the acoustic analog to the more common virtual *visual* environments. They are directly based on auditory display systems, but also contain elements from 3D virtual reality, psychoacoustics and 3D interaction design. The objective of this chapter is to provide a firm basis and to develop an understanding of 3D virtual auditory environments in terms of design, presentation, interaction and application. The primary goal is the development of techniques which allows to establish 3D auditory environments as an equal to visual environments. Task-tailored techniques for an intuitive information sonification and interaction have to be conceptualized, implemented and evaluated. With the words of *Cézanne*, who once recognized the new direction that European art was given in the early fourteenth century by abandoning stylized medieval formulas of representation, the same can be said for and applied to 3D virtual auditory environments as well.

Starting in the next section, the chapter discusses similarities and differences between visual and auditory environments, and defines 3D virtual auditory environments using a formal description of the elements specified above. In the following, the discussion concentrates on the requirements for an intuitive 3D scene auralization and develops the concept of a non-realistic auditory scene design. Very important are also the techniques for 3D scene sonification and interaction, which are elaborated subsequently. Starting with advancements for the sonification of 2D and 3D data sets, the discussion develops and explores new concepts for the sonification of 3D auditory scenes. An integral part in this discussion is the analysis and development of techniques for 3D spatial interaction. These techniques are based on real-world interaction paradigms and performed in 3D space. This chapter concludes with the presentation of an audio-centered framework design and a discussion of promising areas of application.

5.1 VIRTUAL REALITY AND AUDITORY ENVIRONMENTS

The term *virtual* is defined[1] as something that exists only in the mind, as a product of the own, personal imagination. In computer science, *virtual* is associated[1] with simulated and digitally created environments. Milgram et al. define *Virtual Reality (VR)* as:

> *"In general, a Virtual Reality environment is one in which the user is immersed in a completely synthetic world, which mimics the properties of a real-world environment to a certain extent, and which may also exceed the bounds of physical reality by creating a world in which the physical laws governing gravity, time and material properties no longer hold. In contrast, the real-world environments of Augmented Reality systems are obviously constrained by the laws of physics, which necessarily impose certain restrictions on one's ability to interact with the world."* (Milgram et al., 1995)

Nevertheless, it is debated, wether VR is a *technique* that allows to create virtual environments (Heilbrun and Stacks, 1991; Bowman et al., 2004), or if it is

> *"one possible outcome of the biological capacity to imagine, to think in advance, and be prepared to situations to come."* (Hoorn et al., 2003)

1 http://www.thefreedictionary.com/virtual

Although both arguments are true, in this research *Virtual Reality* is seen more as a technique that allows the creation of an imaginary, virtual environment.

The level of *virtuality* is thereby defined along the virtuality continuum, Figure 21, which represents a continuous scale between a real (extreme left) and a virtual (extreme right) environment (Milgram et al., 1994). The area in between is defined as *Mixed Reality (MR)*, and defines a gradual transition that uses elements of both worlds. One application is *Augmented Reality (AR)*, which enhances the perception of a real-world environment through the integration of artificial information, see also the discussions in Chapter 6. An alternative, and currently very attractive field of research is the area of *Ubiquitous Computing*, which in its goals is rather opposite in direction and introduces computers and technologies into the user's environment, rather than forcing the user in a virtual environment enhanced by a computer (Weiser et al., 1999).

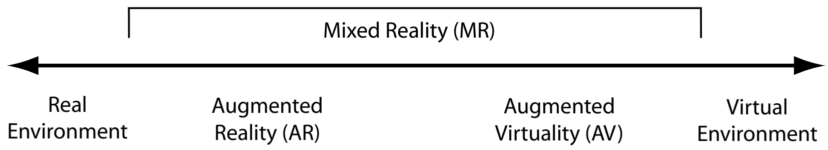


Figure 21: Reality-Virtuality Continuum (Milgram et al., 1994).

The requirements for VR applications are directly bound on the applications goal and the tasks to be performed. In general, these tasks include a travel and exploration (wayfinding) of these virtual environments, possibilities to select and manipulate objects, as well as techniques for controlling the system and – task-dependent – an input of symbolic information (Bowman et al., 2004). In order to perform these tasks, the 3D scenes have to be authored and extended by information, data and techniques to augment the 3D virtual scenery. The interaction with, and the display of these virtual worlds is generally performed using large, often stereoscopic, displays. An alternative presentation of the virtual environment, eg. through sound and acoustics is possible, but the differences in perception and data mapping have to be accommodated. An analysis of these differences is the focus of the following sections. In these discussions, virtual reality and virtual environments are analyzed in respect to their presentation, interaction, degree of realism and display. In a second step, these principles are applied and mapped onto an auditory perception and towards the creation of 3D virtual auditory environments.

5.1.1 *Virtual Reality*

In order to be able to describe the differences between a virtual and an auditory environment, a common basis is required. This basis can be formed using a formal, abstract description, which not only offers a more precise and theoretical study of virtual visual/auditory reality, but also allows to express the processes of interaction and display as a product of various sets and relations. First, a formal definition for virtual reality is developed, and is later extended and transferred to describe 3D auditory environments and augmented audio reality as well. In accordance to the enhanced models described by Hoppe and Ritter, the here developed concept employs a similar modeling (Hoppe, 1998; Ritter, 2005). Figure 22 provides an overview of the formal description along several input and output data streams. The 3D virtual environment is labeled as *Enhanced Environment* $\mathcal{E}$ that contains a set of *Objects* $o \in O$, which are positioned and ordered within the scene using structural information E_S. These objects are defined using sets of geometrical data E_G, as well as through an object-dependent set of symbolic information O_S that provides

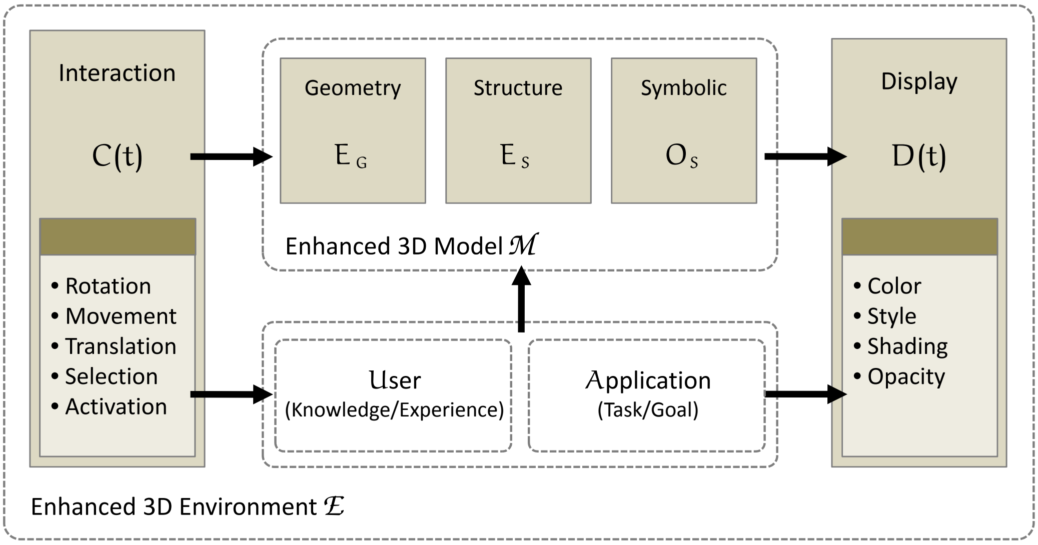


Figure 22: Formal Description for 3D Virtual Environments.

related semantic specifications and settings. Such an enhanced environment also contains a set of interaction relations[2] $C(t)$ and display variables $D(t)$, which are used to represent and display objects contained in the 3D scene:

$$\mathcal{E} = o \in \tilde{E}_G, \tilde{E}_G \subseteq E_G \tag{5.1}$$

GEOMETRICAL DATA (E_G) contains a mathematical description of the underlying geometry that is required for the definition of the 3D objects used in the virtual environment. This data is provided in the form of polygons, triangle sets and/or curves.

STRUCTURAL INFORMATION (E_S) is a set of relations ϕ that assign a position and orientation to each object that is contained in the virtual environment. Structural information maps objects $o \in O$ onto a subset of E_G: $\{\phi | \phi(o, \tilde{E}_G) = o \times \tilde{E}_G, o \in O, \tilde{E}_G \subseteq E_G\}$.

SYMBOLIC INFORMATION (O_S) is a set of relations φ, which provide semantic descriptions for each object: $o \in \tilde{E}_G$, eg.$\{\varphi | \varphi(o, s) = o \times s, o \in \tilde{E}_G, s \in O_S\}$.

INTERACTION DATA ($C(t)$) defines a set of relations – dependent over time t – for an interaction on/with objects: $\{\psi | \psi(c, o) = c \times o, c \in C, o \in \tilde{E}_G\}$. The input depends on the interaction devices used and maps a data vector (input stream) onto the currently selected objects, which in turn provide feedback information with a modification of the objects display settings $D(t)$.

DISPLAY SETTINGS ($D(t)$) are a set of relations ξ that define the appearance and display of 3D objects: $\{\xi | \xi(d, o) = d \times o, d \in D, o \in \tilde{E}_G\}$. These display settings map color, style and other forms of visual representation onto the object's associated parameters. Display styles are activated through either an interaction on objects (eg. $\psi(c, o)$), or through a selection of symbolic information (eg. $\varphi(o, s)$) that requires the object to change its information display.

2 $C(t)$ = from Control

Objects that are parts of the enhanced environment $\mathcal{E}$, eg. $o \in \tilde{E_G}$, are a combination of their geometric representation E_G, their structural information E_S, and their symbolic description O_S. As a result, these objects can be described analogously as *Enhanced Models* $\mathcal{M}$:

$$o = \mathcal{M}; \forall o \in \tilde{E_G} with \tilde{E_G} \subseteq E_G \mathcal{M} = (E_G \times E_S) \times O_S \tag{5.2}$$

Using the above defined sets and relations, the interaction and display of a 3D virtual environment can be expressed using a three-fold relation as:

$$\mathcal{E} = (C(t) \times \mathcal{M} \times D(t))) \tag{5.3}$$

In this equation, $(E_G \times E_S)$ describes the pure geometric representation and the arrangement of objects within the scene, eg. $\phi(o, \tilde{M_G})$. As this virtual scene not necessarily resembles any real world place, no further mapping is required. However, the definition of a mixed reality environment demands an additional, possibly bijective, mapping of the virtual scene objects onto its real-world counterparts, see here Section 6.1 for a more detailed discussion. A secondary projection in Equation 5.3 maps symbolic information (O_S) with additional data and a semantic description for each object onto the existing expression and specifies its specific display $(D(t))$ and interaction $(C(t))$ behavior. To additionally accommodate a user/task-dependency ((A)pplication) and the previous knowledge of a prospective user ((U)ser), Equation 5.3 is modified to:

$$\mathcal{E}_{(User \times Application)} = (C_{(U \times A)}(t) \times \mathcal{M}_{(U \times A)} \times E_S) \times O_S) \times D_{(U \times A)}(t))) \tag{5.4}$$

For a customization on a specific task, not only the underlying 3D model, but especially the interaction, as well as the techniques for the display of information have to be adjusted. A task-dependent interaction on enhanced objects $\mathcal{M}$ can be described as:

$$c \rightarrow \mathcal{M} \rightarrow d, with\ c \in C(t)\ and\ d \in D(t) \tag{5.5}$$

Equation 5.5 describes the mapping of interaction data onto scene objects $\mathcal{M}$. This mapping results in a change of display for this object, or for all objects in this scene.

Using this formal description, a precise modeling of 3D virtual reality is now possible. This formalism not only allows a clear definition of objects and their specific display, but also an accurate modeling of scene interaction and its effect on the entire environment and/or specific objects. Through an exchange of the display variables $D(t)$, different *depictions* of a 3D scene are possible. A substitution by *auditory* display techniques allows a modeling and description of 3D *auditory* environments, refer to Section 5.1.2.

With this formal definition of virtual reality in place, one needs to identify the qualities and characteristics inherent in VR simulations. The terms *virtual reality*, *virtual environments* and *virtual worlds* are often used synonymously, although virtual environments have a stronger association with interaction and a resemblance of a real-world space. Stuart defines *Virtual Environments* in the following way:

> *"An environment is that what surrounds you, the set of conditions and objects you can perceive and with which you can interact. A virtual environment is an interactive computer-generated environment provided by a VR system." (Stuart, 2001)*

Other characteristics that are often used to describe VR systems are the terms *presence and tele-presence*, *immersion*, *involvement* and *flow*. Sheridan introduced *presence* and *tele-presence* to describe a *being somewhere else*, in an imaginary or simulated environment (Sheridan, 1992). Later Witmer and Singer extended these ideas and coined the terms *immersion* and *involvement* as a special state of mind (Witmer and Singer, 1998). The feeling of being fully integrated and being a part of a virtual environment is thereby referred to as *Immersion* (Schirra, 2000), which Coomans and Timmermans describe as:

> *"the feeling of being deeply engaged. Participants enter a make-believe world as if it is real." (Coomans and Timmermans, 1997)*

For Smith et al., immersion is strongly related to the senses that are involved in the perception. Based on these assumptions Smith et al. define a sense-dependent *level of immersion*, which, according to Smith et al., delivers the weakest immersion for audio-only and the strongest immersion for a combined haptic/audio/visual display (Smith et al., 1998). Strongly related to the definition of immersion is also the term *flow*, which was introduced by Csákszentmihályi in 1975, and describes a person that is fully immersed and deeply engaged in a single activity or task (Csákszentmihályi, 1975; Böttcher, 2005). Several implementations of computer games focus explicitly on flow and the achievement of a high level of immersion. Examples are *REZ* and *flOw*, which both obtain this goal using trance-like music and a very simple and intuitive interaction design (United Game Artists, 2001; thatgamecompany, 2008).

Virtual reality has received a high level of attention in the early 1990s through a large coverage in media, movies and books. It was hyped in public discussions as a solution to many problems, but unfortunately not able to deliver the promises. A large problem was the media industry itself, which announced unrealistic future technology developments that could not be implemented at this time. Additionally, much of the hardware technology was in its infants and not enough content was available to fill all of the VR systems with *life* (Murray, 1998; Brooks Jr., 1999). With the increasing realism in computer graphics and acoustics over the last decade, as well as through powerful commodity hardware available, VR might be able to deliver what was once promised promised in a near future (Brooks Jr., 1999). The availability of content and applications that efficiently utilize VR systems and thereby legitimate their employment is of the highest importance. Prospective applications are found in the areas of medicine, 3D visualization, computer gaming, archeology and virtual heritage, as well as in numerous virtual development and training scenarios (Freudenberg et al., 2001a; Bowman et al., 2004; Fraunhofer IFF, 2008).

5.1.2 *Virtual auditory Environments*

As was outlined at the beginning of this chapter, 3D virtual *auditory* environments can be thought of as being the auditory analog to a 3D virtual *visual* environment. With the focus centered on 3D game design, Zizza describes:

> *"the formal definition of an 'audio environment' as defining the parameters and boundaries of the sonic world living in your game." (Zizza, 2000)*

Zizza provides in his article also the specifications for creating a so called *audio design document* that lists and describes the use of all auditory elements for designing a 3D audio/visual computer game (Zizza, 2000; Crawford, 2002). This description, however, perceives *audio environments* merely as a decorative padding for the visually dominating 3D game world, whereas the research in this thesis aims to establish auditory environments as an equal to visual environments. Although the perception, as well as the representation of objects and 3D scenery, differs in many aspects, the majority of ideas and techniques that are applicable in the visual realm can be adopted and transferred towards a 3D scene sonification and an audio-centered interaction design.

The formal model of Equation 5.3 for describing virtual environments can directly be applied to define virtual *auditory* environments as well. Some modifications are, however, necessary in order to accommodate the differences in perception and presentation. This includes techniques for the acoustic display of data and information $(D(t))$ (sonification), as well as a more audio-centered design of interaction techniques $(C(t))$. The display variables $(D(t))$ now comprise loudness, frequency, duration, pulse and many more. In view of an adaptation of these methods, one also has to consider that auditory information is only perceived over time and possesses no latent image. Although a temporal dependency was already integrated in Equation 5.3 for a time controlled interaction and display, this dependency is omnipresent in auditory environments and has to be extended to include the enhanced scene objects as well, ie. $\mathcal{M}(t)$. The variables for display and symbolic information are now centered around an auditory description. This includes different sounds and acoustic parameters for each object $\mathcal{M}$, which now form an *auditory texture* describing each $\mathcal{M}(t) \in \mathcal{E}(t)$. This approach uses a different sound or auditory representation to describe the various conditions $s \in O_S$ of a certain object:

$$D(t) = \mathcal{M} \times s(t), s \in O_S \tag{5.6}$$

A small example shall be used to compare the differences in perception and the varying perceptual spaces for an auditory and a visual environment. This can be seen in Figure 24, which displays a common living room environment that is equipped with a TV set, a desk, a computer, several book shelves, a bed, and a coffee maker. Located in the middle of the room is a virtual avatar that is facing the door. This person's perception and point of view/hearing will from now on be used in several examples throughout this thesis. Additionally, auralizations of example scenes are provided to acoustically enhance the visual depictions. In these auralizations, the listener's position is identical to the avatar's position depicted in the figures. Towards the end of the each auralization, the virtual avatar turns around 360° to provide a better perception of the auditory scene. In the first example, Figure 23 shows both, the visual and the acoustic perceptual spaces, while Figure 24a and Figure 24b compares their respective sensory perspectives. Figure 24a shows the person's visual experience, eg. what can be seen in the room from the avatar's position, while Figure 24b shows the auditory environment in form of a spherical map that highlights audible objects in red. As can be seen from this example, some parts in both environments overlap (eg. clock, telephone, computer), while the majority is disjunct and either perceived visually (eg. door, books, chair), or acoustically (eg. radio, TV set, coffee maker).

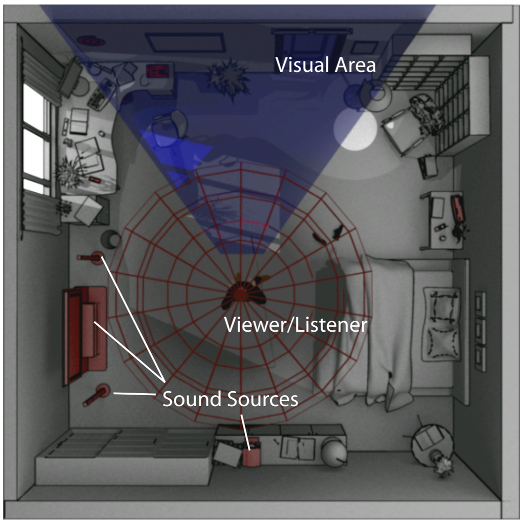
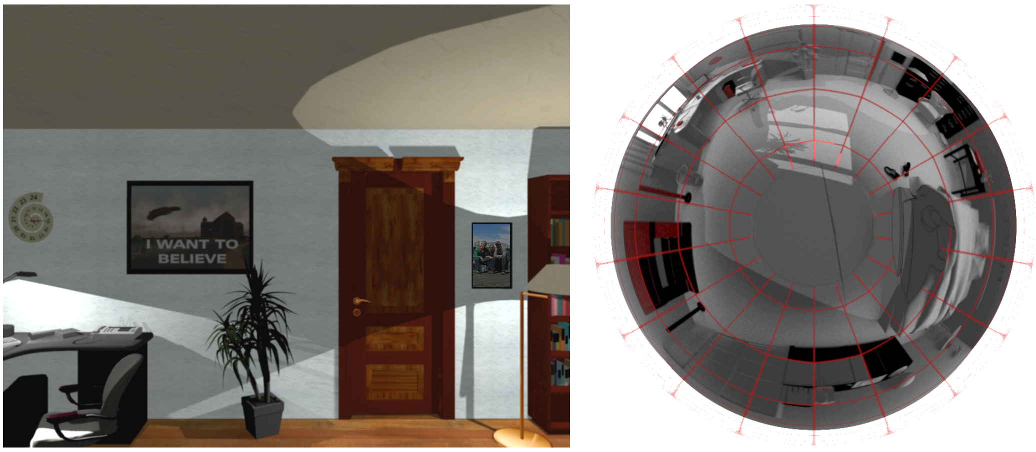


Figure 23: Different Perceptual Areas: Visual (Cone) and Auditory (Sphere).

Auralization of Figure 24b.

Other characteristics of an auditory environment are a possibly higher level of immersion and a strong suitability for narrative presentations (Röber et al., 2006b; Huber et al., 2007). The statement of Smith et al., who classify audio-only presentations as being least immersive, has to be strongly rejected as several studies suggest otherwise (Smith et al., 1998; Röber et al., 2006b). Audio-only applications can, if well designed, be much more stimulant and immersive than audio/visual depictions. The story in a movie or 3D computer game will always be reduced to what is visibly displayed, while in an



(a) Visual Space. (b) Auditory Space.

Figure 24: Visual and auditory Perceptual Areas for a 3D Scene.

auditory presentation, the user fills in the *missing* information through own and personal experiences. This concept of *customization* often provides a much deeper immersion and is often used in games and movies to create an atmospheric suspense (Curtis, 1966-1971; Eidos Interactive, 1998; Konami, 2007).

As stated earlier, one goal of this thesis is to develop and establish auditory environments as an *equal* to visual environments. A key element in this development is the design of techniques that support the listener's perception to receive enough – but not necessarily the same – information as through a visual display. The development explicitly concentrates on the advantages of an auditory perception, eg. a 360° listening and interaction, a multi-modal and non-focussing perception, as well as the possibilities to design an affordable, lightweight and highly portable system.

Definition A *3D Virtual Auditory Environment* describes a 3D virtual environment that can be perceived solely through the means of sound and acoustics. Its representation employs techniques for an audio-centered display, and provides auditory cues to interpret the local environment and to localize 3D objects. The display variables $(D(t))$ include loudness, frequency, duration, pulse and more, as well as support an auditory representation to describe an objects structural information O_S. The interaction $(C(t))$ adheres to natural listening behaviors and is based on 3D head-tracking and a spatial interaction design. The techniques for scene interaction support navigation, orientation, exploration/pathfinding, as well as object selection, activation and alteration tasks.

By examining the example of Figure 24, one can assess the requirements for an implementation of interactive 3D virtual auditory environments. Such a system must create acoustic events that stimulate listening experiences within the user and immerse him acoustically into the virtual environment. These acoustic stimuli have to provide enough information to support a situational awareness, ie. to identify the environment and the own position and orientation within. This requires techniques for an intuitive interaction with the environment, as well as with the objects therein. Several of these tasks can be implemented in a similar way to the previously discussed audio/visual interaction techniques, but have to be centered around an auditory perception that allows the design and utilization of real 3D – spatial interaction metaphors (Bowman et al., 2004). The design of auditory environments also depends on requirements from the application itself, the task to accomplish and the prospective user. As this discussion would lead deep into the area of software design, references to the literature are provided for a further study (Shneiderman, 2004; Faulkner, 1998).

The remaining sections in this chapter discuss and illuminate 3D virtual auditory environments from various perspectives. In analogy to Equation 5.4, the different parts required for the development of auditory environments are explored. Structural and symbolic relationships, ie. $\phi(o, \tilde{M}_G)$, are discussed in Section 5.2 with a focus on the design of non-realistic auditory environments. The following Section 5.3 continues this discussion and highlights sonification techniques for an object/information mapping, ie. $\varphi(o, s)$. Spatial interaction techniques and their mapping onto scene objects $\psi(c, o)$ are examined in Section 5.4, while alternative display styles for an auditory representation of objects $\xi(d, o)$ are covered in all three sections, but especially in Section 5.2.

5.2 AUDITORY PRESENTATION AND DISPLAY

In a visual representation, the properties of an object are mapped onto graphical primitives using a variety of attributes, such as color, style, shading etc. (Schumann and Müller, 2000). For an auditory representation of the same information, these properties are mapped onto *acoustic primitives*, such as loudness, timbre, frequency etc. (Kramer, 1994). Besides a direct audification and *data-to-sound* mapping techniques, other methods can be used to encode information acoustically, such as tempo, rhythm, harmonies and complexity (Stockmann, 2008), see also Section 4.2 for a detailed discussion.

For the physical display of auditory data, three possibilities are available:

- Headphones,
- Surround sound displays, and
- Wavefield synthesis.

Although wavefield synthesis delivers the best performance in sound localization and can be used for several listeners simultaneously, it is still a very complex and difficult technique that requires an awful amount of speakers (Boone, 2001). Surround sound displays are also applicable to larger groups, but with a much smaller *sweet spot* compared to wavefield synthesis (Begault, 1994). A disadvantage of both techniques is the introduction of acoustic artifacts that originate in the listening room's acoustics and the cross-talk cancelation techniques required (Shilling and Shinn-Cunningham, 2002; Vorländer, 2007). Headphones provide a very good sound perception and at the same time permit a direct display of binaural data. Combined with a possible head-tracking technique, this allows a very intuitive and efficient display and perception of 3D auditory environments.

Therefore, the research in this thesis is based on a headphone-specific display and binaural sound rendering.

Another interesting aspect for the display of 3D auditory environments is the degree of realism employed. A related field in computer graphics is the area of so called *Non-photorealistic Rendering (NPR)*, which concentrates on the users perception and an efficient conveyance of an image's underlying information by appealing to and mimicking human drawing techniques (Strothotte and Schlechtweg, 2002; Gooch and Gooch, 2001). It is applied in many areas and has applications in architecture, archeology, scientific visualization, computer games, and many more (Spindler et al., 2006; Freudenberg et al., 2001b). The examples in Figure 25 show two visualizations of the *Magdeburger Kaiserpfalz* – the palace of Otto the Great, one of Europe's most influential medieval emperors

(a) Photorealistic Presentation. (b) Non-photorealistic Presentation.

Figure 25: Varying Realism in a 3D Scene Presentation.

(Nickel, 1973; Meckseper, 1986). A detailed analysis of the stratigraphic sequences in 2000 revealed, however, that the remnants of two buildings from two different periods in time were accidentally mistaken as one (Ludowici, 2000). To portray these findings, Figure 25a is rendered in a cartoon/pen&ink-style drawing technique that also visualizes a decreasing veracity of the buildings shape from the bottom to the top (Röber, 2001; Freudenberg et al., 2001a). A similar concept can also be applied to the display of 3D auditory environments. Here Lodha et al. devised a technique for an acoustic sonification of uncertainty (Lodha et al., 1997). Such a display not necessarily adheres to a strict physical-based sound rendering, but consciously alters an object's acoustic representation by exaggerating certain effects and by including additional information.

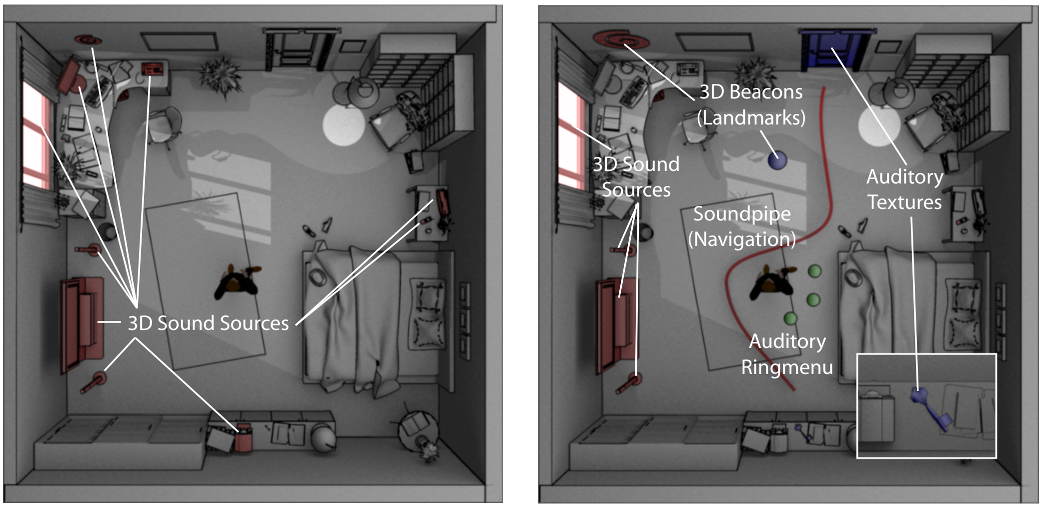
5.2.1 *Scene Auralization*

Auralization describes a process that is used to make a virtual scene *audible*, eg. to map data and information onto acoustic primitives and to display them in the form of *sound waves*. Auralization is often linked with a more abstract, physical and mathematical description, see here Chapter 8, but shall be viewed in this chapter from a more general perspective. Although an auralization of a 3D auditory environment can be performed in number of ways, it can be reduced to just three basic auditory elements (Röber and Masuch, 2004b):

- Speech,
- Music, and
- Sounds and noises.

Each of these elements responds to a different perception. Whereas *Sound* – in its most general definition – often describes a physical process or an action in the local environment, speech and music are both more abstract forms of communication. *Speech* can be used to describe an information very precisely by using many details, while *Music* is often used to convey emotions, atmospheres and moods.

One form of scene auralization that is solely based on the perception of speech can be found in *Audiobooks* and the dramaturgically enhanced *Radio Plays* (Fey, 2003). Both forms of narration enjoyed an increase in popularity over the recent years and can be



(a) 3D Scene Auralization. (b) Non-realistic auditory Display.

Figure 26: Auralization and Realism of a 3D auditory Environment.

used conveniently at many occasions. Audiobooks are, in general, only narrations of a book using a single speaker, while radio plays also employ music and sound effects, as well as the dramaturgical screen play of several actors. Both forms explicitly focus on the advantages of an auditory presentation and narration that highly immerse the listener into the storyline (Fey, 2003; Röber et al., 2006b). Storytelling and narration are also present in other forms of interactive media, eg. in computer games and here especially in the *Adventure* genre. An excellent example is the 2005 released game *Fahrenheit – Indigo Prophecy*, which develops new approaches for an interactive and non-linear storytelling that are consistent with the player's (inter)action (Quantic Dream, 2005). In combination with screen reader devices, old text adventures, such as *Zork*, can be turned easily into an auditory adventure game as well and be played in a very similar way (Infocom, 1982). The possibilities to employ speech for both the output and the input of information are manifold and partially discussed within this chapter. Nevertheless, speech plays only a minor role in this research as the focus is on the design of techniques to acoustically convey abstract information using non-speech sounds.

Pictures at an Exhibition – Promenade.

Music is a very powerful, and also a very emotional form of communication. As it is difficult to describe the semantics of an image or a virtual scene using sounds alone, music can be used to describe the content by appealing towards an evocation of emotions that are expressed in the image or scene. One of the earliest examples is the well known opera *Peter and the Wolf* by Prokofjew, which narrates a story using an orchestra in which each character is described and represented exclusively by a single instrument (Prokofjew, 1936). The interaction between the individual characters and their interplay create a wonderful and enjoyable piece of music that narrates a story primarily through music. A more abstract example is also the composition *Pictures at an Exhibition* by Mussorgski and Ravel (Mussorgski, 1874/1886; Ravel, 1922), and later an electronic interpretation by Tomita (Tomita, 1975). Both pieces express the composers experiences and emotions while looking at the paintings of the Russian artist Viktor Hartmann.

Correct blending of Music.

Computer games, as a modern form of storytelling, also employ music to express emotions and to enhance the perception of the story and the game play. As this form of interactive narration is not bound by a timed schedule as in operas and film, the transition between scenes and the blending between two pieces of music is often not predictable and can not be timed in advance. Therefore, most implementations utilize a gradual transition between two pieces, which often destroys the immersion due to the introduction of disharmonies that are caused by blending artifacts. A solution was found by examining the problem from a musician's perspective and through the development of techniques that allow a harmonic blending between various pieces of music (Berndt et al., 2006). The music is now composed and divided into several pieces, which can be blended correctly at certain points. This allows a much smoother presentation, and to maintain the listener's immersion.

The use of sounds and sound effects to describe a 3D environment is the most direct, but in terms of perception and clarity also the most difficult approach. Descriptive sound elements are used to identify certain objects or actions. These sound elements can also be spatialized using HRTF filters, which allows to determine an object's position and distance relative to the listener. However, not all objects can be intuitively described by sounds, eg. a table or a door, and therefore, this technique is initially applicable to certain elements only. In these cases, additional sound elements in the form of auditory icons and earcons can be employed and learned as a description for *non-sound objects*.

Auralization of Figure 26a.

Auralization of Figure 26b.

Figure 26 continues the example scenario previously introduced and shows a 3D scene auralization containing sounds, music and speech. The majority of sounds are thereby positioned in 3D space to pinpoint the location of the objects described. Other auditory elements, such as parts of speech and music, are presented for a diotic display (in-head localization), eg. they are not filtered through HRTFs and presented binaurally with the same level (Shilling and Shinn-Cunningham, 2002). The audible and 3D positioned elements of the scene are highlighted in red. The scene contains the following elements:

- Speech (TV, Radio, Telephone Answering Machine)
- Music (TV, Radio, or Scene Music)
- Sounds (Computer, Clock, Coffee Maker, TV, Street Noise, ...)

This itemization shows some of the audible elements that can be perceived in this scene. However, not all elements can be audible at the same time, as this would clutter the display and only leave a meaningless noise. Some of these objects are well identified using descriptive sounds, while others are more difficult.

5.2.2 *Non-realistic Auditory Display*

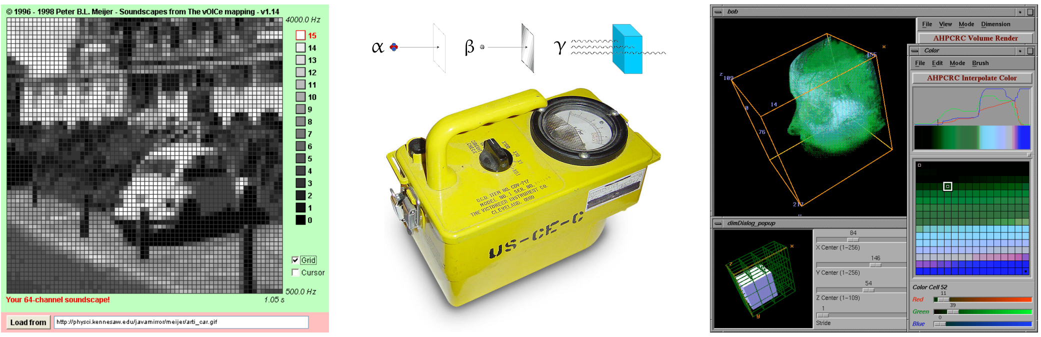
A first definition for a *Non-realistic Sound Presentation (NRS)* is found by Walz, who defines NRS as a method that consciously alters the physical representation of objects to better describe their characteristics, position and possible velocity (Walz, 2004a; Röber and Masuch, 2005a). This definition explicitly restricts itself to a non-realistic physical modeling, eg. the exaggeration of acoustic effects such as the Doppler, but ignores the inclusion of additional non-object-based sound sources and an alternative acoustic representation of objects, eg. *display styles*. Therefore this initial definition is recoined to:

Definition *Non-realistic Sound Rendering* describes a principle for the display of 3D auditory environments that focusses on an intuitive and *non-realistic* auditory presentation of 3D scene information. This presentation is not required to adhere to the laws of physics, instead it concentrates on the most intuitive and direct presentation of information available, by employing additional sound sources and artificial cues, as well as through a deliberate alteration of an objects acoustical appearance. The appearance of objects thereby might include different auditory representations, as well as a change of their physical attributes, such as Doppler factor, velocity, or distance. The basic fundamentals identifying such a *non-realistic* auditory presentation for a 3D scene are:

- A non-physically based acoustic presentation of the 3D scene,
- Additional non-object sounds,
- An exaggeration or reduction of certain physical parameters/laws, as well as
- Situation-based auditory display styles for object descriptions.

Figure 26 shows a comparison of a regular 3D scene auralization and a non-realistic auditory display of the same environment. The normal auralization applied in Figure 26a shows the familiar living room scene with a virtual avatar in its center. The majority of sound sources in this room is based on electrical devices, such as the computer and the TV set. These devices emit sounds from their operation, which can be used for an object identification, but also to determine the objects and the own position. However, these devices are not always switched on, and can not display additional, hidden information that highlights a possible interaction. Figure 26b on the other hand shows a *non-realistic* auditory display of the same scene. This scene display is enhanced by additional sound objects, as well as by several sonification and interaction techniques later to be introduced, refer to Section 5.3.2. The main focus of employing a non-realistic display is to enhance the perception of the 3D scene. Regular 3D sound objects, similar to those in Figure 26a, are displayed in red. However, the loudness of the clock on the left rear wall is amplified to utilize it as a beacon to identify the user's orientation within the virtual room. Another sonification to support this task are so called *North Beacons*, which are employed if no natural sound objects, such as the clock, are available to fulfill this task. The clock can therefore be described as a natural *Auditory Landmark* in this room. Another example is the *Soundpipe* concept, a technique that allows an intuitive and controlled movement through the 3D environment. Very versatile is also the concept of an *Auditory Texture*, which allows a variable acoustic description that is consistent with an objects function and state. In Figure 26b, both the door and the key object have an auditory texture assigned. The example setting displayed assumes a scene within an auditory adventure game, in which the door is locked and the protagonist has to find the key to unlock and open it. In this setting, the door can be sonified as being *locked, opened up* and *open*, whereas the key has to emit a sound that describes it as *key*, and which enables the user to actually find it. Accompanying this example are two auralizations that sonify both environments. To avoid a cluttering of the auralization of Figure 26b, it first displays the general acoustics with the the two beacons, the artificial north beacon in front and the amplified clock towards the front/left. After this, the auditory textures for the key and the door objects are presented, followed by a sonification of the soundpipe and at last a sonification of the 3D menu system. The auditory texture of the key and the door utilize a similar sound that directly denotes a dependency. The auralizations in this example are here included for completeness, with all sonifications explained in more detail in the following section.

The above examples highlight the most important characteristics for a non-realistic auditory design. In essence, it enhances the auditory perception of a 3D scene by deliberately altering the physically correct auditory appearance of 3D scene objects. Through the inclusion of additional sound objects and auditory descriptions, a listener becomes able to identify and interact with the environment. An exaggeration of certain physical parameters allows under given conditions a better perception of the 3D environment and can also be used to highlight specific scene objects using a technique similar to an *Auditory Lens*.



(a) The vOICe System[3] (Meijer, 2008). (b) Radiation Counter[4] (Rutherford and Geiger, 1908). (c) 3D Volume Sonification (Rossiter and Ng, 1996).

Figure 27: Data Sonification Examples.

The last two sections described some of the basic principles for an acoustic display of 3D virtual environments. The next section continues this discussion towards data and 3D scene sonification techniques, and develops methods that aid an intuitive perception and understanding of 3D scenes.

5.3 DATA AND 3D SCENE SONIFICATION

Sonification refers to a long existence in art & science, yet its definition remained rather fuzzy and was interpreted differently depending on the task and the area of application. A definition commonly used today, that serves as a basis for this research, was coined in a research report by Kramer et al. in 1997:

> *"Sonification is the use of non-speech audio to convey information. More specifically, sonification is the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation." (Kramer et al., 1997)*

A major goal of this research is the development of tools and techniques to support an intuitive perception and display of 3D virtual auditory environments. This requires techniques for the sonification of scene and object information, as well as techniques for interaction and scene/object manipulation. Sonification is employed to convey the underlying information, to acoustically describe the environment, the objects, as well as the interactions possible. Sonification, ie. the acoustic display of information, and interaction, ie. the physical act of inputting information, are thereby strongly intertwined and are both part of the interaction/feedback loop (Crawford, 2002; Bowman et al., 2004).

To allow a more precise discussion and a task-related evaluation of the individual techniques, sonification and interaction are both discussed in distinct sections, in which each concentrates on the respective characteristics. The following sections describe a variety of sonification techniques, with the focus on the development of acoustic representations for 3D virtual auditory environments. The succeeding Section 5.4 refers back to the sonification techniques that were developed here and combines them with a spatial and/or speech-based interaction design.

---

3 http://www.seeingwithsound.com/
4 http://www.wikipedia.com/

Technique	Presentation (Analyt./Synth.)	Complexity (# of Parameters)	A/S Continuum (Scale between 0 – 10)	Application (Monit./Analys.)
Audification	Both	$\leqslant 3$	$1-3$	Both
Auditory Icon	Analytic	$\leqslant 4$	$6-7$	Monitoring
Earcons	Analytic	$\leqslant 5$	$7-8$	Monitoring
Hearcons	Both	$\geqslant 16$	$3-5$	Monitoring
Harmonics	Synthetic	$\leqslant 4$	$7-9$	Both
2D/3D Scanline	Analytic	$\leqslant 3$	$2-4$	Analysis
Volume Chimes	Analytic	1	$1-3$	Analysis
Speech	Analytic	1	$9-10$	Monitoring

Table 3: 2D/3D Data and Volume Sonification Techniques.

The first part in this discussion concentrates exclusively on 2D/3D data and volume sonification techniques. The section discusses important principles and techniques, and employs them as basis for the later introduced 3D scene sonification techniques. Although data sonification is not the primary goal of this research, this section develops and discusses several improvements to existing data sonification techniques. The second part of this discussion later extends these techniques and develops 3D scene sonification techniques to acoustically describe global and local scene/object information. Both parts aim at providing rules and guidelines to define and select specific techniques depending on the sonification task aspired.

5.3.1 *Data Sonification*

Geiger Counter.

The primary goal of data visualization and sonification is the acquisition and the conveyance of knowledge through an exploratory analysis. The same principles that govern a graphical visualization of data can be applied to an acoustic sonification as well. Through the use of varying sonification techniques, a user gains knowledge about the structure, the layout, as well as trends and characteristics hidden inside the data set. The mapping of data elements towards acoustic primitives and an acoustic display is here always the first, but also often one of the more difficult steps. An example that uses a very simple design is the sonification of radiation data using a Geiger counter (Rutherford and Geiger, 1908), see also Figure 27b. The employed mapping technique directly auralizes the data using a pulse encoded sonification, ie. the system acoustically *ticks* for each decaying particle that is detected. The number of ticking sounds perceived thereby directly refers to the current radiation present. The underlying information is, as is the sonification itself, strictly one-dimensional. Other data mapping techniques allow an acoustic representation of 2D and 3D data sets, as well as a parallel sonification of several streams simultaneously.

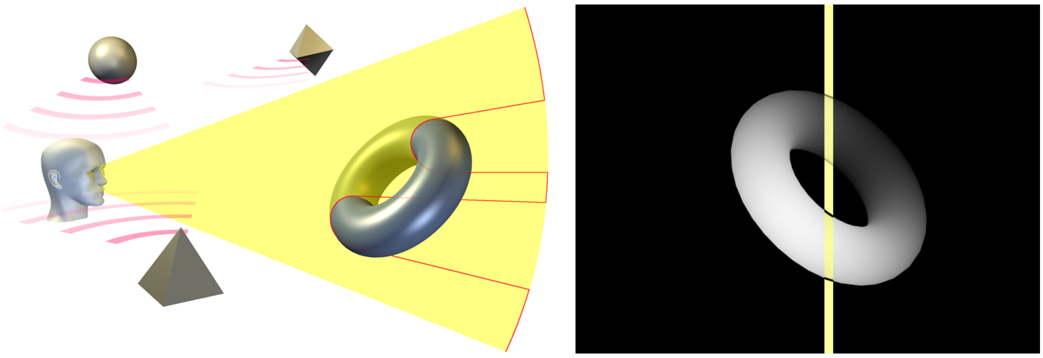
An intuitive sonification is able to directly represent the semantic information and establishes a link between the underlying data and the sounds displayed (symbolic representation). Contrary to visualization, a sonification of data is always perceived over time and requires a certain amount to interpret the sounds heard. Depending on the techniques used, this time ranges from a few milliseconds to seconds and even minutes for very complex sonifications that utilize harmonics and music in their display. Sonifications that are based on physical attributes, such as loudness, pitch, or timbre, are very fast to perceive and analyze. As a result, many sonification examples employ such a direct audification based on these three parameters. In some cases, however, it is also necessary to integrate an evaluation within the sonification itself; to not only display the raw data, but to additionally inform the listener about trends and assessments. An example is the sonification of stock market data, in which a different evaluation of the falling and rising of stocks may be employed depending on the ownership. Music and harmonic melodies can be easily used for such a task and display a falling stock using a minor, and a rising one with a major scale (Janata and Childs, 2004; Stockmann, 2008). Dissonances can be employed to sonify an imminent danger that requires an immediate and fast interaction (Roberts, 1986).

More complex sonifications can be achieved using a combination of techniques and a sequential display. Applicable for such a task are Blattner et al.'s Earcons and Bölke and Gorny's Hearcons (Blattner et al., 1989; Bölke and Gorny, 1995), refer also to the discussions in Section 4.2. Parameters for an encoding of additional information are pulse, tempo, rhythm and length. An accentuation can, for instance, be achieved through the deliberate use of the parameter *tempo* to create a dynamic sonification that enhances certain parts, while others are suppressed (Webster and Weir, 2005; Palomäki, 2006). Palomäki examined in this respect several adjective pairs, such as (*positive–negative*), regarding an acoustic representation using different rhythms (Palomäki, 2006). Also of importance is an appealing presentation and design of the auditory display itself, in which one has to find an adequate balance between function and aesthetics (Vickers and Hogg, 2006).

The selection of a certain technique depends on the sonification goal and the characteristics of the underlying data. Table 3 provides an overview of several sonification techniques that are applicable for an acoustic display of 1D/2D and 3D data sets. The overview provides details for the technique's presentation and complexity, its position within the A/S continuum, as well as for its primary area of application. The presentation differentiates between an analytic and synthetic display; in other words between an active and a passive presentation. This determines whether a technique is suited for a focussed, or a contextual display of information. Complexity describes the number of possible parallel data streams that can be sonified simultaneously, while A/S continuum describes the technique's position along the analogic/symbolic continuum (Kramer, 1994). The last row in Table 3 classifies the technique's primary area of application, ie. monitoring vs. analysis. A combination of techniques is easy to perform, and allows in most cases a higher number of segregable parameter streams. An example is sound spatialization, which enhances the perception and the complexity of each sonification method.

Shapes and Images

The sonification of one-dimensional signals, such as radiation data, stock market information or temperature curves, is relatively easy and straightforward. The acoustic representation of 2D information, such as shapes and images, is much more difficult. Several artists have explored this problem and developed innovative systems and techniques for the sonification of scientific data sets and 2D images (Quinn and Meeker, 2001; Quinn, 2007). For the sonification of image data, Quinn places line elements at certain positions in the image, which are then sonified using different instruments and a varying loudness, see also Figure 27a. Although the system is very interesting, both in function and in the acoustics synthesized, an intuitive understanding and interpretation of the underlying image remains difficult. One approach for the sonification of 2D image data that is employed often is the use of a scanline that traverses the image in a given direction and sonifies the area beneath, or the area in close vicinity to the scanline (Meijer, 1992, 2008). A digital image is thereby continuously *scanned* from left to right, and the pixels



(a) 3D Scanline. (b) Depth Buffer.

Figure 28: Scanline Sonification for 3D Objects.

are acoustically encoded using sinusoids representing intensity and position. An example of the *vOICe* system can be seen in Figure 27a. According to Meijer, the system can also be employed to sonify live video streams and to play interactive visual 3D computer games. However, to conclude from the auditory representation to the underlying image or video is almost impossible. Furthermore, the scanline used in the vOICe system moves automatically and continuously, whereas a user controlled scanline would allow a much more precise sonification, as well as permit an easier understanding of the arrangement of certain image structures.

Demonstration of the vOICe System.

The two important components for a 2D image sonification are an acoustic representation of color and shape. Commercial implementations to aid the visually impaired often restrict themselves to represent color alone, as is the case with the vOICe system (Meijer, 2008). The *Eye-Borg* system maps the color of the visible spectrum to the 12 semitones of an octave, which allows a detailed classification of the underlying data (Girvan, 2005). For a sonification of color, also symbolic mappings using auditory icons can be used. In this case, an icon represents a certain color and describes it through a symbolic association: such as a dripping sound of water to represent *blue*. However, the exact color is often not important and it is sufficient to display the information using a so called *warm/cold tone mapping* technique (Gooch and Gooch, 2001). Referring back to the discussions of music and harmonics in the last section, a sonification of color using a major/minor scale not only allows to determine the color of an image, but also provides an (emotional) impression of the content displayed.

2D Shape Sonification.

After these discussions regarding a sonification of color, the following paragraph concentrates on an acoustic representation of 2D shapes and the sonification of homogeneity in 2D images. The discussion of Section 3.1 has shown that besides a direct visualization of color, the presentation of shape and structural information is of higher importance. The vOICe system is based on a sonification of color, and therefore represents a white and a grey circle differently, although the primary feature *circle* is probably of higher interest. An edge based sonification, in which the user can switch between a color and a shape representation, is the best approach possible. Noisy data has to be processed with a smoothing filter in advance to reduce the number of *false* edges. An acoustic representation of an edge can be mapped to the parameters of loudness and frequency, in which loudness describes the edge's strength, and frequency its position along the y-axis. Listen to the 2D shape example on the left. The scanline itself can be moved freely by the user, which allows a more precise listening at complex areas. Figure 28 visualizes the principle for a 3D implementation of this techniques.

3D Objects and Data Volumes

The sonification of 3D objects and 3D volumetric data sets is a bit more complex, but based on the same principles as the sonification of 2D images. Examples can be seen in Figure 28 and Figure 29, and are discussed in more detail along an evaluation of the techniques developed here in Section 9.2.

The sonification of 3D objects can be performed analogously to the acoustic display of 2D shapes. For this purpose, the scanline is extended towards 3D and uses a depth-buffer edge detection as basis for the sonification. Figure 28a visualizes the principle of the technique and Figure 28b shows the listener's point of perception with a depth-buffer of the object and the 3D scanline inscribed. A depth buffer edge detection is employed to eliminate inflictions resulting from shading and lighting effects. The interaction is one of the key aspects of this sonification technique and employs a magnetic tracking system (Polhemus FASTRAK) with two orientation-sensitive devices, see also Figure 70 in Section 9.2. One device is used for the interaction itself, to orient the object, while the second device is mounted to the listener's headphones to determine the orientation of the user's head. This is required, along with the interaction and positioning of the data, for a correct binaural synthesis and 3D representation of the sonification result.

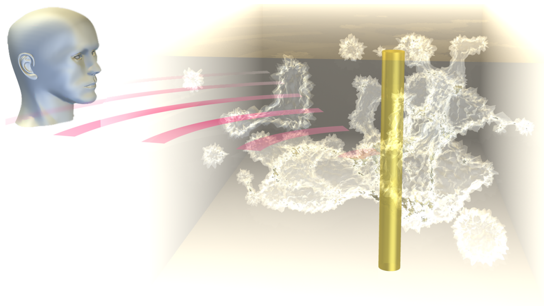


Figure 29: 3D Volume Sonification using an interactive Chimes.

The sonification of volumetric data sets is even more complex, as not only 3D shapes, but also the interior of a 3D object has to be presented acoustically. The interaction and sonification is similar to the acoustic display of 3D shapes, except that the sonification scanline now changes into a volumetric chimes that can be moved freely through the data volume, refer to Figure 29. The same interaction device that was used previously to rotate and position the 3D object is now employed for moving the sonification chimes through the 3D volume. Thereby the user acoustically explores the data set using the 3D chimes, while at the same time developing an understanding of the structure (density) of the data set. Kramer stated that

3D Volumetric Data Sonification.

> *"spatialized sound can, with limitations, be used to (...) analogically represent three-dimensional volumetric data."* (Kramer, 1994)

The sonifications from the 3D chimes are additionally filtered using HRTFs, depending on the chimes' relative position and orientation. The resulting sound is then binaurally displayed using an orientation-tracked headphone system. This combination allows a very precise and intuitive analysis of medium-complex volumetric data sets, and has been studied in more detail using a user evaluation which is discussed in Section 9.2. Earlier attempts of volumetric data sonification concentrated on a multi-variate display of the information to achieve a higher immersion, and to allow a better presentation of the underlying information (Minghim and Forrest, 1995; Rossiter and Ng, 1996). Figure 27c shows a screenshot of a system developed by Rossiter and Ng, which traverses the volume data and sonifies pre-segmented areas using different instruments and frequencies (Rossiter and Ng, 1996).

While the sonification of 2D images and scientific data sets is interesting and rewarding, the main focus of this research is the sonification of 3D virtual auditory scenes.

5.3.2 *3D Scene Sonification*

Some of the techniques that were discussed in the last section can directly be applied to a sonification of 3D virtual auditory environments as well. The beginning of this chapter discussed a formal description for an abstract modeling of VR environments. Equation 5.3 illustrated here the relation between an enhanced 3D environment $\mathcal{E}$ and its containing 3D scene objects $\mathcal{M}$. These objects are characterized by a (time-dependant) display $D(t)$, which visualizes the object's function and current condition:

$$D(t) = \mathcal{M} \times s(t), s \in O_S \qquad (5.7)$$

The actual *display* is thereby arbitrary, and can employ either visual or auditory techniques. For an auditory representation of 3D scene objects, techniques of 3D scene sonification are used. These comprise of various methods to acoustically describe different objects, as well as to represent their functionality and current state. Earcons and hearcons can be well employed for an acoustic description of 3D scene objects and to convey information within listener assistance systems. An example is the already introduced north beacon, which supports the user's orientation in a similar way than a visual compass. Parameters for the sonification and display of 3D virtual auditory environments have been partially discussed in Section 5.1 and Section 5.2. This section continues these discussions, as well as develops techniques for an auditory display of 3D objects and scene information in accordance to Equation 5.7.

A requirement for a 3D scene sonification is that every object within an auditory environment must be audible, otherwise it passes unnoted and could be removed. However, not all objects can be audible at the same time, and should rather be activated depending on the user's action and/or the intent of the virtual environment. The sonification objectives can be classified into three groups:

- Global information regarding the 3D scene and environment.
- Orientational and navigational information for wayfinding and scene exploration.
- Local information that describes objects, their state and possible interactions.

The sonification techniques employed must maintain an adequate balance between the display's function and an aesthetic design, as well as adhere to a natural listening behavior. The different characteristics of the information can be expressed trough various auditory means, for example, emotions and the setup of a certain atmosphere are best conveyed using music (Kiegler and Moffat, 2006), while more abstract information can be sonified using auditory icons and earcons, as well as through speech. The majority of the here discussed sonification techniques benefit from an additional (3D) interaction, which are introduced and explained later in Section 5.4.

The above discussions lead to the definition of 3D scene sonification, as it is employed in this research:

Definition *3D Scene Sonification* describes a set of methods and techniques that auralize and acoustically represent a 3D virtual environment. These techniques support the interaction, orientation, navigation and wayfinding, and thereby intuitively convey local and global information about the virtual environment and the objects therein. The techniques include a precise auditory representation of 3D scene objects, which displays their function, state and possible interactions. The 3D scene sonification is based on a non-realistic auditory design that aims at an intuitive display of semantics, connections and 3D space by employing solely auditory means.

Technique	Application (Global/Local)	Presentation (Analyt./Synth.)	Perception (Active/Passive)
Hearcons	Local/Global	Synthetic	Passive
Scene Object Grouping	Global	Synthetic	Passive
Auditory Landmarks	Global	Synthetic	Passive
North Beacon	Global	Synthetic	Passive
Soundpipes	Global	Analytic	Passive
Sonar/Radar	Local/Global	Analytic	Active
Magic Wand (Cane)	Local/Global	Analytic	Active
Auditory Texture	Local	Analytic	Active/Passive
Audio Lens	Local/Global	Analytic	Active
Attracting and Repelling Sounds (Guides)	Global	Synthetic	Passive
Buddy (Guide)	Local/Global	Analytic	Passive
Speech (Narrator)	Local/Global	Analytic	Passive

Table 4: 3D Scene Sonification Techniques.

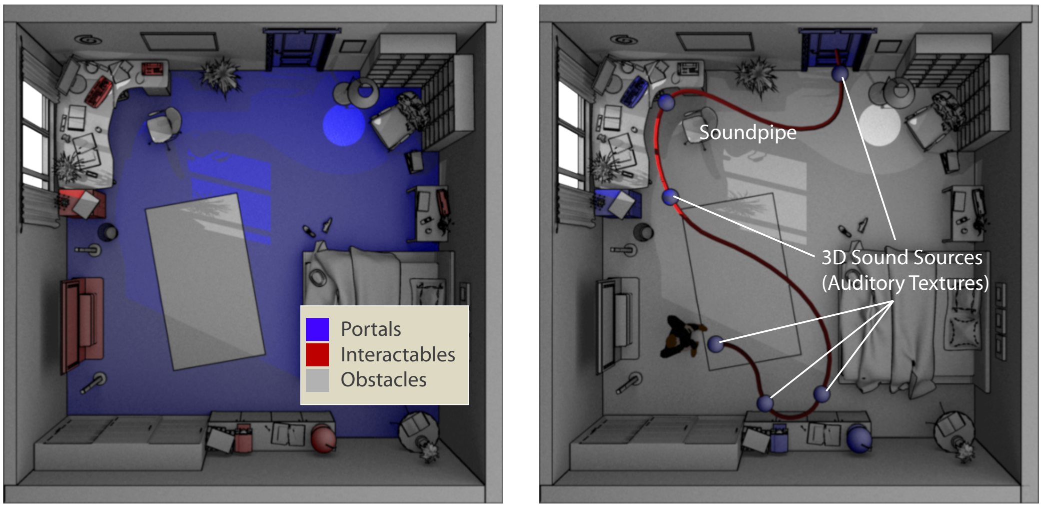
An overview of the employed 3D scene sonification techniques can be seen in Table 4. The list shows the major characteristics for each technique, such as application, presentation and perception. Application thereby describes how the technique is applied, ie. for sonifying global/environmental, or local/object-based information. Presentation is used in correspondence to Table 3, and describes the technique's display as either analytic (focus) or synthetic (context). Several of the techniques also require an additional interaction for an object selection and/or to change parameters of the sonification itself. This characteristic is described here through perception and its related interaction techniques, which are discussed in Section 5.4.

Global Sonification Techniques

Global sonification techniques aim at the representation of *global* attributes of the 3D environment and convey information that is required for a comprehensive understanding, as well as for navigation and orientation tasks. The majority of the here developed sonification techniques is based on non-speech cues. As can be seen in the examples in Figure 30, a global sonification outlines the 3D environment, highlights important objects and possibilities for interaction, as well as provides navigation and orientation cues.

The non-realistic sound rendering of Section 5.2.2 consists of elements of both, global and local sonification. Artificial hearcons and beacons are added to the auditory environment and support the user in terms of navigation and orientation. An example are *North Beacons*, which can be utilized as an auditory compass, as well as which describe an *Auditory Landmark* that identifies important and notable objects within the global/local 3D auditory environment. Both are implemented in the form of hearcons that represent the depicted object acoustically using a descriptive spatialized sound, refer to Figure 26b.

As depicted in Figure 30a, objects in a 3D virtual environment can be differentiated into three groups: *Portals* (blue), *Interactables* (red), and *Obstacles* (grey) (Röber and Masuch, 2004b). *Portals* are all objects within a virtual scene that allow a user to change the



(a) Scene Object Classification. (b) Pathway Sonification using *Soundpipes*.

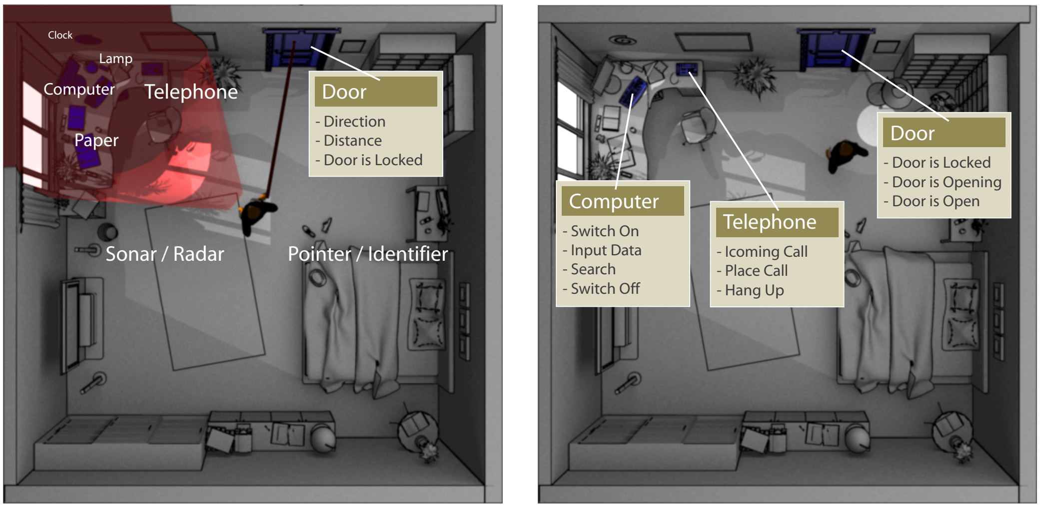
Figure 30: Global 3D Scene Sonification Techniques.

position in the environment, like doors, escalators, stairs or teleporters. This also includes the ground floor, whose auditory description provides additional information about the material it is composed of. Transitional sounds can be employed to describe the passing through a portal in more detail. This includes changes in elevation, as well as the passing through a door. *Interactables* are objects with an added functionality that the user can explore through interaction. These objects often exist in different states/conditions that changes upon interaction. The TV set, or the computer in Figure 30a can be, for instance, switched on and off and used – in the setting of an auditory adventure game – to search for additional story related information and hints. A detailed description of scene objects along with their various states and interactions possible can be achieved using *Auditory Textures*, which are discussed in the following section. The group of *Obstacles* describes barriers that interfere with a free exploration of the 3D environment. The only interactions possible are collision and obstruction. Object bound sounds can be employed to identify obstructions with an increasing volume for an approaching barrier.

Auralization of Figure 30b.

Several techniques are available to support a user's navigation and orientation within 3D virtual auditory environments. Newly developed techniques include so called *Soundpipes*, *Guide*-based systems, as well as an *Auditory Lens* (Röber and Masuch, 2004b, 2006). All these systems are able to connect different areas in a 3D environment, and permit a listener an easy *traveling* in between. The most direct approach is the *Soundpipes* implementation, in which moving sound sources guide a listener – similar to a public transportation system – through the auditory environment. Figure 30b shows a visualization of this principle using the common scene setting. The sound example on the left sonifies this setting from the user's perspective. One can hear the soundpipe moving along its path, with the additional objects (blue spheres) activated and displayed through their respective auditory icons. In this example, a soundpipe spans the entire room from the listener's position to the door, and features additional sound sources (blue spheres) along the way to highlight interesting objects (blue) in close vicinity. These objects are now easy to reach without the possibilities of getting lost.

A very efficient form of user assistance can be implemented using *Guiding* systems. The simplest approach uses so called *attracting* and *repelling* sounds, which either draw



(a) Sonar/Radar and 3D Object Identifier.

(b) Three Auditory Textures.

Figure 31: Local 3D Scene Sonification Techniques.

a listener towards a certain location or away from it. The sounds themselves are implemented as hearcons and are chosen depending on the task and content of the auditory environment. Attraction and repellence can be implemented using pleasing/unpleasing sounds or through harmonic/disharmonic music, listen to the sound example on the right. A more direct approach for a guiding and assistance system is an artificial avatar that helps and directs the listener in difficult situations using speech samples. Such a system should be based on speech synthesis to cover the largest range of actions possible. The scope of such a versatile assistance system includes the entire setting, and ranges from a global navigational aid to a local assistance system for close object examinations.

Sound and Music Guides.

An *Auditory Lens* can be employed for either global/environmental sonifications, or for local, close-object examinations. The auditory lens is constructed as a hearing frustum that only permits sounds from a given direction and distance, see also Figure 34b. This technique is therefore well suited also for a precise examination of a local 3D virtual auditory environment. Using an extension, the technique can be modified to only allow certain types of sounds to be audible, eg. to highlight all interactables within a scene, or to only display environmental sound sources.

Local Sonification Techniques

Whereas global scene sonification techniques are employed to portray an overview of the auditory environment, local sonification techniques are used to examine single objects to denote their function and current state. Examples can be seen in Figure 31, which shows in Figure 31a several sonification technique for an examination of the local surroundings, and in Figure 31b a visualization of the *Auditory Textures* approach.

The *Sonar/Radar* technique, which is depicted in Figure 31a, is inspired by the echolocation of bats and other animals (Surlykke and Kalko, 2008). The example visualized is based on an interaction through head-orientation and/or the use of a 3D pointing device. Both are based on the commonly used *Magic Wand* devices, which are used in virtual reality environments to interact with and to manipulate 3D virtual objects (Bowman and Hodges, 1997; Ciger et al., 2003). In this adaptation, these techniques are employed for sonifying information and to find and identify objects in the local environment. The

Auralization of Figure 31a.

sonar technique, for instance, identifies objects in the listeners *field-of-view* through either the use of speech or descriptive sounds, and groups them according to their distance and relative position using hearcons. The object identifier is based on direction and 3D pointing, and highlights objects that are detected, such as the door in Figure 31a. If objects of interest are detected, a listener can perform a further exploration and/or interaction using *Auditory Textures*. The sound sample on the left demonstrates three examples. The first one shows the auditory radar, in which the five focussed objects in the top left of Figure 31a are acoustically displayed and distance encoded using loudness. The second example demonstrates the Sonar/Cane approach, which is here used to find interactable object, while the last example displays an auditory texture for the selected door object.

Auditory textures are very similar to the texture approach used in computer graphics, and are used to acoustically describe an objects function and state. The approach described here is based on the work of Mynatt, who nested symbols within symbols to acoustically represent different states of menu items in an auditory user interface (Mynatt, 1992). The example in Figure 31b displays three auditory textures, for the door, the telephone and the computer on the desktop. As can be seen in these examples, auditory textures are basically a collection of different sound files, which describe the object in different states and for its various forms of interaction:

- A general descriptive object sound,
- Several action and/or status changed sounds,
- A call sign for the sonar/interactor,
- A speech-based description.

The general sound is the standard acoustic representation for an object in its normal state, usually a descriptive auditory icon, refer to Figure 32. Action and status changed sounds are used to characterize a current activity or changing situation for this object. These sounds are eventually played only once, eg. clicking a button, but can also activate a different general representation for this object, eg. changing a state like switching on the computer. The sonar/interactor call signs and the verbal description are two additional representations that are used for further description and to identify the object. These are activated on request and only played once. The sound selected depends on the listener's interaction, as well as on the content and state of the 3D auditory environment. In the setting of a 3D auditory adventure game, a second layer, such as a story/game engine, can trigger events and change an object's auditory description by selecting an alternative representation to control the story and to advance the gameplay. The sound sample on the left sonifies example auditory textures for the objects depicted in Figure 31b. Each auditory texture is played twice. The first example that is heard denotes the computer object, and is composed of four auditory icons that

Auditory Texture

General Sound	Silent or auditory Icon (ringing)
Status changed	Incoming call (loud ringing)
Status changes	Broken (auditory icon)
Action	Pick up phone (auditory icon)
Action	Dialing (beeps)
Action	Talking (silent)
Action	Ring off (auditory icon)
Radar call	Auditory icon (ringing)
Verbal description	Speech: "Telephone"

Figure 32: Auditory Texture for a Telephone.

Auralization of Figure 31b.

display a possible interaction: *switch on*, *input information*, *search* and *switch off*. Note the additional earcons ahead or behind the auditory icon which closer denote the type of action. The second example displays three icons for the telephone object: *incoming call*, *place call* and *hang up*. The last example demonstrates the auditory texture for the door object. The first sample displays a locked door, but also denotes a possible interaction. The other two examples describes the actions/state: *door opening* and *door is unlocked*.

Summary

The last sections discussed several techniques for a sonification of 3D virtual auditory environments. A focus in these discussions was to develop techniques that can be used to convey global/environmental and local/object information to a listener by solely using auditory means. The discussed global 3D scene sonification techniques are:

- *Hearcons* for the identification of 3D objects, their position and orientation
- *Interactables (Object Grouping)* to classify types of objects/areas in a scene
- *Auditory Landmarks* to highlight specific objects and to improve the user's navigation and orientation
- *Soundpipes* for an efficient and intuitive navigation in large scenes and environments
- *Guiding Systems (Voice and Music)* to guide the user using speech and attractive/repelling sounds
- *North Beacon* to improve a user's orientation in a complex 3D scene
- *Auditory Lens* to focus on a specific area and on specific object types

The discussed local sonification techniques are:

- *Radar & Sonar* to identify objects and their position in a local environment
- *Magic Wand (Cane)* to find, identify, select and interact with the local environment and specific scene objects
- *Auditory Textures* to acoustically denote the various states/functions of a scene object

The border between global and local sonification is not fixed, and several of the here developed techniques can be employed for both tasks, refer also to Table 4.

After the discussion of 3D scene sonification techniques, the next step is a combination with an intuitive audio-centered interaction design that enables a seamless integration of sonification and interaction techniques within a 3D virtual auditory environment framework.

5.4 Interaction Concepts

3D virtual auditory environments represent complex and dynamic 3D spaces that require techniques for an efficient and intuitive interaction. The goal of this section is to develop interaction techniques $C(t)$ for a selection and control of enhanced 3D scene objects $\mathcal{M}$. A selection, as well as an interaction with an object highlights and displays specifics of its symbolic information, eg. $D(t) = \mathcal{M} \times s(t), s \in O_S$. This section concentrates on the techniques $c \in C$ to select and activate this information:

$$\{\psi | \psi(c, o) = c \times o, c \in C, o \in \mathcal{M}\} \tag{5.8}$$

Technique	DOF[5]	Usability (poor/low (1) – great/high (5))	Mobility	Areas of Application
Keyboard	1	3	2	Movement, Orientation, Navigation, Symbolic Input
Mouse	2	3	2	Movement, Orientation, Navigation, Symbolic Input, Gestures
Gamepad	3	4	5	Movement, Orientation, Navigation, Symbolic Input
Space Mouse/Navigator	6	3	2	Orientation, Navigation, Movement
Head-Tracking	6	5	3	Orientation, Navigation, Gestures
3D Pointing (Stylus)	6	4	2	Selection, Activation, Pointing, Gestures
3D Interactor (3Ball)	6	4	2	Selection, Activation, Gestures
Video (Webcam)	2	3	2	Selection, Gestures
Accelerometer	1(3)	4	4	Movement, Orientation, Navigation, Gestures
Speech	1	3	5	Selection, Activation, Symbolic Input

Table 5: 3D Scene Interaction Devices and Techniques.

Bowman et al. group the possible interactions with virtual environments into three layers (Bowman et al., 2004):

- *Navigation* is described as moving and wayfinding, in which moving is divided into exploration, search and the navigation of rooms and areas, while wayfinding is the cognitive component that uses the perceived information to derive a specific path for traveling.
- The *Selection* and *Manipulation* of objects allows to interact with single objects, to change their appearance, to influence the environment, as well as to derive further scene/object information.
- *System Control* and *Symbolic Input* both describe the uppermost layer of the environment and are used to change the user interface and to input abstract symbolic information.

Several of these points have already been addressed in the previous sections, focussing on an efficient sonification to convey the information required for a performance of these tasks. The type of interaction is thereby dependent on the content and the task of the virtual (auditory) environment. A narrative application, such as a 3D computer game,

5 DOF = Degree of Freedom

requires a different interaction than a virtual reality training simulation, or a guiding system for the visually impaired.

An important requirement for the design of interaction techniques is that they are founded in a natural behavior that mimics an interaction with the environment and objects in the real world. In awkward listening situations, for instance, humans tend to slightly tilt their head to perceive new spectral listening cues that allow a more precise estimate of a sound source's direction and distance (Gaye, 2002; Goldstein, 2007). This interaction enables one to determine the origin of a certain sound, at which direction in a second step other senses are focussed to gather more information. Based on such observations, the interaction design required for 3D virtual auditory environments can be defined as:

Definition *3D Auditory Scene Interaction* describes a set of methods and techniques which complement *3D Scene Sonification*, and allow a listener to interact with a 3D virtual auditory environment. The task of these interactions is to select and manipulate specific 3D objects, to perform moving and navigation operations, as well as to input abstract, symbolic information. The techniques adhere to a natural listening and interaction behavior and are based on real-world spatial interactions. The two primary components are 3D user head-tracking, as well as 3D point and selection techniques to perform spatial interactions based on 3D gestures. Secondary interactions comprise of speech recognition and common 3D scene interaction and are based on techniques used in 3D computer games and in entertainment applications.

An overview of several applicable interaction devices and techniques can be found in Table 5. Table 5 presents a variety of interaction devices and assesses their potential with a discussion of their degree-of-freedom (DOF), their usability and mobility, as well as through their possible areas of application. The following three sections examine these interaction techniques in more detail, with a focus on the implementation of 3D spatial interaction techniques to improve the 3D scene sonification methods developed in Section 5.3.2. The majority of the discussed spatial interaction techniques are realized and implemented using a 6 DOF Polhemus FASTRAK (Polhemus, 2008). The items *Stylus* and *3Ball* in Table 5 are a cylindrical, respective a sphere-shaped, 3D input device for the Polhemus FASTRAK, refer also to Figure 33c.

5.4.1 *Common 3D Scene Interaction*

A basic interaction with 3D virtual environments is required in many applications, such as in 3D computer games and edutainment systems that are based on VR technology. These applications are designed for either desktop-based computer systems, or for (mobile & transportable) gaming consoles. For the interaction with 3D environments, such as computer games, guidelines exist and common interaction patterns have evolved (Salen and Zimmerman, 2003). A play and interaction with these environments is generally performed using either a keyboard/mouse combination, or by using a gamepad or joystick. A gamepad is thereby just a modern implementation of a classic joystick and consists of several analog/digital joysticks/bars, as well as a number of free programmable buttons. Both forms of interaction can very well be employed to control 3D virtual auditory environments as well, and utilize the user's previous experiences with regular audio/visual virtual environments, ie. computer games. Whereas the control of a character via a keyboard/mouse combination is simple and easy to implement, this approach is not suited for a mobile implementation and also restricts the play and interaction to be in front of a computer screen. A gamepad, on the other hand, can be easily integrated into mobile de-

(a) 5DT Data Glove[6]. (b) 3D Space Navigator[7]. (c) Polhemus FASTRAK[8] (Polhemus, 2008).

Figure 33: Spatial Interaction Devices.

vices and also separates the gameplay away from the computer, thus *leaving the screen* for an interactive play within 3D virtual auditory environments (Röber and Masuch, 2005a).

Both devices, the keyboard/mouse combination and the gamepad, are employed and used to interact with 3D auditory environments and are evaluated and assessed among other forms of interaction in Section 9.3. However, as this form of interaction limits the possibilities of 3D auditory environments, the next section focusses on a more audio-centered 3D spatial interaction approach.

5.4.2 *Spatial Interaction*

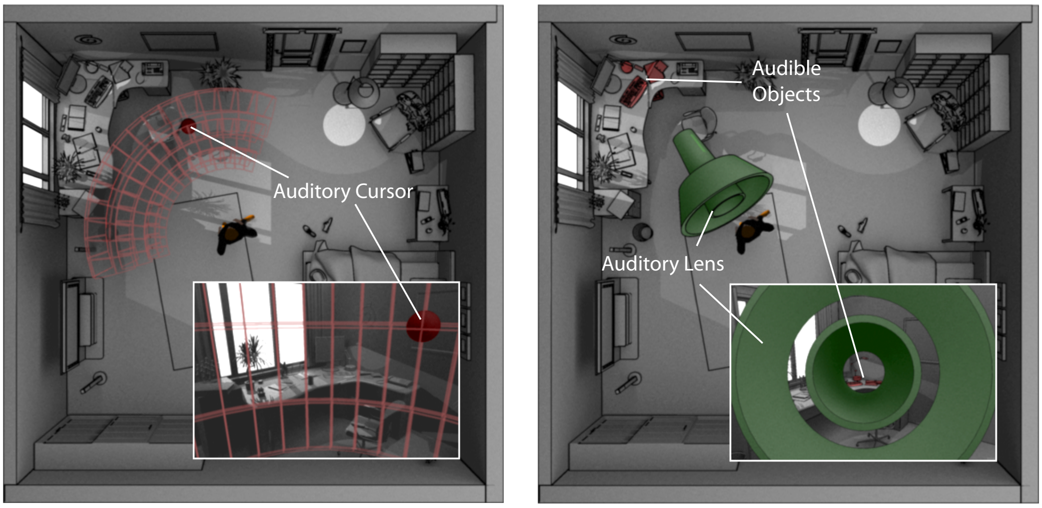
A spatial interaction design that mimics a real-world interaction is in many cases more desirable, as it more intuitively maps the interaction onto the virtual environment:

> *"The quality of the interaction technique that allows us to manipulate 3D virtual objects has a profound effect on the quality of the entire 3D user interface." (Bowman et al., 2004)*

While the interaction with virtual game environments did not change over a long period of time, the last years have seen several new interfaces based on an unconventional and more natural interaction behavior. Two of the most influential examples are Sony's *Eyetoy* interface, as well as Nintendo's new game console the *Wii* (Sony Entertainment, 2003; Nintendo Europe, 2006, 2007). Both vendors employed new hardware to devise new interaction paradigms. In the case of the Eyetoy system, a webcam was added and allowed a direct and full body interaction with the entire game world (Sony Entertainment, 2003). The Wii console additionally integrated several accelerometers and an infrared camera system that allows a basic user tracking and positioning, as well as an interaction using gestures (Nintendo Europe, 2006, 2007). Further advantages of this hardware are that it is available for a very low price and that it can be easily integrated into own projects and within non-gaming applications (Lee, 2007; Seznec, 2007).

Other hardware examples that allow an implementation of spatial interaction metaphors can be seen in Figure 33. Figure 33b displays a so called space mouse/navigator that provides 6 degrees of freedom and which is mainly employed for navigating and manipulating 3D virtual environments. A more flexible application permit so called 3D tracking devices, which are depicted in Figure 33a and Figure 33c. The technology allows a very precise position and orientation tracking using magnetic sensors, as well as permits an

6 http://www.3dconnexion.com
7 http://www.5dt.com
8 http://www.polhemus.com



(a) Auditory Cursor and 3D Interactor. (b) Auditory Lens.

Figure 34: Spatial Interaction Techniques.

implementation of gesture recognition and other forms of 3D spatial interaction (Röber and Masuch, 2004b). Head-tracking is a very important form of interaction as it simulates a natural listening behavior and thereby allows a deeper immersion into the auditory space (Gaye, 2002). In the implementation of the here discussed interaction techniques for 3D virtual auditory environments, head-tracking is employed in all applications, as it improves the perception of auditory environments by a large factor.

A few of the earlier introduced 3D scene sonification techniques are perceived passively and do not require any interaction. However, the majority relies on and benefits from an added spatial interaction, and allows to perceive the information displayed more intuitively. Some examples for spatial interaction can be seen in Figure 31a and Figure 34. These images show visualizations of 3D scene sonification and interaction techniques using the familiar living room environment. The here depicted examples are implemented using a magnetic field tracking system (Polhemus FASTRAK), refer to Figure 33c. This system supports up to four sensors, which are employed to perform user head-tracking, as well as for 3D spatial interaction.

Auralization of Figure 34a.

Figure 34a shows the example of an *Auditory Cursor*, which is basically an extension of a regular computer cursor and based on a 3D pointing technique (Röber and Masuch, 2004b). Technically, the auditory cursor is aligned along several spheres that are centered around the listener. The cursor itself is represented through a 3D sound object (hearcon) that snaps onto the grid and can be moved along for an intersection and object selection (red dot in Figure 34a). The cursor's direction is clearly audible using sound spatialization, while the cursor's depth is encoded through pitch and loudness variations (auditory depth cuing), listen to the example on the left. The direction of the auditory cursor is determined using the direction of the 3D pointing device (3Ball) relative to the user's position (second sensor is mounted on user's head for 3D head-tracking). The distance can be input using a second interaction device, but is in this implementation mapped to an additional button on the 3D interactor, which, if pressed, changes the auditory cursor's depth.

Another interaction technique is visualized in Figure 34b, which highlights a so called *Audio Lens*, or hear-frustum. This audio lens enables the user to perceive selected sound

Auralization of Figure 34b.

sources only. The selection process is based on direction, radius and distance, but can also include the various types of sound sources available, eg. object sounds, environmental sounds, beacons and so forth. The direction and distance of the auditory lens can be specified in a similar way as for the auditory cursor in the last paragraph, using a tracking sensor that is equipped with an additional button. The example on the left sonifies the concept. In the first half, all objects are audible from the listener's position depicted in Figure 34b, while in the second half, the auditory lens is activated and *zooms* the depicted sound objects into focus.

Auralization of Figure 31a.

Figure 31a shows two visualizations, of which one is an implementation of a blind man's cane, while the other demonstrates the principle of a radar/sonar-based 3D scene sonification. The interaction with the cane is similar as with the auditory cursor, and is based on a 3D pointing technique. If an object is in the direction of the 3D pointing device, eg. the door in Figure 31a, the door object identifies itself using an auditory icon, or a short verbal description. The sonar/radar technique can be employed in a multitude of ways. Similar to a magic wand device, it can be used as a flashlight to highlight objects in a certain direction of the scene. The *lit* objects would reveal themselves in a similar fashion as to the white cane, but distance encoded, as is visible in Figure 31a. The acoustic representation of the wall clock is less pronounced than the auditory icon of the nearby telephone, listen to the example on the left. One can also devise a sonification technique that is based on a *real* echolocation, in which the interaction device is used to *emit* a high pitch sound that is reflected by the scene's objects. As this application requires a very sophisticated simulation of sound wave propagation, the discussion is postponed till Chapter 8.

An advantage of spatial interactions is the possibility to mimic a real-world interaction behavior. One example are gestures, which can be utilized for an interaction with 3D auditory environments in several ways. Both, the head-tracking system, as well as the 3D interaction devices can be used for an implementation of gestures. Gestures for the head-tracking system can be reduced to very basic interactions, such as nodding and negation, whereas a 3D interaction device allows the implementation of much more complex 3D gesture movements. These gestures can be used to interact with virtual objects in a natural way, as well as to change parameters within a virtual menu system. Such a menu can be implemented using a ring-topology, in which the menu is arranged and centered around the listener. Several auditory widgets can be employed, and are intuitively to operate using 3D interaction devices and the techniques described.

Another interesting interaction is the use of force-feedback systems, which can be employed for a redundant and multi-modal presentation of 3D scene information. An example is the use of a so called force-feedback headphone system, which in the context of its realization can also be classified as sonification technique (Evergreen Technologies, 2005). The headphones employed look very similar to regular headphones, but vibrate and rumble at lower frequencies. This feature can be employed to sonify object collisions and to drag the listeners attention to certain locations, therefore be used for highlighting specific environmental and 3D object information.

5.4.3 *Speech-based Interaction*

A speech-based interface provides a very intuitive and also very flexible form of communication, and can be used in certain cases as an alternative for the interaction with 3D virtual auditory environments. Using speech synthesis and recognition, both channels, ie. the output and input of information, can be controlled using speech. Dedicated technology exists for both applications and can be easily employed in own implementations.

Whereas speech synthesis, ie. text-to-speech, has made large improvements over the recent years and is able to synthesize natural sounding voices and sentences, the recognition of human speech is still often deficient. Furthermore, every speech recognition system requires additional conditions, such as a previously trained voice and a silent environment, which makes their general application and use still difficult (Wendemuth et al., 2004). However, if only a few words have to be recognized that can be segregated in recognition space, a speech-based interface can well be employed within these conditions. A drawback, however, is that an excessive use of speech can be tiresome, especially for tasks that are not as suited for a speech control.

A more natural application for speech recognition and synthesis is the communication with other avatars in a virtual environment, in which speech recognition and synthesis can be used in a similar way as for real-world communication. Speech synthesis itself, however, has more applications and can be employed in a broader scope. It can also be used for a speech-based summary of a 3D scene and the objects therein, therefore covering the part of verbal descriptions within the concept of auditory textures. Other applications include the control of text-based adventure computer games, to make them accessible for the visually impaired (Malyszczyk and Mewes, 2005; Atkinson and Gucukoglu, 2008).

5.5 FRAMEWORK DESIGN

The next step after the layout of 3D virtual auditory environments along their associated 3D scene sonification and interaction techniques is a discussion of possibilities for an actual implementation. This is the focus of the following section, which first summarizes the requirements for the design of such a system and afterwards discusses the development of an audio framework highlighting important design essentials. The goal in the design of this framework is an implementation and evaluation of the previously introduced sonification and interaction techniques, as well as an exploration of possibilities to devise an immersive and convincing 3D virtual auditory environment.

5.5.1 *Design Essentials*

Some of the key components for the design of an interactive 3D sonification framework were already outlined in Section 2.2. The research of this chapter, through the examination of possibilities and applications for 3D virtual auditory environments, provided a deeper understanding of the requirements and the acoustic modeling and design required for 3D auditory spaces. In order to perceive enough information to aid the user's orientation, navigation and interaction, a 3D auditory environment must exhibit certain qualities and techniques. The components for the design of a multiple-applicable audio framework are therefore:

- A 3D (polygon-based) virtual environment managed by a scenegraph system.
- Possibilities for a 3D graphics-based visualization of the 3D scene.
- A 3D audio engine that supports a non-realistic auditory design.
- 3D sound rendering and binaural display techniques.
- Dedicated methods for 3D scene sonification and spatial interaction.
- Intuitive user-input and interaction devices, with
  - User/Head-tracking capabilities, and

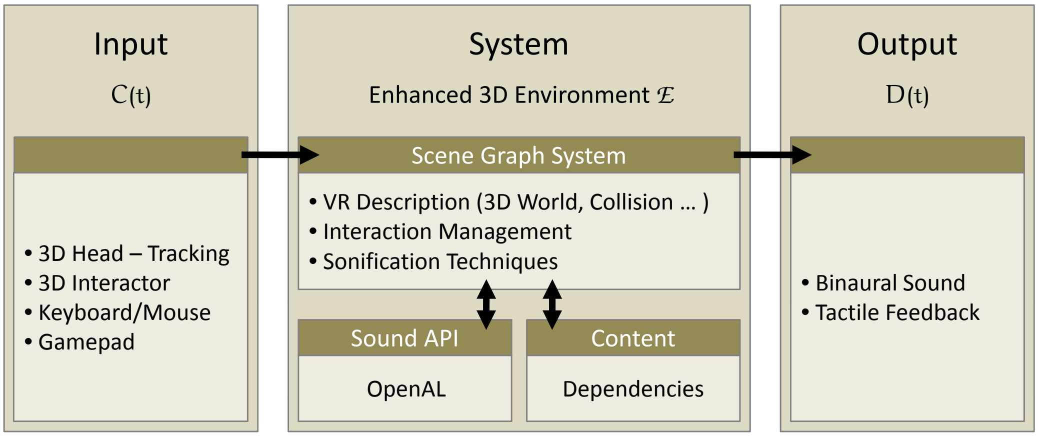


Figure 35: Framework Overview.

  – 3D pointing and interaction devices.

- As well as an authoring environment for the design and setup of 3D virtual auditory environments.

An overview of the designed system can be seen in Figure 35. The system has been envisioned with the targeted areas of application in mind and in analogy to the design of regular 3D game engines used for the development of 3D audio/visual computer games (Boer, 2002b; Salen and Zimmerman, 2003). The core component is a scenegraph system for the representation of the enhanced environment $\mathcal{E}$. The scene sonification and interaction techniques can be implemented on top of this system and be used in conjunction with a non-realistic auditory display for the scene. Major hardware requirements include tracking capabilities to measure the user's head orientation, but also to implement the previously described spatial interaction techniques:

- A PC system equipped with 3D sound hardware (Creative Labs X-Fi),
- A tracking system (Polhemus FASTRAK),
- A wireless gamepad for default interactions,
- A microphone for speech input, as well as
- Regular (force-feedback) headphones.

The devised system is based on standard PC hardware, as this allows an easier development and evaluation of the entire system and the single techniques. However, at several points throughout this research, alternatives that allow an implementation on mobile hardware and devices are discussed (Stockmann, 2007). Although it was shown that regular sound hardware has certain difficulties and inefficiencies with 3D sound spatializations and the simulation of room acoustics, it serves as an initial basis to gather first results. For the tracking of the user's head-orientation, as well as for the performance of spatial interactions and 3D gestures, a Polhemus FASTRAK system is employed. This is a 6 DOF magnetic field-based tracking solution that allows the use of up to four independent sensors (Polhemus, 2008), see also Figure 33c. The final binaural sound display and rendering is performed using regular HiFi headphones. The headphone systems employed consist of a regular high-quality HiFi system (*Hear0999 Audiosphere*

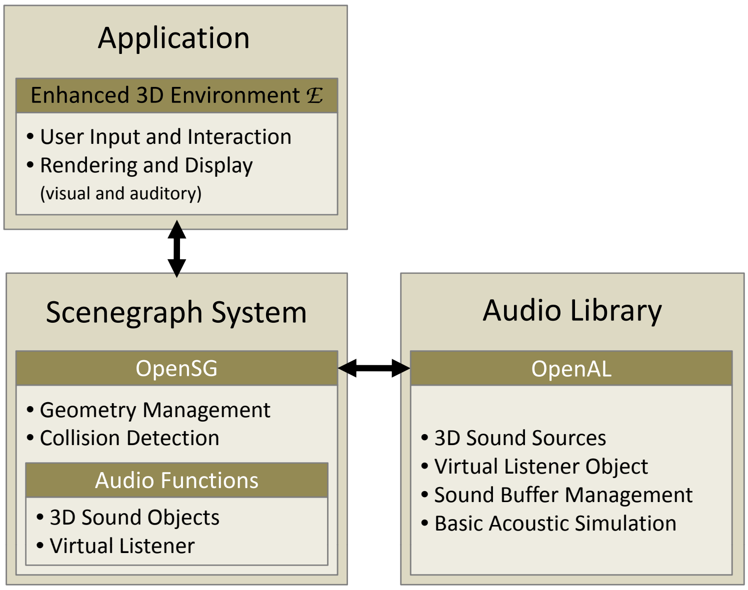


Figure 36: Framework Implementation.

(AKG Acoustics GmbH, 2008)), and a so called force-feedback headphone solution that rumbles at low frequencies (Evergreen Technologies, 2005).

5.5.2 *Implementation*

The major components of the audio framework devised can be seen in Figure 35 and Figure 36, which display the scenegraph system along its modules as the core of the system. The implementation and the design of the audio framework is centered around OpenSG, a modern 3D scenegraph system that has many applications in research and science, as well as in entertainment solutions (Reiners and Voss, 2008). OpenSG already features many requirements for an implementation of 3D virtual environments, and therefore provides a good starting point for an implementation of 3D virtual auditory environments. Similar to other scenegraph systems, OpenSG uses the common description languages VRML and XML, which both provide the grammar and the descriptors required to define and model 3D virtual auditory environments. VRML allows the integration of audio nodes within the definition of environmental geometry and 3D objects. These audio nodes are easy to implement, as well as can be extended for a more precise description and modeling of 3D auditory environments (Hoffmann et al., 2003), see also Listing 5.1. OpenGL, which is directly integrated into OpenSG, is used to visualize the content of 3D auditory environments for the purpose of analysis and control.

The devised audio framework is implemented using C ++ , and as already mentioned, centered around an extension of OpenSG. The extensions include sound rendering capabilities, collision detection, as well as an implementation of auditory textures along the previously introduced sonification and interaction techniques. An overview of the implementation can be seen in Figure 36. The audio framework (OpenSG) controls and calls all connections to the audio API (OpenAL) employed. Several audio functions have been implemented into OpenSG and can be connected to the scenegraph to represent virtual 3D sound sources and listeners. The audio API is thereby disconnected from the core system, as can be seen in Figure 36. This allows to substitute the sound rendering API by a more capable system, refer to the discussions of Chapter 8.

The framework also includes visibility tests for a very basic modeling of room acoustics, as well as an implementation of the in Section 5.4 developed interaction techniques. These techniques utilize an external 3D interaction device (Polhemus FASTRAK), which returns the position and the orientation of the employed sensors. This data is used to determine and implement the 3D user head-tracking, as well as to perform a virtual object picking and selection within the virtual scene. On top of these interactions, all 3D scene sonification and interaction techniques are implemented.

The modeling and design of 3D virtual auditory environments can be performed using tools such as 3DStudioMax, from which the geometry is exported and stored as VRML data file. Sound nodes, as well as auditory textures are integrated into this scene description and loaded later into OpenSG. An update of the 3D virtual auditory environment within OpenSG performs now a visual, as well as an auditory rendering of the 3D scene. Listing 5.1 shows an overview of the integration of audio nodes within VRML objects (Hoffmann et al., 2003; Walz, 2004a). This definition provides several virtual speaker parameters, such as direction, position, intensity and distance attenuation.

Besides an integration of audio nodes into the scenegraph environment, also the system's connection with user interaction techniques $C(t)$ and a development of auditory display styles $D(t)$ is required. The definition and setup of the enhanced 3D objects $\mathcal{M}$ requires a mapping of scene geometry E_G with structural scene information E_S and symbolic information O_S, refer to Section 5.1. Much of this data can efficiently be implemented using auditory textures, for which now the concept of dependency modeling is introduced.

The dynamics of 3D virtual auditory environments as well as the animation of objects can be specified using dependencies and an implementation using auditory textures (Deutschmann, 2006), refer Section 5.3. The most important dependencies for the modeling of a dynamic environment are:

- Position Dependencies,
- Object Dependencies, and
- Time Dependencies.

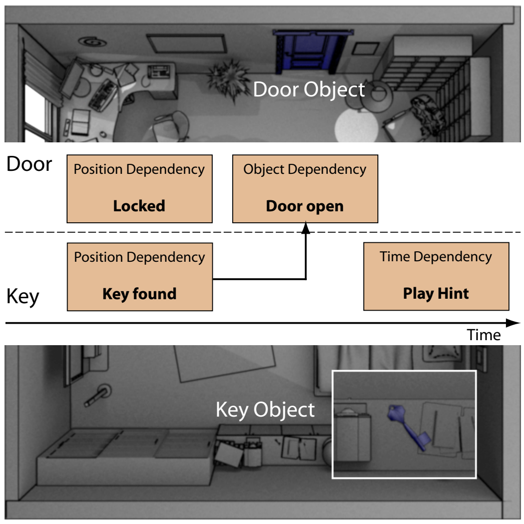


Figure 37: Interaction Dependencies.

A *Position Dependency* describes the influence of the listener's position onto surrounding objects within the virtual auditory environment. Depending on the listener's position, certain objects are audible, while others are silent. A position dependency is described through a threshold ϵ – the distance to a certain object in the scene. If the listener approaches this object and if the distance is below this threshold ϵ, a certain action is evoked. Position dependency is very general and applicable in many forms, most notably for augmented audio reality applications, which will be introduced and discussed in Chapter 6. An *Object Dependency* describes an inter-object dependency, in which one object can change the state and appearance of another object or group of objects. This dependency is very helpful for using additional *dummy* objects to model dependencies based on the connection of the environment with other secondary systems. A *Time Dependency* is simply a counter that triggers an event after a certain amount of time has lapsed. This dependency can also be used for the modeling of additional narratives in the form of a state machine. Figure 37 displays a simple dependency graph for the door and the key object. In this familiar setting, a small part of a virtual 3D auditory computer game is described. The task of the user in this example is to find a missing *Key* object to unlock and open the *Door*. Both objects are described through a position dependency, which, if the user approaches, display their current condition. Here the door appears as being *locked*, while the key identifies itself as key to open the locked door. The unlocking itself is performed using an object dependency that unlocks the door after the key is found. If the user has difficulties finding the key, a hint is played after a certain amount of time trough an added time dependency. The concept of dependency modeling allows a very broad design and implementation of various auditory environments.

```
DEF Orgel Transform {
    (...)
    Sound {
        direction 0.0 0.0 1.0           // source direction
        location  0.0 0.0 0.0           // source position
        intensity 1.0                   // source intensity
        minBack 1.0                     // distance attenuation 1
        maxBack 5.0                     // distance attenuation 2
        minFront 1.0                    // distance attenuation 3
        maxFront 40.0                   // distance attenuation 4
        spatialize true                 // turn on spatialization
        source DEF ogrelclip AudioClip {
            description "BWV 1080"      // sound description
            url "sound/BWV-1080.wav"    // load sound data
            loop FALSE                  // loop on/off
            }
        }
    (...)
}
```

Listing 5.1: Definition of VRML AudioNodes (Hoffmann et al., 2003).

As the framework is classified as a 3D auditory display system, techniques for the spatialization of sound, as well as for the playback of speech and music are imperative. The qualities of the playback, sound spatialization, as well as for acoustic simulations are thereby of the highest importance. Although currently available audio APIs still have several limitations, they can, nevertheless, be employed in an initial prototype to evaluate the devised techniques. In this framework, OpenAL was chosen due to its large availability and an active development community (OpenAL, 2008). Although OpenAL works properly for the majority of audio/visual applications, some listeners experience difficulties in the localization of virtual 3D sound sources. In addition to OpenAL, a mobile DSP oriented sound API has been developed and can be employed for a spatial sound rendering on multiple portable devices as well (Stockmann, 2007; Huber et al., 2007).

Another key component of the system is the connection to the various interaction devices that are employed, see also Figure 35. A low latency of the tracking equipment is thereby essential, and must be below 80ms for the head-tracking system, as otherwise perceptional artifacts occur (Brungart et al., 2005). The implemented spatial interaction techniques and 3D gestures also require a low latency ($\leqslant$ 200ms). For an application towards music and the design of virtual instruments, the latency must be as low as 30-40ms, as otherwise a continuous and focussed play is difficult to achieve (Stockmann, 2008; Stockmann et al., 2008). For the systems connection to the tracking equipment, the VRPN library was employed, which supports several VR solutions and is controlled via a network system (Taylor II et al., 2001, 2008), refer to Figure 70. The latency of this API is low enough for a regular interaction with 3D auditory environments, but required a re-implementation due to an application for the design of virtual computer music instruments (Stockmann et al., 2008). The additional gamepad, as well as the keyboard and mouse connections to the system have been implemented in a straightforward manner using the Direct Input API from Microsoft[9].

5.5.3 *Areas of Application*

Similar to auditory display systems, the areas of application for 3D virtual auditory environments are manifold and quite diverse. This section briefly introduces some of the more interesting areas, while Chapter 9 discusses and studies them in more detail using a variety of user evaluations.

The areas of application for 3D virtual auditory environments that are considered within this research are:

- The exploratory analysis and sonification of abstract 1D, 2D and 3D data sets.
- Audio-centered entertainment applications, such as audio-only computer games.
- Auditory narration, in a combination of audiobooks and computer games.
- Enter-, Edu-, and Infotainment scenarios for
  - Guiding systems and training simulations for tourists and the visually impaired.
  - Augmented audio reality applications.

The sonification and data mapping techniques that were discussed in Section 5.3.1 can be well employed for an auditory display of abstract 1D, 2D and 3D data sets. Clear advantages for using sound to convey data values are a non-focussed perception, spatialized presentations 360° around the user, as well as a simpler implementation using less rigid hardware requirements. An extension of existing graphics-oriented visualization systems towards a multivariate – audio/visual – data display thereby allows to enhance the perception through an added redundancy and the perception of information over multiple channels. Section 9.2 discusses this approach as well as several examples for an acoustic presentation of stock market data, 2D and 3D shapes, images, and 3D volumetric data sets.

As the research in this thesis is conducted in close proximity to entertainment applications, the framework developed is highly suited for a presentation of entertainment content in the form of audio-only and mixed reality computer games. The scenegraph system can be used to represent 3D environments as the fundamental basis of a virtual world, as well as allows the integration of 3D scene sonification and spatial interaction techniques. The majority of these techniques are evaluated in Section 9.3 using small examples and a user evaluation. Section 9.4 takes a closer look in the direction of audio-only computer games, and compares several existing audiogames with the possibilities that are facilitated through this framework. This section discusses three action games, as

9 http://msdn.microsoft.com/en-us/directx/default.aspx

well as one auditory adventure, which are all based on the audio framework developed and which employ spatial sonification and interaction techniques. Another advantage of auditory displays is the possibility to achieve a high level of immersion in narrative presentations. Therefore, Section 9.6 explores a new form of interactive narrative called *Interactive Audiobooks*, which combines audiobooks and radio plays with interactive elements form computer games.

The framework designed is not only applicable to entertainment scenarios, but can be employed in a number of *serious* applications as well, such as in guiding and training simulations for tourist and the visually impaired. The differences between an entertainment and a serious application are marginal, and primarily reside in the authored content, as well as in the user interface designed and the methods of interaction and scene sonification that are used. Several aspects of these areas are examined in Section 9.3 and Section 9.5, which both focus on a general application of sound and acoustics to improve everyday routines and processes. Section 9.5 thereby evaluates specifically the potential of augmented audio reality for aiding the visually impaired, as well as for an implementation of an augmented audio computer game.

5.6 SUMMARY

This chapter discussed and defined 3D virtual auditory environments in the context of virtual reality and 3D auditory display systems using an abstract definition of VR/MR environments. This *new* definition focusses on an audio-centered design and employs a non-realistic auditory scene description. Starting with methods for 2D/3D data sonification, a number of 3D scene sonification and 3D spatial interaction techniques were devised and discussed, including the concepts for an auditory cursor, -guides, -landmarks, -lens, a sonar/radar and a soundpipes system. Additionally, the concepts for a dependency modeling and an auditory texture were developed, as well as an audio framework conceptualized to implement the approaches discussed.

The following Chapter 6 continues the previously started discussion on augmented audio reality, and extends the in this chapter designed framework towards a 3D augmented audio display system.

6 AUGMENTED AUDIO

Each thing we see hides something else we want to see. — René Magritte

AUGMENTED Audio Reality (AAR) describes a system that *enhances* the acoustics of a real world location through the display of additional auditory information, thus allowing others to see things as *we* want them to be seen. Augmented audio is an extension and a special case of the 3D virtual auditory environments that were discussed in the last chapter. The focus of this chapter is a detailed analysis of augmented audio reality in terms of acoustic presentation, listener immersion and the development of techniques that enable a perception of both environments as one. The goal is to transfer and adopt the 3D scene sonification and spatial interaction techniques developed in the last chapter for an application within 3D augmented audio reality scenarios. Therefore, the chapter first reviews the fundamentals of mixed reality applications and defines the principles and requirements of 3D augmented audio reality. In a second part, the chapter develops and implements spatial and location-aware interaction techniques. These techniques are integrated within a self-designed low-cost augmented audio system that is based on an extension of the framework developed in the last chapter.

6.1 AUGMENTED REALITY

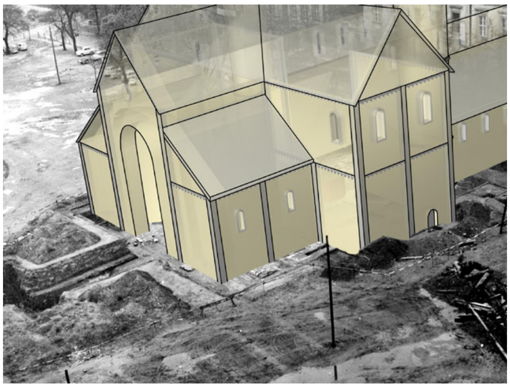
Section 5.1 introduced the term 3D virtual auditory environment and discussed it within the context of virtual reality and the virtuality continuum. This continuum describes the varying degree of virtuality and reality within VR applications, and arranges a real environment on the extreme left and a virtual environment on the right hand side (Milgram et al., 1994), refer to Figure 21 in Section 5.1. The area in between is classified as Mixed Reality (MR) and consists of elements of both environments, real and virtual (Milgram et al., 1994; Azuma, 1997). While the discussion in the last chapter solely focussed on entire virtual auditory environments, this section concentrates on mixed reality and its applications. One of the first definitions of augmented reality – already in respect to the development of augmented *audio* – was provided by Cohen et al.:

> *"Augmented reality is used to describe hybrid presentations that overlay computer-generated imagery on top of real scenes. Augmented audio reality extends this notion to include sonic effects, overlaying computer-generated sounds on top of more directly acquired audio signals. (Cohen et al., 1993)*

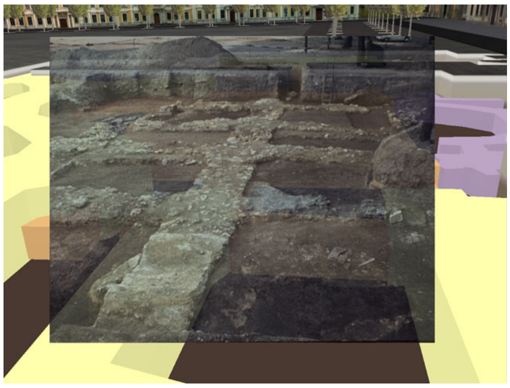
The goal of this section is to identify the position of augmented audio reality within the virtuality continuum, and to connect it to the abstract definitions of an enhanced scene $\mathcal{E}$ that were developed in the last chapter.

6.1.1 *Mixed Reality Systems*

The area of mixed reality ranges from Augmented Reality (AR) towards Augmented Virtuality (AV), and is differentiated by the level of virtuality present (Milgram et al., 1994), see also Figure 21. Augmented reality describes a system that enhances and augments a real world environment with additional artificial information (Caudell and Mizell, 1992; Feiner et al., 1993). Augmented virtuality describes the opposite, and defines a virtual environment that is augmented by real world data. The technological and semantical



(a) Augmented Reality.



(b) Augmented Virtuality.

Figure 38: Image-based Mixed Reality (Freudenberg et al., 2001a; Röber, 2001).

characteristics are in both systems very similar to pure virtual environments with one major difference. As virtual environments not necessarily resemble a real world place, the (physical) attributes that describe this environment can be arbitrary. This is not true for augmented reality, which is bound to the physical laws that are present at the real world location (Milgram et al., 1995). Accordingly, the term augmented reality has been generalized and redefined by Azuma:

> *"Augmented Reality (AR) is a variation of Virtual Environments (VE). VE technologies completely immerse a user inside a synthetic environment. While immersed, the user cannot see the real world around him. In contrast, AR allows the user to see the real world, with virtual objects superimposed upon or composited within the real world. Therefore, AR supplements reality, rather than completely replacing it." (Azuma, 1997)*

The major goal of mixed reality systems is to present the real and virtual elements in a way that they are perceived as one. Azuma defines three primary attributes for the classification of augmented reality systems (Azuma, 1997):

- Combination of virtual and real elements,
- Interaction in realtime,
- Real and virtual objects are arranged in 3D space.

Examples for augmented Reality/Virtuality.

Two examples for an image-based mixed reality can be seen in Figure 38. In the first example, a virtual reconstruction of the *Magdeburger Kaiserpfalz* is composed over an original photograph of the excavation site (Figure 38a), while the example in Figure 38b shows a photograph of the archeological remains blended over a virtual reconstruction of the excavation site (Freudenberg et al., 2001a; Röber, 2001). According to the definition above, Figure 38a describes an augmented reality scenario, while Figure 38b shows an application for augmented virtuality.

The formal model that was developed in Section 5.1 for describing 3D virtual auditory environments can be easily extended to include augmented audio reality environments as well. The main difference is an additional mapping of the structural information E_S onto a real world location. A bijective mapping can be described as a homomorphism that contains a complete one-to-one mapping of the virtual environment onto a real

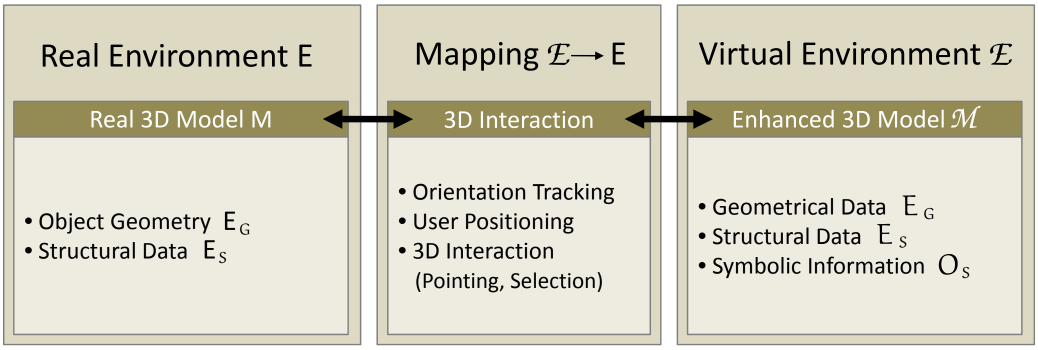


Figure 39: Augmented Audio Reality – Principle.

world place. However, the definition of augmented reality only requires this mapping for at least one object, eg. $\exists \mathcal{M} \in \mathcal{E} : \mathcal{M} \rightarrow \mathrm{M} \in \{\mathtt{real\ world}\}$, see also Figure 39. Furthermore, augmented reality enhances and augments not only existing objects, but also adds additional artificial objects and information to a real scene, hence $\mathcal{E} \not\leftrightarrow \{\mathtt{real\ world}\}$. The space and the setting (ie. physics) of the augmented environment, of course, have to match the ones that are present in the real world, as otherwise, the environment will be classified as virtual reality (Milgram et al., 1995).

The definitions of flow, immersion, presence and tele-presence, as were discussed in Section 5.1, are one-to-one applicable to mixed reality systems as well. One major difficulty is the combination of virtual and real objects in a way that they are perceived as existing in one single environment. One important factor is the display quality of the virtual objects, another is the latency of the interaction with the system. To perceive a virtual object as *real*, it has to be presented alike to the real objects in the scene. While computer graphics achieved a high degree of realism at interactive rates in recent years, the latency of mixed reality systems is still an issue (Bowman et al., 2004; Bimber and Raskar, 2005). Latency is thereby a combination of the time required for various tasks. It includes the time for interaction, the time required for the system to respond, as well as the time necessary for the final display of the updated scene. Besides the rendering and display of the output, the interaction devices employed often consume the majority of time (Bowman et al., 2004). The overall latency must be below the update rate of the system, which is for visual applications in most cases around 25fps (ie. latency $\leqslant$ 40ms).

6.1.2 *Augmented Audio Reality*

One of the commercially most successful implementations of an augmented audio reality (AAR) system are *Audio Guides* such as the Sennheiser *guidePORT* (Sennheiser, 2008). These portable devices are employed in museums throughout the world to guide visitors through an exhibition and to display specific information for certain exhibits. Newer systems also feature an automatic user positioning, in which the system automatically detects the user's location by using encoded magnetic fields that are associated with certain exhibits (Sennheiser, 2008). However, as the interaction and user-localization is point based only, these audio guides do not classify as AR system in the strictest sense of its definition (Azuma, 1997).

The fundamentals of augmented audio reality were developed at the same time and together with *visual* AR technologies (Feiner et al., 1993; Cohen et al., 1993; Cohen, 1994). Interesting to note is that augmented audio reality never received very much attention.

In most cases, it is only developed to complement visual AR systems, without a clear focus of the possibilities of an *audio-only* augmented reality. Although an AAR system requires significantly less resources compared to a visual system, difficulties still apply with an efficient and accurate display of artificial sound sources that are superimposed over a real world acoustics. More recently, Härmä et al. describe an augmented audio environment as:

> *"The concept of augmented reality audio characterizes techniques where a real sound environment is extended with virtual auditory environments and communication scenarios. An augmented audio environment is produced by superimposing a virtual sound environment onto the pseudoacoustic environment."* (Härmä et al., 2003)

The acoustic display of the real environment is here integrated into the system using an in-ear microphone/speaker combination. As this especially alters the spatial perception of sound, Härmä et al. describe it accordingly as *pseudoacoustic environment* (Härmä et al., 2003). A more intuitive presentation can be achieved using so called *nearphones*, speakers that are mounted in close vicinity to the ear but without blocking the pinna and the ear canal (Cohen, 1994; Kanno et al., 2006). An interesting alternative are so called bone-conducting headphones, which transmit sound via skin and bone, refer to Figure 45a (Vonia Corporation, 2008). Both systems allow an unaltered perception of the real worlds acoustics, and at the same time a presentation of additional, artificial sound sources.

Auralization of Figure 40.

A small example visualizing an augmented audio reality scene can be seen in Figure 40. It shows the familiar living room environment, this time displayed in a photo-realistic way, with a virtual avatar in its center. In this example, the setting is part of an augmented audio adventure game scenario, in which the main character has to find the exit through the front door. Similar to the examples discussed previously in Chapter 5, the door is initially locked, and the task is to find the key and a certain document to unlock and open the door. The objects which are depicted in red are common 3D sound sources, while the objects in blue are assigned auditory textures, refer to the discussions in Section 5.3.2. The clock on the wall is used as an orientation beacon for the virtual auditory environment. The door's auditory texture conveys in this example the state of being locked, as long as a missing (virtual) document and the key are not found. To achieve this task, several auditory clues are provided as an aid, such as a virtual character (ghost), or additional auditory beacons.

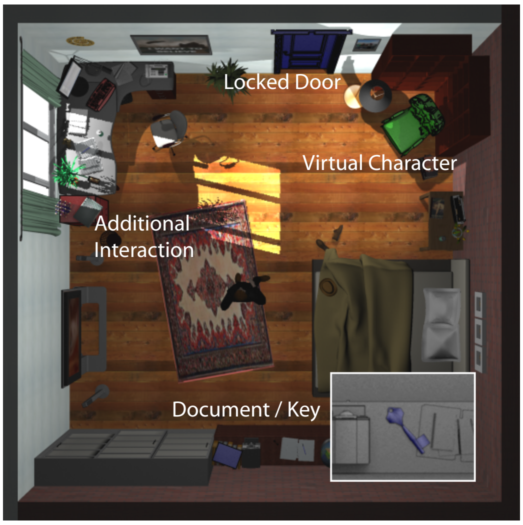


Figure 40: Augmented Audio Example.

An experience of this virtual auditory environment, placed within a real location, can be, depending on its content and implementation, very stimulant and immersive. Although the feeling of presence is directly related to the sophistication of the technologies used, an important factor is the absence of (visual) information, which is now substituted by the user's own imagination. Similar to the reading of books and the listening to radio plays, it can be assumed that a user's immersion and the feeling of presence within an augmented *audio* reality is stronger than in any other VR/AR system. Although this statement is difficult to prove, Chapter 9 discusses and evaluates several applications that are concerned with this statement.

Using these discussions and the descriptions of Azuma and Cohen et al. (Azuma, 1997; Cohen et al., 1993), augmented audio reality is defined for the use in this research as:

Definition *3D Augmented Audio Reality* describes a system and techniques that allow an enrichment of an existing real world environment with additional auditory data and information. The system is thereby centered around an auditory design, which describes a strong focus on the benefits of an auditory display and an audio-only application. The superimposed 3D virtual auditory environment thereby complies with the real world place in terms of (physical) characteristics and topology. Spatial interactions are performed similar as to 3D virtual auditory environments, with an added exploration through a real world user positioning and orientation tracking.

The requirements for a 3D augmented audio reality environment are:

- Based on a (non-realistic) 3D auditory display design.
- Employs efficient techniques for user localization and orientation tracking.
- Uses a realtime 3D spatial interaction design (positioning, pointing).
- Is developed around an auditory-centered design.

Requirements

The requirements for the development of an AAR system are in many aspects very similar to the demands of a 3D virtual auditory environment. The most prominent difference is additional user-tracking and positioning equipment, which is required for a localization/orientation of the listener within the real/virtual environment. According to the above provided definition and the discussions thus far, the requirements for the development of an augmented audio reality system are:

- A 3D virtual auditory environment framework with suitable:
  - Task-related sonification techniques,
  - Task-related spatial interaction techniques, as well as
  - A non-realistic 3D auditory display.
- A proximaural sound presentation system.
- User positioning and orientation techniques (tracking equipment).
- 3D spatial interaction devices.
- Realtime sound spatialization and acoustic simulation techniques.
- All integrated within a lightweight and transportable system (wearable computer).

Due to a less sensitive auditory perception, compared to vision, the requirements for the tracking accuracy are not as high as for visual applications. It can be assumed that, depending on the task and application, a tracking accuracy of +/- 2m for the listener's position is sufficient. This, however, is only true if the setting and the objects for interaction are large enough. In small environments, such as an office, a different positioning approach is required, and can be realized using Bluetooth and other proximity-aware technology. Using an intelligent design for the augmented environment, the required accuracy can further be *decreased* to allow the employment of free available, everyday technology. The accuracy for measuring the user's orientation (head-tracking), however, has to be as high as for 3D virtual auditory environments.

The remainder of this chapter will analyze and discuss these requirements in more detail, and extend the audio framework of the last chapter towards an employment for augmented audio reality applications. Section 6.2 concentrates on the required location-aware interaction techniques and discusses both, tracking hardware as well as the methods required for a 3D spatial interaction design.

Applications and Prospects

Several implementations of augmented audio reality already exist and were beneficially applied in many areas. These range from entertainment & edutainment applications, over the guiding of visually impaired, to the support of daily tasks using ambient intelligence and ubiquitous computing (Cohen, 1994; Rozier et al., 2000).

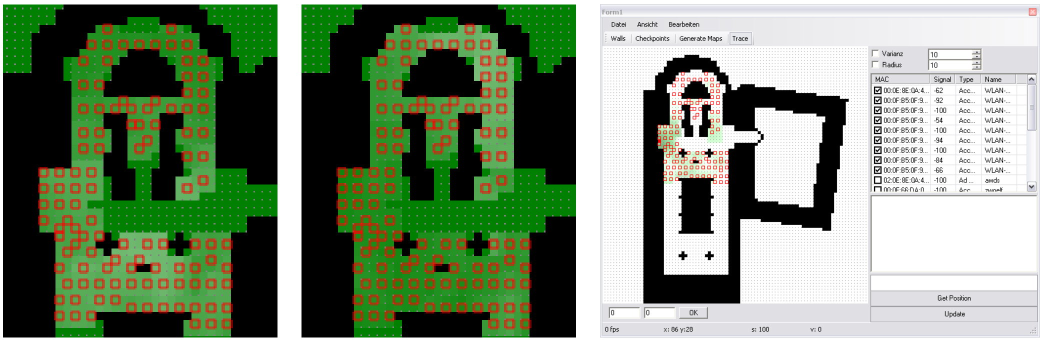
Similar to virtual reality, the possibilities of augmented (audio) reality systems inspire and are explored by artists and scientists together: Augmented audio is most often employed for an exploration of exhibits in a museum (Sennheiser, 2008), for an acoustic examination and exploration of history in the *Berlin Wall* project (Mariette, 2006), as well as for playing computer games (Cohen et al., 2004). Mariette has designed several exhibitions that focus exclusively on the possibilities of augmented audio reality and the perception of a personal, location-sensitive 3D soundscape (Mariette, 2006, 2007a). Due to the less rigid hardware requirements, some implementations of augmented audio on portable devices, such as mobile phones, exist (Ekman, 2007).

The technology of augmented audio reality is in general very well suited to aid guiding and navigation related tasks. Several prototypes – most of them developed for aiding the visually impaired – have been designed to simplify and enhance a navigation and orientation in complex environments, such as cities and large office complexes (Cohen, 1994; Härmä et al., 2003; Walker and Lindsay, 2005). Interesting would be also an application of augmented audio reality to support the performance of daily tasks in the form of audio-centered ambient intelligence and ubiquitous computing. As current developments continue to increase the visual complexity of our environment, augmented audio might be a more suitable technique for a non-intrusive display of data and information.

6.2 LOCATION-AWARE INTERACTION CONCEPTS

A key aspect in augmented reality applications is the possibility to interact with the system within a real-world setting. For this purpose, efficient and accurate user-tracking devices and techniques are required to determine the user's position and orientation. Using these two techniques, the user is able to navigate and orientate oneself within an augmented environment in the search for places of interest for further exploration. At these locations, spatial interaction techniques are required to interact with virtual and/or real objects, to gather more information and knowledge (Bowman et al., 2004; Bimber and Raskar, 2005). Although common interaction techniques using a computer keyboard and mouse would be sufficient in many cases, a spatial interaction design has the advantage of emulating interactions from the real world, and thereby directly improves the perception of both environments as one.

The following two sections are therefore dedicated to the development of efficient and cost-effective user tracking and positioning techniques, as well as to the design and exploration of spatial interaction metaphors. The goal of this section is not only to device



(a) Radiomap AP2. (b) Radiomap AP5. (c) Positioning Application showing the User's Location Probability (green).

Figure 41: WiFi-based User Positioning.

an augmented audio reality system and techniques in accordance to 3D virtual auditory environments, but to design a framework that is efficient, reliable and uses commonly available low-cost components.

6.2.1 *User Positioning and Tracking*

The problem of *registration* is one of the central aspects in augmented reality, in particular for visual AR systems. Dedicated user tracking equipment can thereby also be used for a detection of gestures and movements to permit an intuitive interaction with the virtual environment (Rose et al., 1995; Azuma, 1997). Rolland et al. describe the importance of user tracking and positioning techniques for AR systems as follows:

> *"Tracking for virtual environments is necessary to record the position and the orientation of real objects in physical space and to allow spatial consistency between real and virtual objects." (Rolland et al., 2001)*

Different systems for measuring the user's position and orientation are available, and include inertia-based sensors, time-of-flight systems, mechanical devices, as well as spatial scans and phase-difference sensors (Rolland et al., 2001), see also Figure 42 for some examples. As the focus of this research is the development of a cost-effective system, several of theses technologies already disqualify. The global positioning system (GPS), which is classified as a time-of-flight system, is employed in a few outdoor-based augmented reality scenarios (Cohen et al., 2004). However, GPS systems can not be used indoors and also exhibit a positioning accuracy of only 8–10m (Rolland et al., 2001). Other cost-effective user positioning techniques are based on Bluetooth, or the WiFi radio network (Youssef et al., 2003). An advantage of using Bluetooth is that it can be employed easily and is also very stable and reliable. The disadvantage of this technology is that it only permits, similar to Sennheiser's guidePORT system, a point-based user localization, which does not allow a complete and continues tracking of the user's position (Härmä et al., 2003; Sennheiser, 2008).

A video explaining Figure 41.

A positioning technique that enjoys a continues increase in popularity and research is the so called *WiFi Radio-based Positioning* (Youssef et al., 2003; Youssef and Agrawala, 2005; Ivanov and Schemmer, 2007). This technique uses publicly available WiFi access points to determine a user's location. In a pre-processing step, a radiomap is measured and generated which displays the radio signal strength for each access point and location

(a) InertiaCube3[1] (InterSense, 2004).

(b) OceanServer OS5000-US Compass[2] (OseanServer Technology Inc., 2007).

(c) Gyration UltraGT[3] (Gyration, 2004).

Figure 42: 3D Interaction and User-tracking Devices.

in the map. During the later user tracking, this information is compared to the current signal strength, in which the system generates a probability map to determine the user's current location (Youssef et al., 2003; Youssef and Agrawala, 2005; Otto and Domke, 2007). An example can be seen in Figure 41. Here Figure 41a and Figure 41b display radiomaps for two access points that were used in an augmented audio reality experiment that took place in the Magdeburg Cathedral, refer to Section 9.5 for a more detailed discussion. In this experiment, a total number of 9 access points were used, see also Figure 43. Figure 41c shows a screenshot of the application during the tracking procedure and the determined user's location, which is highlighted by a red square. An advantage for using this technique is the growing number of publicly available WiFi access points, as well as a large variety of WiFi equipped mobile hardware, such as mobile phones, laptops and portable game consoles (PSP).

As the localization is based on signal strength alone, it does not interfere with any security issues and the positioning also works properly with encrypted access points. The accuracy of the positioning varies from 2–6m and is dependent on the environmental geometry, the total number of access points used, as well as on the quality of the pre-measured radiomap (Röber et al., 2006a; Otto and Domke, 2007). This accuracy is sufficient for augmented *audio* reality applications if the setting is distributed over a larger area. The position, however, has to be interpolated and additional measurements have to be employed to avoid ambiguities. As this radiomap is one of the key aspects, an efficient and accurate simulation of radio signal propagation is imperative. Similar to other areas, computer graphics hardware can here be employed to improve the simulation's quality and especially its efficiency (Rick and Mathar, 2007). A close examination of Figure 41a and Figure 41b exhibit a shadowing effect of the radio signal that occurs behind walls and large obstacles.

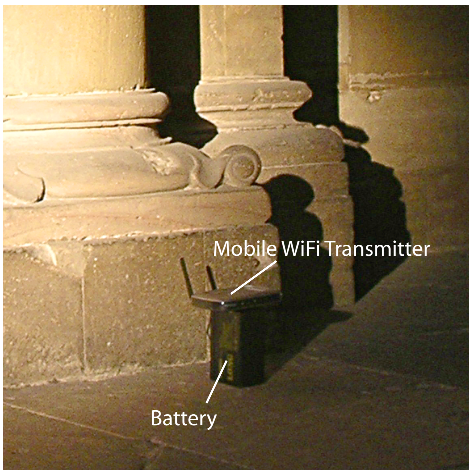


Figure 43: Mobile WiFi Access Point.

1 http://www.isense.com/
2 http://www.ocean-server.com/
3 http://www.gyration.com/

As of this effect, a WiFi-based user tracking is of higher accuracy indoors than outdoors (Röber et al., 2006a; Skyhook Wireless, 2008). For a free field experiment, the access points have to be arranged in a way to create an artificial signal shadowing, ie. that they do not evenly cover the entire area. Newer measurements in Berlin, with a very high number of simultaneously *visible* access points ($\geqslant 25$), suggest a constant positioning accuracy between 1–2m (Otto and Kurth, 2008).

Just as the user positioning, the measurement of the user's head orientation is of high importance as well and has to be performed with a high accuracy and a low latency. Several devices to accomplish this task are depicted in Figure 42 (Rolland et al., 2001). The example in Figure 42a is a rather expensive device, which is based on high accuracy inertia sensors, while the device in Figure 42b shows a digital compass and the example in Figure 42c is a 3D mouse based on a gyroscope (InterSense, 2004; OseanServer Technology Inc., 2007; Gyration, 2004). All devices provide a very high resolution of up to 0.1° and have a latency that ranges between 20–50ms.

For a low-cost implementation of an augmented audio reality system, the digital compass and the gyro-based 3D mouse have the highest applicability. Both devices can be employed in a wearable computer setup and do not require additional hardware or sensor equipment. Also, the WiFi-based user positioning appears to be well suited for augmented audio reality, as this technology can be applied to indoor and outdoor settings. However, care has to be taken in the arrangement of the access points and the measurement of the radiomap.

6.2.2 *Spatial Interaction*

The last section discussed several techniques and hardware devices, which are applicable to perform 3D spatial interactions within virtual/augmented environments. A direct interaction with virtual objects, eg. to acquire information and/or to select/activate objects, requires spatial interaction techniques similar to those discussed in Section 5.4. The focus of this section lies on the development of techniques that efficiently combine both worlds, ie. to define an interaction on virtual objects that are based in a real environment.

An interaction with augmented environments is very similar to the interaction design of entirely virtual spaces. Basic tasks to be performed are selection, manipulation and/or activation of objects, as well as a recall of information. The main difference, compared to an interaction with pure virtual environments, is an interaction in real world space. This requires a precise and congruent mapping of both – virtual and real – environments in order to be perceived as one. The WiFi-based user tracking technique that was discussed in the last section provides here an average positioning accuracy of about 3–4 meters, although experiments with high-quality radiomaps suggest a possible accuracy of around and less than 1 meter (Youssef and Agrawala, 2005; Otto and Kurth, 2008). A tracking accuracy of 4m is sufficient if the augmented audio environment only contains large objects that are adequately separated

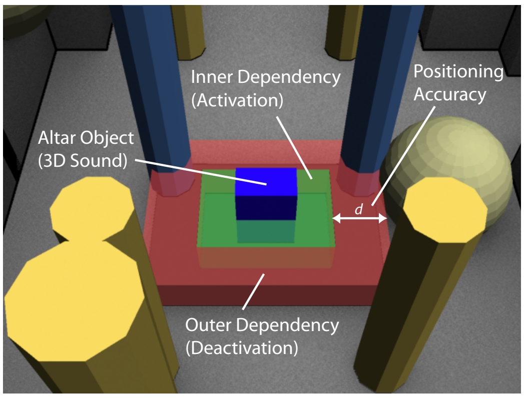


Figure 44: Design for Position Dependency.

from each other, refer to Figure 44 and Section 9.5 for a discussion of examples. The position dependency, as it was introduced in Section 5.5, describes a user controlled interaction with the environment based on proximity. To compensate a less accurate positioning performance additional measurements are required, as is shown in Figure 44. These consist of two bounding boxes of different size, which act as switch and either activate or deactivate a certain object. The inner bounding box is thereby used to activate a position dependency, while the outer bounding box is used to switch it off. The differences in size between the bounding boxes d must be large enough to compensate for the positioning error, ie. for the WiFi based positioning around 2–3 meters. This approach, however, is only applicable to larger settings with sufficiently separated objects. An alternative implementation that circumvents these limitations can be realized through additional proximity sensors, such as Bluetooth and RFID technology. Here an object is activated as long as the user's distance is below a certain pre-defined threshold. The implementation for the other environmental dependencies, eg. time- and object-dependencies, can be directly taken from 3D virtual auditory environments without any modification.

All spatial interactions that were devised in Section 5.4 are directly applicable also for an interaction with augmented auditory environments. The employed 3D interaction devices, however, have to be both, portable and wireless. Although a regular wireless gamepad can be used to model the majority of interactions, an application of a gyro-based 3D mouse is here much more interesting, refer to Figure 42c. Such a gyro mouse is employed in the implementation of this framework as a 3D interaction device. With two buttons and an additional scrollwheel, this device can be well employed in a large variety of 3D spatial interaction scenarios, such as for 3D object selection, pointing and picking, but also to interact with 3D auditory menu systems.

After these discussions, the following section reviews the previously devised audio framework and extends it towards an application for augmented audio reality scenarios. The here discussed techniques for user tracking and positioning, as well as for 3D spatial interaction are thereby integrated into the system.

6.3 SYSTEM DESIGN

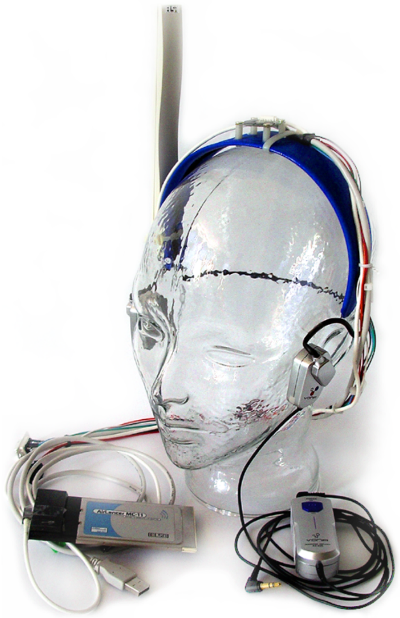
The audio framework of Section 5.5 can easily be extended towards an application for augmented audio reality. The augmented audio reality system designed is meant to be used in a variety of applications, ranging from pure enter- and edutainment, to the development of applications to aid the visually impaired. Several experiments and case studies have been performed and are presented and discussed in more detail in Chapter 9.

The development of the system is divided into hardware-related and framework design issues. The first section describes the selection of specific hardware for 3D user interaction, orientation tracking and positioning, while the second part discusses the modifications of the framework regarding an implementation of the 3D spatial interaction techniques.

6.3.1 *Hardware Requirements*

A major goal for the design was the development of an affordable system based on efficient, but low-cost components. The hardware requirements therefore are:

- A portable and lightweight computer platform (Laptop or PocketPC), with
  - Realtime sound spatialization and acoustic simulation hardware.



(a) User Positioning and Tracking, Bone-phones.

(b) 3D Pointing and Selection.

(c) Play and Interaction.

Figure 45: Augmented Audio Reality – System Hardware.

- A proximaural headphone system,
- User-orientation tracking hardware (head-tracking),
- A user-localization/positioning system, as well as
- A wireless 3D spatial, and a standard (gamepad) interaction device.

An overview of the selected hardware can be seen in Figure 45. It shows in Figure 45a a digital compass for the head-tracking, a special antenna for the WiFi-based user-positioning, as well as bone-conducting headphones as a proximaural display. Figure 45b displays the 3D interaction device, a common gyro-based 3D mouse, while Figure 45c shows a regular wireless gamepad for the input of standard interactions. Additionally, several mobile WiFi access points were employed for the experiments and for a test of the system, refer to Section 9.5.2. These access points were dispersed manually within the environment to ensure a good coverage, but to also introduce signal shadowing artifacts of pillars and walls, refer also to Figure 41.

The selection of this hardware was made for several reasons. The components exhibit all of the required qualities and are additionally highly affordable. The bone-conducting headphones are available for € 100 and can be well employed as an acoustic display in auditory environments (Vonia Corporation, 2008). The results of a short analysis and comparison with a regular HiFi headphone system are available in Section 9.5.1. The digital compass employed in this system is available for € 200, possess several serial interfaces and achieves an update rate of 40Hz ($\approx$ 25ms) with an accuracy of 0.1° (OseanServer Technology Inc., 2007). The WiFi-based user tracking is performed using an external, but regular WiFi computer card that can be equipped with an additional antenna and is available for € 60. The 3D pointing and interaction device that was employed is a gyro-based computer mouse, which is available for around € 50.

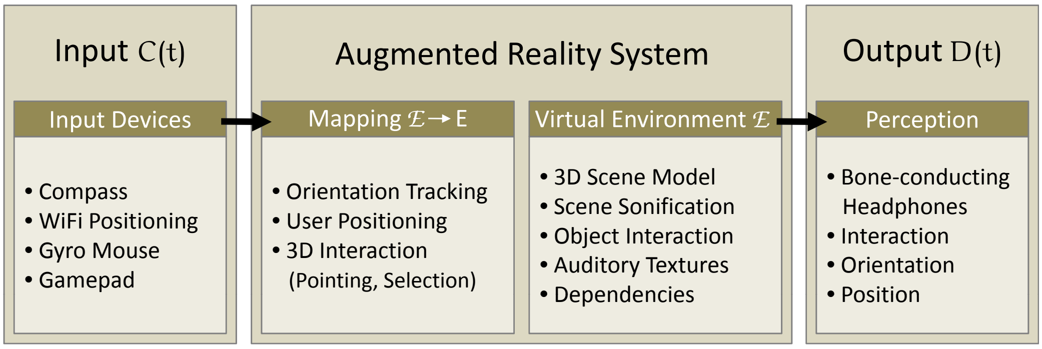


Figure 46: Augmented Audio Reality Framework.

6.3.2 *Framework Extension*

The system itself is an extension and implemented on top of the audio framework that was devised earlier in Section 5.5. It is implemented using C ++ and oriented along the requirements of mobile hardware, ie. laptops and PocketPCs. Although the framework's performance on less efficient hardware, such as a PocketPC running WindowsMobile, has not yet been evaluated, it should perform with an adequate efficiency. As in the previous implementation, the system employs OpenSG as 3D scenegraph system, OpenAL for sound rendering 3D spatialization, as well as OpenGL for an additional 3D visualization of the scene. A freely programmable DSP-based sound API for sound rendering and sound spatialization has already been realized and is available for an application on mobile systems (Stockmann, 2007). Interesting to note is that this API is more efficient than any other sound API available and also allows an exchange of the HRTF filters employed for a personalization of the 3D sound spatialization.

The majority of aspects that were discussed previously in Section 5.5 are directly applicable to an implementation of augmented audio reality as well. An overview of the augmented audio reality system developed can be seen in Figure 46. Based on the original audio framework, an additional mapping of the 3D virtual auditory environment $\mathcal{E}$ onto a real location is performed. This mapping is performed through several input devices and techniques that allow an efficient user tracking and positioning. Additions to the audio framework therefore include a WiFi-based positioning system, as it is depicted in Figure 41, as well as a support for the new head-tracking and 3D interaction devices, refer to Figure 45. Figure 41 shows a screenshot of the developed WiFi-based user positioning system, which allows to determine the user's position within a real environment with an accuracy of about +/- 2m. This system is implemented using C ++ as well, and integrated into the main system as external library (DLL). The digital compass, which is employed as 3D head-tracking device, is connected to the system using a regular serial interface. With an update rate of 40Hz, the compass communicates with the system and transmits three angels: heading, pitch and roll, which are used to determine the user's head orientation. The main interaction with this system is performed using positioning, ie. position dependencies, as well as through a control via a regular gamepad and a 3D pointing and selection device. For the 3D pointing and object selection, a Gyro mouse is employed, which is controlled using a standard mouse device that converts the 2D mouse coordinates into angels for a selection and picking of virtual 3D objects. Feedback is perceived via bone-conducting headphones (auditory output), refer to Figure 45, as well as through the user's position and orientation within the real environment. Section 9.5.1 examines the bone-conducting headphones in more detail and compares the perception of 3D virtual sound sources and virtual room acoustics using bonephones and regular HiFi headphones.

6.4 summary

After the discussions of 3D virtual and augmented auditory environments, the areas of application and the design of actual examples and prototypes moves into the focus of this research. Figure 47 in the following Chapter 7 provides an overview of the entire system and shows the integration of an authoring component within the current audio framework. Additionally, this chapter discusses issues regarding the authoring and design of general applications that are based on 3D virtual/augmented auditory environments. The chapter develops rules and guidelines for various authoring tasks and devises a 3D authoring environment that is used to implement several examples, refer also to Section 9.5.

I paint things as I think of them, not as I see them. — Pablo Picasso

7 AUTHORING AND DESIGN

In addition to the representation and display of 3D virtual and augmented auditory environments, the design, as well as the process of authoring these environments is of high importance. This not only includes the design, selection and assignment of sounds and sound sources to certain positions, but also the definition of interactions, dependencies, auditory textures, as well as the setup of an underlying *context* – the symbolic information E_S. The goal of this chapter is therefore to identify and describe the authoring process for 3D virtual auditory environments, as well as to derive general authoring paradigms and guidelines that are suitable for the design of a large variety of applications and tasks.

The first section of this chapter reviews common design principles for 3D auditory environments. Additionally, the section examines existing approaches for the authoring and design of sound and acoustics as they are employed in entertainment computing, such as for the design of audio/visual computer games. The goal of this discussion is the development of an authoring pipeline that supports the design of 3D virtual/augmented auditory environments. The discussion continues in the following section by providing authoring guidelines and design aspects. The process of authoring is thereby divided into smaller tasks, which are analyzed individually and for which several principles and guidelines are presented. The chapter concludes with the design of a prototypic authoring environment, which is applied to the authoring of an example scene to show some of the developed techniques in practice.

7.1 AUDITORY AUTHORING

Authoring describes the process of creation and the adding of content for a specific medium. Authoring environments are able to aid this process and are employed in many areas to *create* the actual application. A related area to the authoring of 3D virtual/augmented auditory environments exists in the form of designing audio/visual computer games, and in particular, the design of interactive 3D adventures. Despite many rumors, the adventure genre is still very active and kept alive by a large community, who also creates and develops own authoring systems for the design of 2D and 3D adventure games. Outstanding examples are the *3D Adventure Studio*, the *Adventure Game Studio* and the *Adventure Maker* (van der Honing, 2003; Jones, 2008; Giovanni, 2008).

A commonality in all these systems is their room-centered design approach, which allows the authoring and an easy connection between different rooms and environments. The design of a 3D room/environment includes a selection of background graphics, masks and sprites for walkable areas and walk-behinds, hotspots and objects for interaction, as well as a selection of music and sound effects. Through the design of several of such rooms and an integration with story-related users tasks an adventure game is created. Secondary tasks of the authoring include the design of a user interface, the development of an asset management system, as well as the authoring of content, narration and user interaction techniques. The user interface thereby combines all other applications and later manages and controls the player's progress and play. Interesting to note is that the majority of authoring environments provides an integrated runtime component,

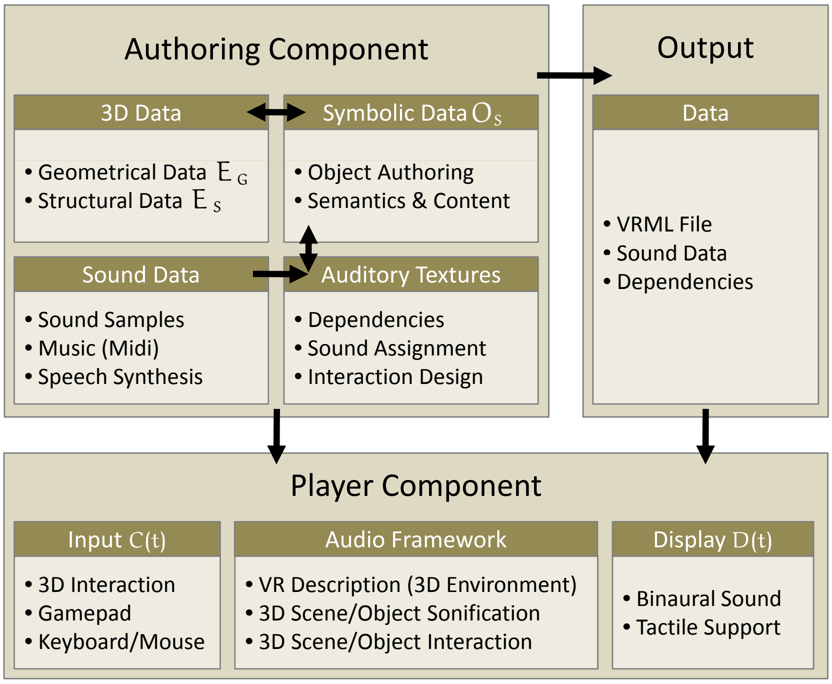


Figure 47: Audio Framework – Overview.

which allows a comfortable preview and evaluation of the game directly from within the authoring application.

The authoring process for 3D virtual auditory environments is not much different than the design of a 3D audio/visual adventure game. However, certain aspects, such as the auditory presentation and perception are quite different and have to be considered during the entire authoring process. With a consideration of an authoring for 3D augmented audio applications as well, also the hardware and sensors employed move into focus. One requirement of the authoring system is therefor to define an interface between the runtime system and the 3D interaction hardware used, see also Figure 47. This is especially important for the design of augmented audio reality applications, which rely heavily on external hardware for user interaction and positioning. As the hardware of AR systems is not yet standardized, a related authoring environment is often only applicable to this specific system (Paelke and Reimann, 2006). The required components for the authoring of 3D virtual/augmented reality applications are described by Paelke and Reimann as:

- Surveying and capturing of an environment for augmentation,
- Modeling of content for augmentation,
- Definition of dynamics (eg. for user interaction), as well as
- Spatial registration of both (real and virtual) environments.

This list, although its authored applications focus on a *visual* perception, can also be employed for the design of an authoring system specific to the creation of a 3D *auditory* environment. Such an audio-centered authoring and design is a non-trivial task and includes much more than just the assignment of sound sources to certain positions in 3D space. The process of authoring itself is founded and based on several design guidelines that were derived during this research, and whose discussion is the focus of the succeeding section.

An overview of the entire audio framework, and the authoring process in general, can be seen in Figure 47. It shows that the framework is basically divided into two parts, an authoring component and a runtime system. The authoring starts with the creation of a 3D model that defines the augmented environment and resembles a real-world place. The 3D model is described by geometric data E_G in the form of polygons and triangles, as well as through structural information E_S that defines the topology and the mapping of the 3D model onto the real-world environment. The second step comprises the design and creation of sounds and sound effects, as well as speech and music samples, which are later added to the virtual environment and assigned to certain objects. Thus far, the authoring was performed using external tools and applications, such as *3DStudioMax* for 3D modeling and *Soundforge* for sound design (Autodesk, 2008; Sony Creative Software, 2008), see also Figure 50. After this *pre-authoring*, the design of the actual 3D virtual auditory environments is performed through the definition and assignment of object-related symbolic information O_S to each object within the scene. The result of this authoring is the transformation of a regular 3D scene E into an acoustically enriched 3D enhanced environment $\mathcal{E}$.

The transformation (authoring) is mainly based on the design of time-, object- and position-dependencies, the creation and assignment of 3D sound sources, as well as includes a selection of task-related sonification and interaction techniques and the definition of auditory textures for the virtual scene objects, refer to Figure 47. After the authoring is complete, the acoustically enhanced environment $\mathcal{E}$ is saved along with all data and scene information within an extended VRML data file. This data can later be loaded into the runtime system, and be used for playback and to explore and experience the authored 3D virtual auditory environment.

Within the context of this research, the authoring and design of 3D auditory environments is defined as follows:

Definition The *Authoring and Design* of a 3D virtual/augmented auditory environment is described as the transformation of a regular 3D scene E into an acoustically enriched 3D enhanced environment $\mathcal{E}$. The authoring is divided into a pre-authoring step, which includes the creation of scene geometry E_G and an authoring of structural information E_S. The actual 3D auditory scene design performs a definition and assignment of symbolic data O_S to the entire scene and to individual objects to design enhanced 3D models $\mathcal{M}$. The authoring pipeline comprises:

- The creation of geometry and a virtual (geometric) 3D scene authoring,
- The design of sound, speech and music samples,
- The definition of dependencies and a selection of sonification and interaction techniques, as well as
- A scene object authoring through the definition and setup of auditory textures.

Besides this technical creation of 3D virtual auditory environments, also the arrangement and authoring of sounds and objects within this environment is crucial and requires a careful design. A well authored application is here always able to intuitively convey enough information to the listener. A badly designed environment might contain empty and silent areas in one place, while it overburdens the listener with too much information and too many sources in other locations.

7.2 AUTHORING GUIDELINES

Several parts of the authoring process are task-related and can vary depending on the content and intent of the final application. An entertainment application, such as an audiogame, has different requirements than an augmented audio application that aims to aid the navigation and orientation of the visually impaired. The similarity of both examples is the underlying authoring pipeline, as well as the design guidelines and authoring techniques used. As discussed in the previous section, the basic authoring is similar for all applications and requires geometrical data, structural information, sound, speech and music samples, as well as content and symbolic information. The differences reside in the authored environments themselves, more specifically in the 3D models, the intent, the content, the selection of sonification and interaction techniques, as well as in the later display and use of the final application.

Chapter 4 described several examples for 2D and 3D auditory displays along details of their implementation and design principles. The majority of these principles are also applicable and can be used for the design and authoring of 3D virtual auditory environments as well. Despite a task-related authoring, some issues are common in the design of all applications. This includes primarily a careful selection and design of sounds and samples to support an expressive and efficient sonification of the environment. But it also includes the principle that not all objects can be audible at the same time, as this would clutter, and render the auditory display inoperable. A procedure has to be employed that controls the playback of sound sources and sonifications consistent to the listener's interaction. Auditory textures can here be used together with distance attenuation techniques to blend out farther away sources, as well as to activate sound sources using proximity and a positional dependency.

To exemplify the process of authoring and to discuss some of the authoring techniques, the familiar 3D environment is employed in the setting of an augmented audio reality game. Figure 48 provides here an overview, as well as shows the required authoring steps to create a small audio-based AR game. In accordance to the previously discussed examples, the environment resembles a part of an augmented adventure game, in which the player's first task is to find an exit to this room. The door, however, is locked, and the player needs to find a key to unlock and open the door. Therefore, both objects, *Door* and *Key*, are assigned auditory textures that describe these objects, as well as implement possible interactions. The door object represents two states, locked or open, while the key emits a descriptive sound that identifies itself as key. Both objects have a position dependency, in which the key's position dependency also activates an object dependency that unlocks the door, refer also to Figure 37. The key object additionally features a time dependency, which is used as an aid for the player to find the key. A further authoring for this scene is required to setup the environment and to provide additional auditory feedbacks, eg. environmental sounds, a narrators voice and music. The clock on the wall

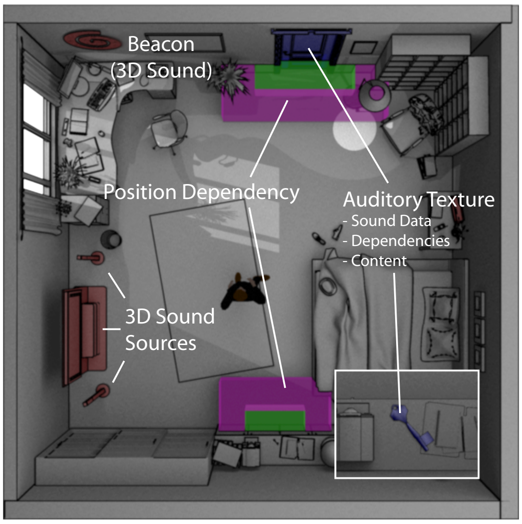


Figure 48: Authoring Example.

can be employed as a an auditory landmark and is used as a *north beacon* to provide additional cues for the user's orientation. These tasks require the specification, setup and positioning of 3D sound sources, as well as the use of dummy objects and additional auditory textures to implement a more detailed narration.

The authoring of the example depicted in Figure 48 is relatively easy and straightforward. The following three sections discuss specific problems within the authoring pipeline, as well as devise guidelines and principles to improve the design process.

7.2.1 *3D Scene Authoring*

3D scene authoring is concerned with the actual physical design and 3D modeling of the virtual environment. This includes the design and creation of individual 3D models and their arrangement within a larger topology. The designed 3D models do not need to be very realistic, as they are mainly used as placeholders for an assignment of auditory textures, as well as for collision detection. Therefore, in most cases simple boxes, spheres and cylinders are sufficient. This allows a fast design and modeling of scene geometry and also to perform changes and adjustments quickly and without many difficulties. Furthermore, low resolution models require less storage space and can also be rendered more efficiently using OpenGL for a possible *visual* scene display. Care must be taken in the positioning of virtual objects for an augmented audio reality application. Depending on the positioning technique employed, objects with a position dependency have to be separated by at least twice the positioning accuracy, refer to Section 6.2. 3D scene authoring also comprises the design and compilation of speech, sound and music samples for their later use and assignment to individual scene objects. Although high quality sounds are required in all cases, some applications may have secondary requirements, such as storage space in mobile applications, or to account for perceptional differences due to the use of a bone-conducting headphone system. For the example depicted in Figure 48, the 3D scene authoring includes an approximate modeling of the room's interior, as well as the compilation of a sound pool for later sound assignments. 3D models can be created using tools such as 3DStudioMAX, from which they can be exported and stored as VRML data file, refer to Figure 50. Objects in this VRML file are used in the next authoring step, in which they are extended by the integration of sound nodes and auditory textures.

7.2.2 *Content and Dependency Authoring*

The second step in the authoring pipeline is the creation of content; the authoring of the very essence of an auditory environment. A large variety of applications can be implemented just through the modeling of various position, time and object-related dependencies. Their application is very versatile and can also be employed in the form of a *state-machine*, in which additional *dummy objects* are used to store boolean variables for the various states. In this setting, and by also utilizing time-dependencies, a state-machine can be employed as story-engine for the construction of a story arc. The boolean variables thereby contain the various story elements, which are activated either using a time-, or user-triggered object dependency. In the above example, a time dependency is added to the key's auditory texture, which emits a hint after a certain amount of time has passed. An object-based dependency is added to the key object as well, which unlocks the door after the key has been activated. Both, the key as well as the door object contain a position dependency as it is shown in Figure 48. This position dependency activates both objects and acoustically describes their current state. If the key object has been previously

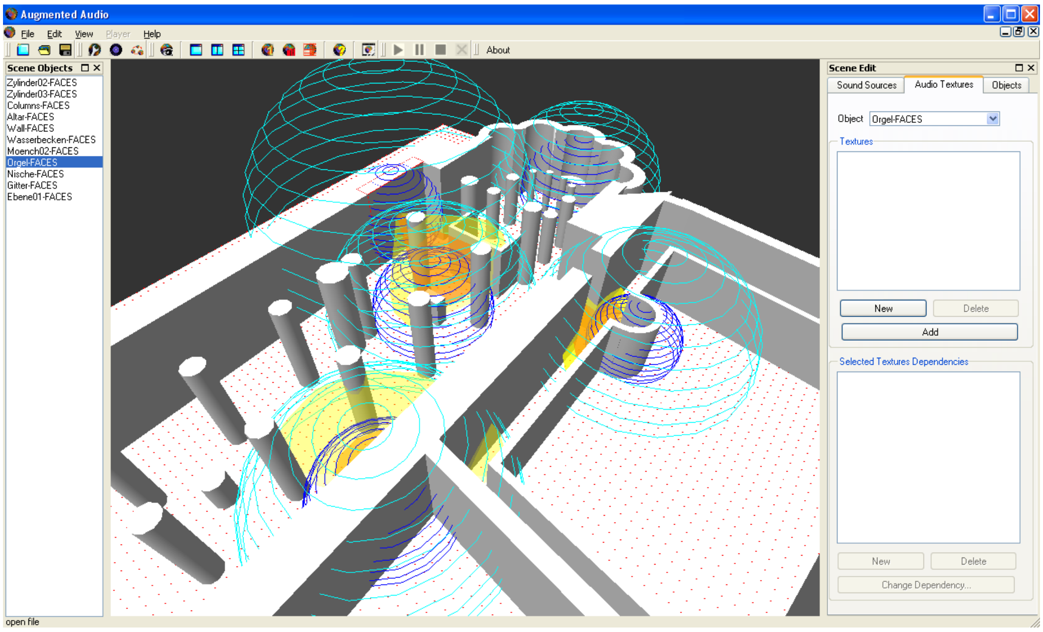


Figure 49: Auditory Authoring Environment.

activated, the door opens, and the player can proceed to the next level. Additional content and story data, which is not shown in this example, can be added to provide a denser storyline and a more realistic atmosphere. This can also be authored simply through the creation of additional 3D sound sources and the assignment of auditory textures.

7.2.3 *Sonification Design*

After the story authoring is complete, the existing auditory textures are extended to implement user interactions and additional object sonifications. The wall clock in Figure 48 is employed as a scene-embedded north beacon, and assigned a sufficiently loud ticking sound. The ticking sound is spatialized with an omnidirectional emittance pattern (point sound source) to be used as an orientational aid. The remaining auditory representations for the key and door object can now be added as well, and might be additionally enhanced by a narrative speech-based description. Finally, all remaining objects are authored and auditory textures are used to define the interaction with these objects and the sonifications used. Music and the narrator's voice are assigned to dummy objects, for which object and time dependencies are used to control their playback.

7.3 DESIGNING AN AUTHORING ENVIRONMENT

An authoring environment that supports the previous discussions and authoring concepts can be developed in accordance with the existing audio framework that was devised in Section 5.5 and extended in Section 6.3. An overview of the entire framework was provided at the beginning of this chapter in Figure 47, which highlights its two major components: an authoring and a runtime system. The runtime system is thereby fully integrated into the authoring component and allows a direct preview of the authored scenes. The authored applications are saved as an extended VRML data file, and are later loaded together with the sound and music samples into the runtime system for playback.

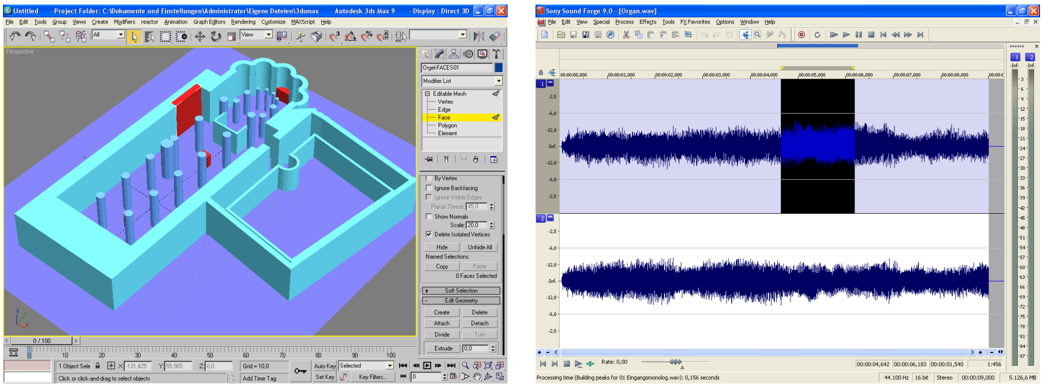
The runtime system can thereby be specific and focus on a certain type of application only, such as 3D audiogames or an augmented audio reality system, but the authoring itself shall be performed in a single, task-overlapping authoring environment, see also Figure 49.

The actual authoring and design starts with an existing 3D model of the environment and a large sound pool of representative sound, speech and music samples. The requirements for developing an authoring system that supports all of the previous discussions are:

- General Application
  - Load/save VRML data files
  - 3D scene visualization (OpenGL)
  - 3D Sound and object visualization
  - Dummy object creation and positioning
  - General environmental acoustics authoring
  - Listener authoring (start position, ...)
  - Menu and mouse based interaction
  - Scene authoring preview
- 3D Sound Source Authoring
  - Select, create and delete sources
  - Sound assignment
  - 3D Alignment and positioning
  - Parameter adjustment (gain, filtering, spatialization, ...)
  - Distance attenuation
- Auditory Texture Authoring
  - Select, create and delete auditory textures
  - Object assignment
  - Time-, object- and position-dependency authoring
  - Sound assignment and individual parameter adjustment
  - Interaction selection (input dependency)

The list differentiates between general authoring requirements, and specific needs for the creation and design of 3D sound sources and auditory textures. A graphical display of the 3D environment is required, as well as a visualization of the individual authoring tasks. This allows a much easier scene design and parameter adjustment. Examples are shown in Figure 49, but are also discussed in more detail in the following sections.

This authoring approach is only suitable for the design of 3D auditory environments, and in this form not applicable for the design of 2D auditory displays and non-topological environments. Section 9.6 later introduces such a *non-spatial environment* in the form of Interactive Audiobooks, and thereby also describes a system that focusses on the authoring of these applications.



(a) 3D Scene Modeling using 3DStudioMAX (Autodesk, 2008).

(b) Sound and Music Design using Soundforge (Sony Creative Software, 2008).

Figure 50: The Pre-Authoring Process.

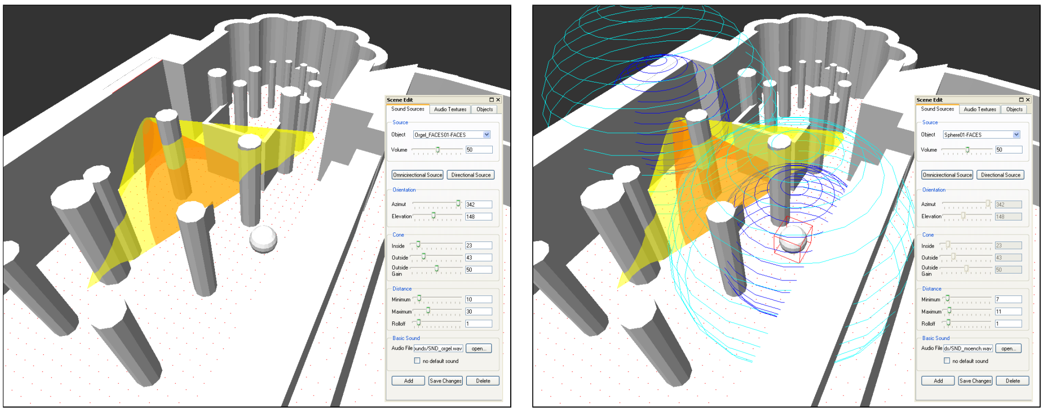
7.3.1 *Implementation*

The implementation of this authoring environment is performed using C ++ , and is based on the framework design as it was introduced in Section 5.5 and Section 6.3. The authoring environment therefore also utilizes OpenSG as scenegraph system and uses OpenAL/EFX for sound rendering and synthesis. OpenSG is used to load the 3D model as VRML data file and displays the 3D scene using an OpenGL visualization, compare with Figure 49. Individual 3D objects can be selected to assign and create 3D sound sources and/or auditory textures. The object data nodes are thereby extended by this information and saved to the VRML file after the authoring process is finished. The user interface for the authoring environment resembles a classic Windows GUI and was developed using the Qt library. The interface and 3D scene visualization is customizable through a selection of specific view screens and floating toolbars. These toolbars allow the creation, positioning and assignment of 3D sound sources and auditory textures to each virtual object or additional dummy objects, see also the following section.

7.3.2 *Authoring Process*

As a conclusion for this chapter, an example is provided that shows the individual steps of the authoring process using the pipeline devised and the authoring environment developed. The example, as can be seen from Figure 49, is part of the authoring for an augmented audio reality application, which will be examined and evaluated in more detail in Section 9.5.2. The focus of this section is to discuss its design and construction and to illustrate the various aspects of the authoring process.

The design starts with the pre-authoring and the modeling of the 3D environment, see Figure 50a, as well as with a selection and mastering of suitable sound, speech and music samples, refer to Figure 50b. The result of this pre-authoring is a VRML data file that contains the geometry and topology of the 3D scene, as well as a pool of sound data for the sonification and interaction authoring. The VRML data is loaded into the authoring environment, compare with Figure 49, in which now the assignment of sound sources to certain scene objects, as well as the design of auditory textures starts.



(a) Creation of a Sound Source and Specification of the Source Direction.

(b) Specification of Distance Attenuation using two Spheres.

Figure 51: Sound Source Creation and Parameter Adjustments.

Sound Sources

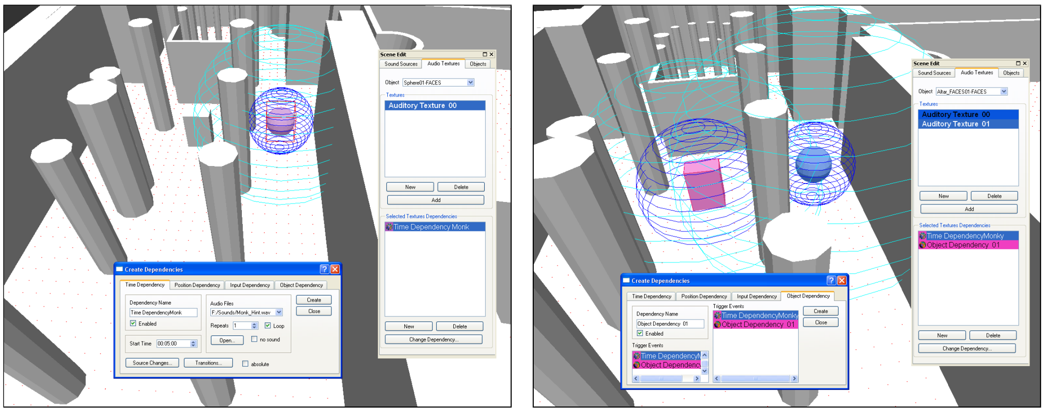
The first part of the authoring is the creation of 3D sound sources and their assignment to individual scene objects. Figure 51 shows the creation of sound sources, as well as the setup of several source specific parameters, such as direction, distance attenuation, loudness and rolloff factor. As OpenAL is employed as underlying sound engine and used for the sound rendering and spatialization, the parameters and settings used are directly mapped to the OpenAL API (Hiebert, 2006). Additional source parameters are visualized using two semi-transparent cones, refer to Figure 51a, in which the transparency defines the loudness/gain and the orientation and size of the cone the direction and radiation pattern for this source. The application allows two different gains to be defined, one for the inner cone (red), and a second for the outer cone (yellow). A distance attenuation model describes the attenuation in between using the rolloff factor specified, compare with Figure 51b.

Authoring of Sound Sources.

Auditory Textures

The definition and assignment of auditory textures is one of the most fundamental, but also one of the most flexible authoring steps, and is depicted in Figure 52. Figure 52a shows the definition of a time-dependency, while Figure 52b shows the dialog for authoring an object dependency. The sphere on the right hand side in Figure 52 is a so called dummy object, which serves as a placeholder and represents in this example the position of a virtual monk. This monk is represented acoustically through a snoring sound, because he is napping and the task of the player is to identify and activate this sound source later. The activation is triggered through proximity, in which an additional position dependency changes the state of the source (ie. wakes up the monk) and initiates a short dialog, refer also to Section 9.5.2. In this example, the time dependency is set to 5 minutes, and if activated, an auditory hint is played to guide the player in the direction of the monk, see also Figure 52a. This time dependency also triggers an activation of the altar object through a defined object dependency, refer to Figure 52b. Time and object dependencies can be employed in many different tasks and are a simple and intuitive way to integrate content into the 3D auditory environment and to narrate a storyline.

Authoring of Auditory Textures.



(a) Authoring of a Time Dependency for the Sphere Object (Monk).

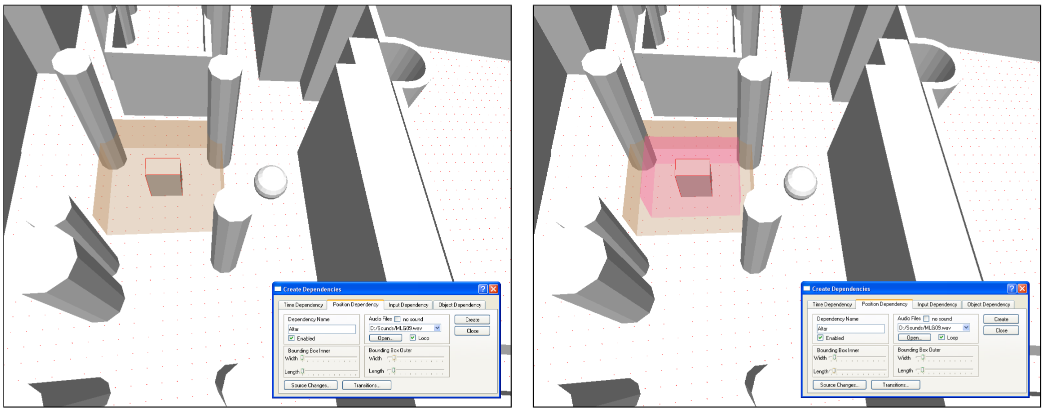
(b) Authoring of an Object Dependency between the Sphere (Monk) and Altar Objects.

Figure 52: Authoring of Auditory Textures.

Position Dependency

Authoring of a Position Dependency.

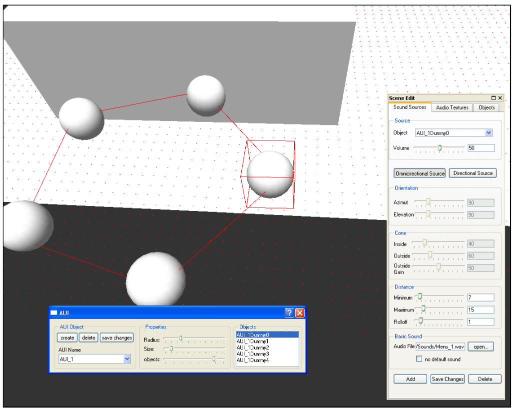
Another part of the auditory texture authoring is the creation of position dependencies, which are shown in Figure 53. Position dependencies are used as a proximity sensor to activate an object, and/or to trigger another object dependency. The authoring of position dependencies for the augmented audio example requires, due to the inaccuracies of the WiFi-based user positioning, the specification of an outer and an inner rim, refer to Figure 53. The inner rim is thereby used to activate an object, while the outer rim is used to deactivate it, see also the discussion in Section 6.2. The differences in length of both *boundary boxes* must be wide enough to accommodate the maximum error of the positioning system. Otherwise, a constant activation/deactivation of an object occurs, which might result in a choppy and very disturbing playback of sound. The same approach can also be used for the authoring of 3D virtual auditory environments, but as the avatar *positioning* is here controlled using a gamepad, the distance between the two bounding boxes can be minimized.



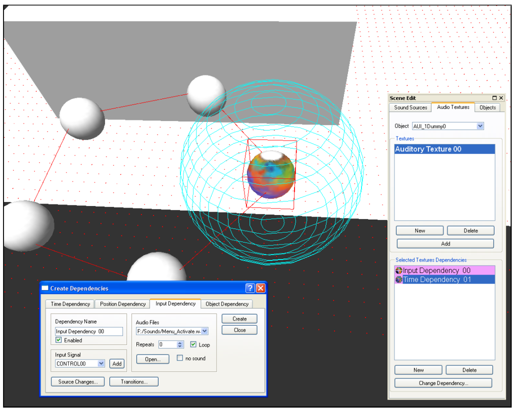
(a) Definition of the outer Rim (Source Deactivation).

(b) Definition of the inner Rim (Source Activation).

Figure 53: Authoring of Position Dependencies.



(a) Creation of a fivefold Ring Menu System.



(b) Definition of Auditory Textures with Interaction Dependencies and Assignment to the Menu Objects.

Figure 54: Authoring of a Ring Menu System.

Ring Menu System

The last example in Figure 54 shows the authoring of a ring-based menu system. Figure 54a shows here the creation of a five-fold circular menu system, for which one node is exemplary authored using an input dependency, see also Figure 54b. Each of these nodes is represented by a dummy object to which an auditory texture is assigned. This auditory texture is similar to the auditory textures discussed before, but can additionally handle an input dependency, eg. for user interaction. Upon interaction and an activation of a node, a certain action is executed. This action can be implemented using additional dependencies and also activate/change other scene objects using secondary dependencies. This section presented several of the previously discussed authoring techniques using an example from augmented audio reality. The authoring for other applications and tasks is very similar, and the main techniques and guidelines apply there as well.

Authoring of a Ring Menu System.

7.4 SUMMARY

This chapter discussed the requirements for an authoring and design of 3D virtual/augmented auditory environments. Several principles and design guidelines for the authoring of auditory environments have been discussed and were presented in the form of an authoring pipeline. Additionally, a 3D authoring environment has been devised and was implemented on top of the audio framework discussed in Chapter 5 and Chapter 6. This authoring environment allows the design of acoustically enriched 3D enhanced environments $\mathcal{E}$, and can be applied to a large variety of areas of application. Exemplarily, the design of an augmented audio reality scenario was presented and discussed.

This chapter basically concludes the discussion of 3D virtual/augmented auditory environments along their authoring and design. Later Chapter 9 returns to a discussion of 3D auditory environments with a presentation and evaluation of several applications and prototypic implementations. As an efficient auditory representation of these environments requires both, a realistic 3D sound spatialization and simulation of room acoustics, the following Chapter 8 is dedicated to a discussion of acoustic rendering techniques. Thereby it explores the possibilities of utilizing computer graphics rendering techniques and hardware for an effective and more realistic sound rendering and simulation.

8 ACOUSTIC RENDERING

Parallel straight lines do not meet one another in either direction. — Euclid

AURALIZATION describes the process of employing physical and mathematical models based on Euclidean geometry to render a virtual auditory scene audible. Thereby a binaural sound output that simulates the acoustic experience for a certain location and for a specific listener in this virtual environment is created. Auralization is an important factor for the research in this thesis, as sound and acoustics are both used as the main carrier to display abstract information. The majority of current applications that employ 3D sounds and room acoustics simulations use freely and commercially available APIs, such as OpenAL/EFX, FMOD or AM3D (Firelight Technologies Pty, Ltd, 2001-2008; AM3D A/S, 2008; Hiebert, 2006; Peacock et al., 2006). Although these APIs are easy to deploy and achieve quite good results for audio/visual presentations, they often fail in audio-only applications as too many approximations are applied regarding the human auditory perception and propagation of sound waves. This chapter takes a closer look into acoustics and 3D sound rendering with a focus of sonifying 3D virtual/augmented auditory environments. Here not only several techniques are discussed, but also advanced and transferred to a graphics-inspired sound rendering technique to achieve a more realistic and efficient simulation.

The chapter is divided into five sections, of which the first one takes a closer look on the requirements of sound rendering for virtual and augmented auditory environments and discusses the most important concepts and techniques. The following section continues the discussion in the direction of a graphics-based sound rendering and debates the utilization of graphics hardware for general (sound) signal processing. Based on these findings and developments, the following two sections present and discuss ideas for a graphics-inspired sound simulation through a ray- and a wave-based acoustic simulation. As each of these techniques has its respective advantages and drawbacks, the last section discusses concepts towards a possible unification.

8.1 AURALIZATION AND SOUND RENDERING

This first section is dedicated to the auralization of spatial auditory environments and discusses their requirements for sound rendering and auditory design. Auditory environments and augmented audio reality have been introduced and discussed in Chapter 5 and Chapter 6, although in these sections with a focus centered around their definition and application. Section 5.2.2 already motivated the needs for a *non-realistic auditory design* for 3D virtual auditory spaces, at which point this section continues this discussion from a more technical, auralization-centered, perspective.

Essential for an auralization of 3D auditory environments are three *types* of sound:

- Spatialized 3D sound sources,
- Non-spatialized sounds, and
- Room acoustic simulations.

A discussion of their application can be found throughout this research, but is highlighted especially in Section 3.4 and Section 5.2. 3D spatialized sounds are required for all objects with a defined position within a 3D virtual/augmented environment. The spatialization of sound allows later to determine the sounds origin and distance relative to the user, which are both encoded using directional and distance cues. For this task, techniques of HRTF/HRIR convolution (direction) and low-pass filtering (distance) are applied, refer also to Section 3.2.2. Non-spatialized sounds, eg. mono or stereo sounds with an in-head localization, are employed for additional descriptions that are not assigned a specific position within the 3D environment. Examples include the narrators voice from the off, as well as ambient music and feedback sounds from an non-spatial auditory menu system. Room and environmental acoustic simulations are applied to spatialized sounds only, to further integrate and represent them within their local surroundings. As an example, Figure 55 shows an overview of the sound rendering required and also identifies some of the various sound types used. Figure 55 displays 3D positional sound sources as red objects in the scene, as well as shows their direct (red) and secondary (blue) reflections on the room's walls. The two blue objects (window and door) show the position of additional environmental sound objects, which both describe the outer exteriors, eg. the acoustic space on the other side of the window/door.

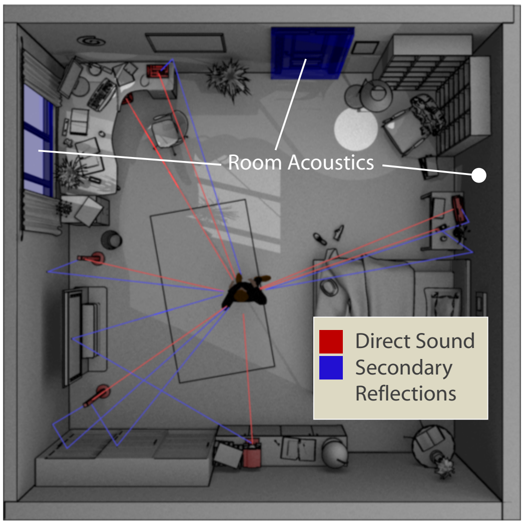


Figure 55: Room Auralization.

An accurate 3D sound rendering along a realistic simulation of a room acoustics is still a very difficult and computationally intensive task. Yet many applications, such as 3D computer games and virtual training scenarios, rely on a realistic acoustic presentation that is able to complement the visual depictions. The difference to audio-only applications and 3D virtual auditory environments is that in these cases the visual image delivers the key information, which is only complemented by the auditory display. That means that an accurate description of an auditory environment requires techniques with a higher quality for both, 3D sound spatialization and acoustic simulation. Almost every application that currently employs either of these techniques is based on 3D sound APIs, with their shortcomings and approximations described earlier (Boer, 2002b). Although the examples and prototypes in this research are also mainly based on OpenAL/EFX and AM3D, this chapter takes a closer look on techniques for a higher quality and more efficient sound rendering and simulation.

The propagations of sound and light seem on a first glance to not have very much in common, but they share, nevertheless, several similarities that can easily be exploited. The following sections examine the possibilities to employ computer graphics techniques and commodity graphics hardware for sound rendering and simulation. Some of the most promising techniques are thereby discussed in more detail and adapted towards a graphics-based sound rendering approach. The motivation for this research is twofold, as on one hand computer graphics and graphics hardware can be used to qualitatively and quantitatively improve existing sound simulations, but also provides a glimpse into a possible future of *programmable* sound hardware. Currently available PC sound hardware is rather fixed in its pipeline and functionality, although some steps were already made into this direction (Aureal, 2000; Creative Labs, 2005). Nevertheless, a fully customizable sound rendering pipeline, similar to the development of graphics hardware over the last decade, would benefit many applications and allow sound simulations and custom effects with a much higher realism by the use of personalized HRTFs, thus increasing the level of immersion in all applications (Röber et al., 2006c, 2007).

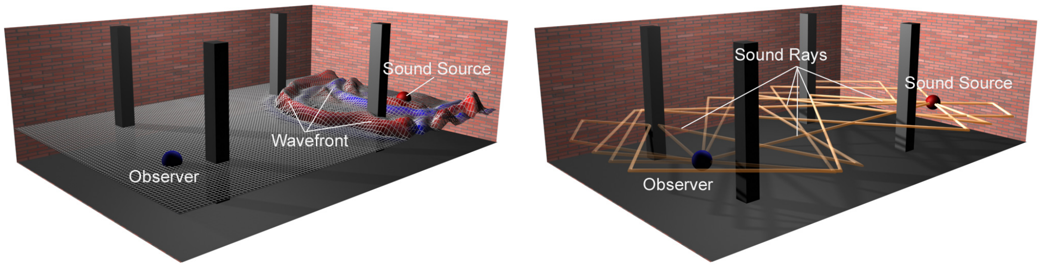
8.1.1 *3D Sound Rendering*

One of the key aspects that is required is an efficient and high-quality 3D spatialization of monaural sound sources. The perceptional and physical principles for this were outlined and discussed in Chapter 3. The focus of this section is to introduce several concepts and techniques to actually perform 3D sound spatialization, with the goal to identify possibilities that can be exploited to improve 3D sound synthesis in both, quality and efficiency.

A very interesting approach is here described using so called *perceptual rendering* techniques, which acoustically display the virtual scene tailored to the auditory senses. This approach can, in conjunction with Section 5.2.2, be described as a non-realistic auditory scene design, as it only considers those parts of the environment that are clearly audible from the user's current position. Research in this area has been conducted by Funkhouser et al. and Tsingos and Drettakis, who both looked at certain techniques to increase the richness of 3D auditory scenes by grouping and classifying sound sources depending on their importance and perception (Funkhouser et al., 1999a; Tsingos and Drettakis, 2004; Tsingos, 2007). Their results can be applied to many areas and used to enhance the display and the perception of 3D virtual auditory environments. However, efficient techniques for the spatialization of monaural sounds are still required.

Most 3D sound systems employ here a convolution using HRIR/HRTF filters, which describe the transformation of a monaural sound into a binaural signal representing the source with its current position, orientation and distance. This function is based on sound reflections along/within the torso/shoulder/head system, and therefore exhibits a strong personalization effect. Current HRTFs that are employed in sound hardware and sound spatialization APIs are based on generalizations using so called *standard ears*. This, however, causes several perceptional artifacts and especially front/back and up/down conversions. A solution to this problem would be the measurement or simulation of personal HRIR filters to increase the accuracy and efficiency for 3D source localization. However, some researchers have also shown that listeners can – in certain cases – adopt to non-personalized HRTFs, which allows the development of *training* applications to accommodate the listener's hearing to adjust to given HRTFs (Hofman et al., 1998; Richardson and Kaiwi, 2002).

The personalization of HRIR filters is still an active and not yet fully resolved area of research. As the measurement is complex and requires dedicated equipment, the majority of approaches focus on a simulation using 3D geometrical models (Kahana et al., 1999; Richardson and Kaiwi, 2002). Two techniques exist, which are often applied for sound simulations are the wave-based and the ray-based approach, see also Section 8.1.2. Two often employed techniques are based on either boundary element methods (BEM), or finite element methods (FEM) (Ise and Otani, 2002). However, as the simulations are very complex and time consuming, an alternative solution has to be found. An idea that was already expressed at the beginning of this chapter is the use of computer graphics hardware and graphics-based techniques for the simulation. The remaining sections of this chapter further develop this idea and explore two possibilities to employ graphics technology to enhance the display of 3D virtual auditory environments, both qualitatively and quantitatively.



(a) Wave-based Approach. (b) Ray-based Approach.

Figure 56: Acoustic Simulation Techniques.

A different approach for 3D sound rendering is the use of ambisonics, or B-Format, for the recording and playback of real/virtual sound fields. The technology was developed in the 1960s and describes a multi-channel recording/playback system that can be used for spatial recordings, but also for the presentation of virtual 3D sound fields (Cooper and Taylor, 1998; Pope et al., 1999). Ambisonics are based on a decomposition of a sound field using spherical harmonics and describes the sound pressure along several different gradients. The first-order B-Format is based on just four channels, but more can be added to increase the accuracy of the system. A great advantage for using ambisonics is that they can efficiently be simulated using computer graphics as well. Dempinski and Viale describe an implementation in 3D computer graphics to illuminate 3D objects using global illumination models (Dempinski and Viale, 2005). The exact same principles, with slight modifications of course, can also be applied for a rendering of 3D sound sources and virtual sound fields, which makes this technology very interesting to enhance any spatial auditory display system.

Besides the rendering and synthesis of 3D sound sources, also the simulation of environmental, or room acoustics is very important. It conveys information about the local surroundings, but it can also be used to detect objects and obstacles in close vicinity. The next section therefore summarizes two of the most widely applied techniques for the simulation of environmental acoustics.

8.1.2 *Acoustic Simulation Techniques*

The two most often used approaches for the simulation of room acoustics are 3D waveguide meshes and ray/beam tracing techniques, see also Figure 56 for a comparative overview. Figure 56a displays a visualization of the waveguide technique, a more physically correct sound propagation model that is based on differential equations. The acoustic energy (eg. pressure) is distributed along nodes using difference equations, which emphasize the applicability of this technique to the simulation of wave-based propagation effects, such as diffraction and interference. Due to its computational high complexity, it is usually only employed for the lower frequency end.

Figure 56b visualizes an alternative approach that is based on energy acoustics and uses ray tracing techniques. This method is based especially on techniques from computer graphics, and is, due to the approximation of sound waves to *sound rays* only applicable to the middle and higher frequency parts.

Over the past years, graphics hardware has inspired several researchers to also deploy it in a large variety of non-graphics applications, including sound rendering and sound simulation (Whalen, 2005; Aszódi and Czuczor, 2002; Jedrzejewski, 2004; Röber et al., 2006c, 2007). Jedrzejewski uses the GPU for simple 2D geometric room acoustics using ray tracing and regular specular reflections (Jedrzejewski, 2004), while Kapralos et al. and Deines et al. employ a particle-based system to adopt the photon mapping technique towards a *phonon tracing* approach (Kapralos et al., 2004; Bertram et al., 2005; Deines et al., 2006b). The aforementioned ray- and wave-based techniques for sound simulation posses a great potential for a hardware-accelerated graphics-based implementation as well. Section 8.3 and Section 8.4 refer to these techniques and discuss a GPU-based implementation that enhances both methods in terms of quality and simulation efficiency.

8.2 Sound Signal Processing

With the availability of programmable graphics hardware and high-level shading languages, graphics programming moved into the focus of many research communities and improved their scientific computations (Owens et al., 2005). The GPU as the core of current graphics hardware can be characterized as a massively parallel streaming processor that has applications in many research areas. The advantages of a GPU-based implementation for sound rendering and simulation are obvious: Not only that the propagation of light and sound shares several similarities which can be easily exploited, but also because the GPU can be straightforwardly turned into a freely programmable DSP for general (sound) signal processing.

Digital signal processing is concerned with the alteration and modification of digital signals, and involves a frequency-dependent amplification or attenuation of certain parts. Digital filters can perform virtually any operation, but are limited by the filter's cost and execution speed. Several different filter types exist in time-domain sound signal processing, with the most common being linear, causal, time-invariant and FIR filters (Zölzer, 2002). A digital filter can be described by its impulse response or its transfer function in the Z-Domain:

$$y(n) = \sum_{i=0}^{N} h_i x(n-i) \tag{8.1}$$

$$H(z) = \sum_{n=0}^{N} h_n z^{-n} \tag{8.2}$$

Here, Equation 8.1 shows the familiar time-based convolution of an input signal with a finite impulse response (FIR) filter of size $N+1$, while Equation 8.2 displays the filter's transfer function within the Z-Domain. When plotted along the Z-plane, $H(z)$ visualizes the zeros and poles of this filter operation and thereby directly the areas of amplification and attenuation of the frequency response (Zölzer, 2002). This is very useful for the design of digital filters, in which poles and zeros can simply be placed along the unit circle to quickly evaluate the filter's frequency response.

Sound signal processing has many applications for the sonification of 3D virtual auditory environments. Impulse responses are used to convolve dry sound files to create a spatialized, binaural impression (HRIR) and to simulate a room's acoustics (RIR). Due to the large number of sound files, the convolution is thereby required to be performed in the most efficient way possible. Nevertheless, several approximations are available and can be used to reduce the computational cost with a minimal impairment of the signal's accuracy. Currently available sound APIs employ a large variety for *emulating* room acoustics and sound spatialization. OpenAL, for instance, uses low-pass filters and several delay lines to simulate obstruction and occlusion effects (Hiebert, 2006; Peacock et al., 2006). Hall and echo, again with low- and band-pass filters applied, are employed to create an illusion of room acoustics.

Using the earlier discussed advantages of commodity graphics hardware, the following section explores the possibilities for utilizing computer graphics techniques and high-level shading languages for a general (sound) signal processing.

8.2.1 *The GPU as Digital Signal Processor*

The GPU as a stream-based and freely programmable processor is highly suited for a general time-domain signal processing. Two publications that already discuss these possibilities have been presented by Gallo and Tsingos and Whalen (Gallo and Tsingos, 2004; Whalen, 2005). Gallo and Tsingos employ the GPU for 3D sound spatialization and measured a slight increase in performance compared to a regular CPU implementation (Gallo and Tsingos, 2004). Whalen concentrated his efforts on a classic DSP approach and implemented several convolution-based signal processing effects using fragment shaders (Whalen, 2005). A direct comparison with a CPU implementation revealed, however, that the CPU outperformed the GPU at several occasions. The work of both authors is, however, based on earlier GeForceFX GPUs, which also both describe as inefficient for audio processing due to limitations in texture-access modes, shader-length and -complexity, as well as a too slow AGP bus connection (Whalen, 2005; Gallo and Tsingos, 2004). Since then, several improvements have been made and current graphics technology with its unified shader architecture and high-speed PCIe bus connection performs very well for the manipulation and filtering of digital signals.

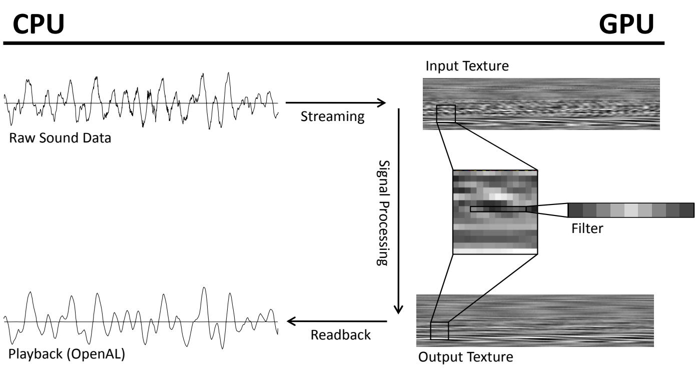


Figure 57: GPGPU-based Signal Processing.

Utilizing the GPU for time-domain DSP filtering operations is very similar to the general GPGPU[1] approach, as is illustrated in Figure 57. The input signal is transferred into graphics memory and loaded as a floating point texture with each texel representing one sample. A second texture holds the convolution kernel, while a third texture is used for the filter result. The convolution itself is implemented using fragment shaders and applied per texel of the input texture.

GPU-based Signal Processing.

A shader example that performs a box average can be seen in Listing 8.1. In this example, `texture` contains the 1D input signal and is stored in a 2D texture array with `length` being half the filter's size. The return value is the original texture, in which the red channel now holds the averaged filter signal. Additional examples for employing the GPU as general DSP can be found in Section A.1, which also discusses shader examples for the later discussed GPU-based sound simulations.

The enormous efficiency advantage of a GPU-based implementation is primarily the result of a parallel computation. This approach, however, does not permit a realization of all types of digital filters in graphics hardware. Some have to be implemented as several

1 GPGPU = General Purpose computations using a Graphics Processing Unit (GPU)

```
float4 convolution(float2 coords : TEX0, uniform sampler2D texture) : COLOR
{
  float4 s1 = tex2D(texture, coords);

  //----- median filtering with size 2*length+1 -------//
  float tmp = 0.0;
  for (int i=(-1*length) ; i<(length+1) ; i++)
    tmp += (0.5-tex2D(texture, sample(coords, i)));

  float tmp2 = 1.0 + (((length*2)+1) / 100.0);
  s1.r = saturate(0.5+(tmp2*(tmp / (1+(2*((length*2)+1))))));

  return s1;
}
```

Listing 8.1: GPU Signal Processing (Convolution).

rendering passes, or require a pass-in of additional parameters/computations performed on the CPU (Micea et al., 2001; Röber et al., submitted). Table 6 shows the efficiency of several examples, which have been implemented and tested on the CPU, as well as in graphics hardware using fragment shaders and the approach displayed in Figure 57. The efficiency is measured in *fps*, and refers to the number of sound samples (44.1kHz resolution, 1s length) processed per second. It shows the raw processing efficiency only, without data streaming or any other computations. Two GPUs were evaluated to additionally assess the differences in available graphics hardware.

The results in Table 6 show clearly that older graphics technology is not always able to compete with the CPU (Whalen, 2005), but also that newer graphics hardware with its unified shader architecture possesses enormous computational capacities.

Time-domain signal processing filters are easy to implement and can also be combined with sound simulation techniques to perform a binaural sound rendering. However, convolutions with large kernel sizes are still very time consuming. A possible solution is a filtering in the frequency domain, at which the following section discusses an alternative approach using an implementation of frequency band decompositions and a signal filtering with adjustable frequency weights.

Filter Type	CPU-based	GPU-6800GT	GPU-8800GTX
Volume	3,400 fps	11,200 fps	90,500 fps
Normalize	3,150 fps	10,600 fps	96,250 fps
Convolution (5)	1,600 fps	1,320 fps	43,500 fps
Convolution (25)	501 fps	300 fps	11,300 fps
Convolution (125)	90.5 fps	64.1 fps	2,560 fps
Pitch / Resampling	1,700 fps	1,880 fps	58,500 fps
6-Tap Delay	1,450 fps	1,600 fps	25,800 fps
Chorus / Flanger	2,650 fps	3,600 fps	70,000 fps
Compressor / Limiter	820 fps	1,120 fps	30,300 fps

Table 6: CPU vs. GPU - Signal Processing Efficiency (fps per 44.1kHz/1s).

f_j	f_{range_j}	f_{center_j}	λ_{center_j}
f_0	22 Hz — 44 Hz	31.5 Hz	10.88 m
f_1	44 Hz — 88 Hz	63 Hz	5.44 m
f_2	88 Hz — 177 Hz	125 Hz	2.74 m
f_3	177 Hz — 354 Hz	250 Hz	1.37 m
f_4	354 Hz — 707 Hz	500 Hz	0.68 m
f_5	707 Hz — 1,414 Hz	1,000 Hz	0.343 m
f_6	1,414 Hz — 2,828 Hz	2,000 Hz	0.172 m
f_7	2,828 Hz — 5,657 Hz	4,000 Hz	0.086 m
f_8	5,657 Hz — 11,314 Hz	8,000 Hz	0.043 m
f_9	11,314 Hz — 22,627 Hz	16,000 Hz	0.021 m

Table 7: Frequency Bands f_j.

8.2.2 *Frequency-based Filtering*

Two approaches are applicable for a frequency-dependent sound signal processing, either with or without the use of a Fourier transform. The straightforward approach uses complex textures and a CPU- or a GPU-based FFT for a frequency domain conversion of the signal data (Ritter et al., 1999; Govindaraju and Manocha, 2007). The phase and amplitude are stored within complex textures and can be processed and manipulated using fragment shaders, analog to Figure 57. After filtering and an inverse transformation, the signal data is streamed back to main memory for further processing or playback. An alternative approach is to decompose the signal in a pre-processing step using windowed sinc filters into several frequency bands (Table 7), and to process each of these bands individually using a frequency-dependent weighting function. This still allows an efficient processing and implementation, but in respect to a filters frequency behavior and without the necessity of time-consuming FFT conversions.

Frequencies of the audible spectrum are classified and described by frequency bands (octaves) according to human psychoacoustics (Vorländer, 2007; Goldstein, 2007). Table 7 provides an overview of the different frequency bands, along their index number, frequency range f_{range_j}, center frequency f_{center_j} and center wavelength λ_{center_j}. The audible spectrum is therefore defined as the sum of these 10 frequency bands:

$$A_{spectrum} = A_s = \sum_{j=0}^{9} f_j \qquad (8.3)$$

HRIR convolutions are a good application to describe this filtering approach. HRIRs are a collection of FIR filters that are dependent on direction, distance and time, and are used for 3D sound spatialization. Prior to the binaural rendering of a scene, all HRIRs and footage sounds need to be decomposed into these 10 frequency bands. Later, the acoustic energy of each direction is filtered and delayed with its associated HRIR. In a second step, the energy of all bands are accumulated according to Equation 8.3. Compared with the last section, this approach is preferable if a frequency-dependent weighting is required. Besides the pre-processing step for the initial signal decomposition, this technique is still very efficient, easy to implement, and allows a frequency-dependent modeling of materials and sound/object interactions.

8.3 3D WAVEGUIDE TECHNIQUE

As outlined in Section 8.1, the waveguide technique is a commonly employed method in room acoustics to simulate the propagation of sound waves using numerical techniques and time-domain difference models. Waveguides have been originally developed for the simulation of string-based musical instruments (VanDuyne and Smith, 1993; Smith, 1992), and were later also applied to model the vibrations of air in room acoustic simulations (Savioja et al., 1995). Waveguides and 3D waveguide meshes are very well suited for the simulation of sound wave propagation, but require a huge amount of sampling points (nodes) in order to achieve realistic results. This makes this approach technically only applicable to the lower frequency end, as with higher frequencies the necessary sampling distance decreases and the number of nodes required rises cubically.

As motivated throughout the last chapters, the quality and efficiency of sound rendering is of the highest importance, especially for an acoustic display of 3D auditory environments. Here not only techniques for 3D sound spatialization are required, but also methods for a realistic simulation of room acoustics and an auditory presentation of the local environment. The waveguide technique offers here a promising approach that can be improved in terms of quality *and* efficiency by using computer graphics and commodity graphics hardware. Current graphics applications typically require the processing of huge amounts of data, for which graphics hardware has been optimized to support using a highly parallel design. This makes this hardware, along with its high-level programming techniques, very interesting for parallel computing problems, such as wave propagations using waveguide meshes. The idea is to develop a technique that allows a fast and – if possible – real-time implementation of 3D waveguide meshes for the lower and mid frequency ranges. The simulation results can then directly be applied for an environmental presentation of 3D auditory spaces. Using additional techniques for a non-realistic auditory design, this setup should provide the environmental information required to navigate and orient oneself through 3D virtual auditory environments.

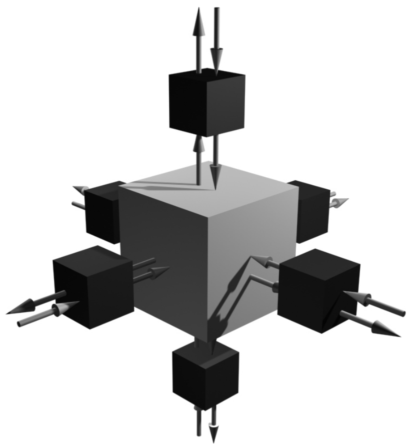


Figure 58: 3D Waveguide Node.

8.3.1 *Waveguide Meshes*

Waveguide meshes are an extension of the waveguide technique and constructed by bi-linear delay lines that are arranged in a mesh-like structure (VanDuyne and Smith, 1993). Each node in the mesh is defined as scattering junction and acts as spatial and temporal sampling point for the wave and energy propagation. Scattering junctions are thereby of equal impedance with two main constraints in effect:

- The sum of all inputs is equal to all outputs, and
- The pressures in each crossing waveguide are equal at the junction.

By assuming a lossless scattering, the *acoustic pressure* v_J is determined by adding all incoming wave components v_i^+ according to Equation 8.4 (VanDuyne and Smith, 1993).

The relationship between the incoming v_i^+ and outgoing v_i^- components is expressed by Equation 8.5.

$$v_J = \frac{2\sum_i R_i v_i^+}{\sum_i R_i} \tag{8.4}$$

$$v_i^- = v_J - v_i^+ \tag{8.5}$$

For a homogenous N-dimensional rectilinear mesh, in which each junction connects to 2N neighbors, Equation 8.4 is reduced to:

$$v_J = \frac{\sum_i v_i^+}{N} \tag{8.6}$$

By discretizing time and space one obtains the difference equations that govern the wave propagation within an N-dimensional rectangular mesh with:

$$v_{J,k} = \frac{\sum_l v_{J,l}(n-1)}{N} - v_{J,k}(n-2) \tag{8.7}$$

In this equation, k identifies the current node, n represents the discretized time steps and l is associated with neighboring nodes. In order to simulate boundary conditions, so called 1D termination nodes with only one neighbor are employed to simulate phase-reversing and -preserving reflections, but also non-reflective, anechoic walls (Savioja et al., 1995). Certain atmospheric absorption effects can be emulated using an additional factor, but are here ignored due to an application in interior acoustics only.

A major problem with digital waveguide meshes is a non-isotropic speed for energy propagation. This is also known as frequency dispersion and varies between different mesh topologies. The dispersion error for the rectilinear grid ranges from zero along the diagonals to its maximum extent along the coordinate axes, and is quantified as:

$$k_d(\underline{\beta}) = \frac{c'\underline{\beta}}{c} \tag{8.8}$$

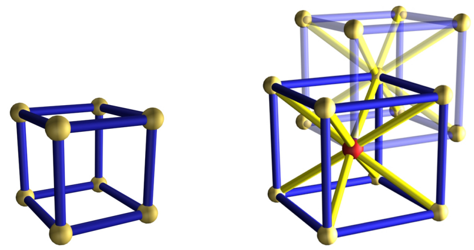
with c being the speed of propagation in a dispersion free environment, and $c'\underline{\beta}$ the actual speed in the direction of $\underline{\beta}$. The analytical expressions for k_d can be derived using Von-Neumann analysis (Bilbao, 2004; Campos and Howard, 2005; Fontana and Rocchesso, 2001), and show that the dispersion error for the 3D rectilinear mesh ranges on a spherical surface between $0.927 < k_d < 1$, with $k_d = 1$ along the diagonal axes. This causes a distortion of the initially spherically bandlimited signal along the coordinate axes, which also impairs the simulation results.

8.3.2 *Optimal Sampling*

The rectilinear Cartesian lattice is in terms of sampling efficiency not the most optimal grid available. Under the assumption of an isotropic spherically bandlimited signal, hexagonal lattices provide a much better packing density. A denser packing of spectra in the frequency domain translates to an increase in sampling distance in the spatial domain. The Body-Centered-Cubic grid (BCC) represents such a hexagonal lattice, see also Figure 59b. A direct comparison with the Cartesian grid in Figure 59a reveals that a BCC node has 8 nearest neighbors and is constructed with an additional sampling point in the cell center. Due to the more optimal sampling, the internodal distance can be increased to $\sqrt{1.5}$, which in 3D results in roughly only 70 % of the sampling points required to represent the same information (Conway and Sloane, 1976). Hexagonal grids are also very common in nature, as bees, for example, build their honeycombs using hexagonal cells and as a result achieve a minimum expenditure of wax.

An advantage of the BCC grid is that it can be decomposed into two cubic grids that intertwine and are offset by half the sampling distance, see Figure 59b. More generally, the N dimensional hexagonal lattice can be constructed by two N-1 rectilinear grids that are shifted by $\sqrt{2}$ in all N dimensions. This characteristic allows an easy indexing of the data and is also of high importance for a later implementation in graphics hardware. The BCC lattice is already known and used in computer science for image processing and scientific visualization (Theußl et al., 2001b; Röber, 2002), but has its main roots in chemistry and crystallography (Jackson, 1991; Wells, 1984a).

The BCC lattice offers several advantages for the simulation of room acoustics using digital waveguide meshes. The inherent optimal sampling requires only about 70 % of the original sampling points, and thereby directly reduces the computational load and data storage required. Additionally, as the BCC grid has 4 delay units per node, a different and lower frequency dispersion error can be expected.



(a) Cartesian Lattice. (b) BCC Lattice.

Figure 59: Cartesian and Body Centered Cubic Lattice.

The equations that govern the propagation of sound waves using the BCC lattice are based on difference equations derived from the Helmholtz equation by discretizing time and space. With now 4 principal axes they are transformed to:

$$\begin{aligned} p(t+1,w,x,y,z) &= \\ &\tfrac{1}{4}[p(t,w+1,x,y,z)+p(t,w-1,x,y,z) \\ &+p(t,w,x+1,y,z)+p(t,w,x-1,y,z) \\ &+p(t,w,x,y+1,z)+p(t,w,x,y-1,z) \\ &+p(t,w,x,y,z+1)+p(t,w,x,y,z-1)] \\ &-p(t-1,w,x,y,z) \end{aligned} \tag{8.9}$$

in which p is the pressure at point (w,x,y,z) at time step t. Because of the increased sampling distance, the unit length in the BCC lattice is extended to $\sqrt{1.5}$, which also changes the update frequency f_{update} to:

$$f_{update} = \frac{c\sqrt{2}}{\Delta x} \approx \frac{480.8}{\Delta x}\ \text{Hz} \tag{8.10}$$

As a result, the BCC lattice propagates sound waves *faster* than the rectilinear Cartesian lattice, which has to be considered in virtual impulse response measurements. The

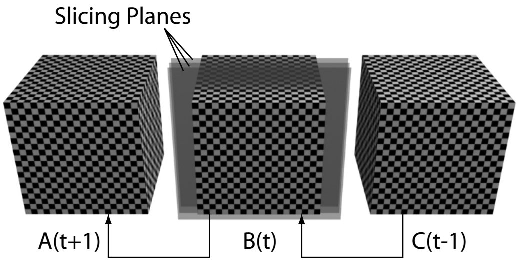


Figure 60: 3D Waveguide Mesh – Rendering Principle.

frequency dispersion factor k_d has for a spherical surface a range of $0.953 < k_d < 1$ (Campos and Howard, 2005). At its maximum extent, the error is only 4.7 %, compared to the 3D rectilinear grid with 7.3 % in the direction of $\beta_n = \pi/2$. The spatial bandwidth can be determined by decomposing the BCC lattice into two rectilinear grids of spacing d. The sampling efficiency of the BCC lattice compared to the 3D rectilinear grid therefore is:

$$\frac{\mu_{BCC}}{\mu_{CC}} = \frac{1}{\sqrt{2}} \approx 0.707 \tag{8.11}$$

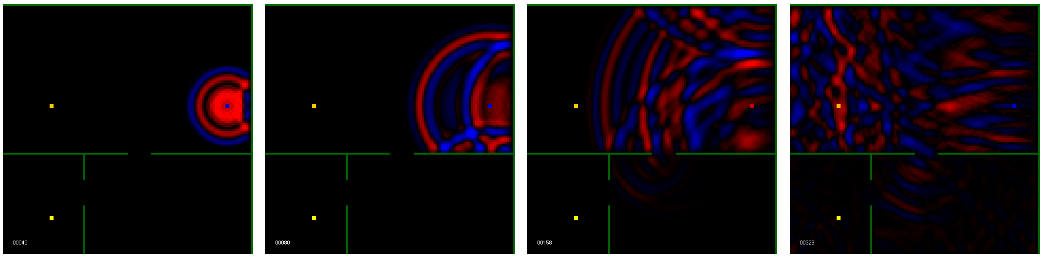
Although, in a direct comparison with the rectilinear Cartesian grid, the computational effort per node is slightly higher, overall it is still much more efficient and also exhibits a less pronounced dispersion error.

8.3.3 *GPU-based Implementation of 3D Waveguide Meshes*

) 3D waveguide meshes are easy to implement and realize in graphics hardware using a high-level shading language. The technique is mainly based on 3D 32-bit floating point textures, fragment shaders, as well as OpenGL's framebuffer objects (FBO). The BCC waveguide mesh can be decomposed into two rectilinear 3D textures, which are loaded and stored separately into texture memory. Both grids with time frames $t-1$ and t reside here in just one single RGBA texture. The base grid is loaded into the *Red* and *Green* channel, while the offset grid is placed in *Blue* and *Alpha*. This allows to compute two nodes, eg. one BCC cell, in one step. The channels are directly rendered into a framebuffer object in an alternating fashion, having one texture attached to it as the primary render target. Figure 60 visualizes the method's principle. The three textures A,

Mesh Size	2D-CC CPU	2D-CC GPU	3D-CC CPU	3D-CC GPU	3D-BCC GPU
$64 \times 64 \times 24$	32,000 fps	9,800 fps	238.0 fps	1,358.0 fps	1880.0 fps
$128 \times 128 \times 24$	7,500 fps	6,330.0 fps	16.1 fps	990.0 fps	1240.0 fps
$256 \times 256 \times 24$	2,000 fps	5,024 fps	4.1 fps	322.0 fps	430.0 fps
$512 \times 512 \times 24$	456.6 fps	2,670.0 fps	0.9 fps	88.3 fps	121.0 fps
$768 \times 768 \times 24$	213.7 fps	1.377 fps	0.28 fps	39.9 fps	55.2 fps
$1024 \times 1024 \times 24$	123.06 fps	830.0 fps	– fps	19.1 fps	31.6 fps
$2048 \times 2048 \times 24$	29.93 fps	221.0 fps	– fps	– fps	– fps
$4096 \times 4096 \times 24$	7.9 fps	56.5 fps	– fps	– fps	– fps

Table 8: Waveguide Mesh Efficiency – CPU vs. GPU (fps).



(a) Wavefront at $t = 40$. (b) Wavefront at $t = 80$. (c) Wavefront at $t = 160$. (d) Wavefront at $t = 320$.

Figure 61: 3D Waveguide Meshes (BCC Lattice).

B and C contain the waveguide node data at their respective time frames. Texture B is *sampled* using texture aligned slicing planes that have the same resolution as the texture itself. During this sampling, and for every time frame, the fragment shader solves the difference equations according to Equation 8.9 for each voxel in the waveguide mesh and stores the results into the next buffer (eg. A).

A second channel contains additional scene information, such as scene geometry and material specifications to model basic boundary conditions. An example of the fragment shader discussed can be found in Section A.2, which shows the implementation using both BCC nodes, an implementation of a sound source, as well as simple phase-reversing reflections.

8.3.4 *Results and Efficiency*

Several experiments have been performed to evaluate and assess the quality and efficiency of a graphics-based implementation of 3D waveguide meshes. The experiments have been performed on a regular PC equipped with a P4 3GHz processor, 1GB of main memory and an *n*Vidia GeForce 8800GTX graphics accelerator. The performance results can be seen in Table 8, although some results are missing due to insufficient texture memory and CPU performances. For comparison reasons, also the efficiency of a simple, non-optimized CPU implementation is shown. The results clearly show the advantages of a graphics-based implementation, as especially for larger meshes, the CPU is outperformed by a substantial factor. Table 8 also shows that the implementation of the BCC lattice is faster by a factor of approximately $\sqrt{2}$, the number of samples *saved* due to the more optimal sampling used. An earlier implementation of this approach was impaired by the non-availability of 3D)framebuffer objects, which massively decelerated the performance due to a forced computation in 2D and a heavy texture copying (Röber et al., 2006c). Although wave-based room acoustics can be computed much faster and with the same accuracy using this technique, one limitation that applies is the texture memory available. Dedicated GPGPU hardware provides today up to 1.5GB of RAM, and allows thereby with this technique for over $100,000,000$ waveguide nodes (Nvidia, 2007). This should be sufficient for the modeling of most scenarios and can even be pushed further using more efficient texture packing techniques.

3D Waveguides CC.

3D Waveguides BCC.

Figure 61 shows several time frames for a setting of three rooms with reflecting walls and ceiling. It was used to compare the quality, as well as the efficiency of both implementations. The sound source is marked by a blue dot, the two microphones by yellow dots, while walls are identified by green color. The data set in this example has a size of $128 \times 128 \times 24$ for the rectilinear Cartesian grid and a size of $92 \times 92 \times 34$ for the BCC lattice. The sound source, walls and microphone positions are adjusted accordingly to fit the BCC's dimensions. The rendering efficiency is for the rectilinear grid on average of 990 fps, and for the BCC lattice 1,240 fps. Both meshes were excited using a single sine wave, whose frequency was adjusted for the BCC lattice according to Equation 8.10.

After this detailed examination of wave-based acoustic simulations, the following section explores the applicability of a graphics-based implementation for ray-based acoustic simulations.

8.4 RAY/ENERGY ACOUSTIC SIMULATIONS

Geometrical acoustics is often referred to as ray/energy acoustics, and is in this respect very similar to optical models used in computer graphics. Ray acoustics thereby approximates sound waves as particles that are moving along directional rays and adopts existing ray tracing techniques for sound simulation. Due to this approximation, wave-based propagation effects, such as diffraction and interference, are usually not considered and therefore not part of the simulation. Ray acoustics is therefore only applicable to frequencies, whose wavelength are much shorter than the dimensions of the enclosure or any object within (Everest, 2001; Kuttruff, 2000).

Ray tracing is a long known area of research in computer graphics and has seen many improvements and enhancements. Lately, with the introduction of dedicated hardware, ray tracing shifted from an offline simulation towards an interactive and realtime rendering system (Purcell et al., 2002; Purcell, 2004; Moreno-Fortuny, 2004). Many advancements from the visual realm can also be mapped and beneficially applied to ray acoustics simulation as well. Similar to the last section, computer graphics and graphics hardware can aid ray-based sound simulations to perform faster and with a higher quality. However, due to several differences in light and sound wave propagation, spectral and temporal effects have to be integrated into the simulation model and accounted for by the ray tracing system.

Several articles about ray/energy acoustics have been published, of which some already discuss the realtime possibilities of a ray-based acoustic simulation system (Funkhouser et al., 2002c; Savioja et al., 2002; Tsingos and Drettakis, 2004). The majority, however, concentrates only on specular reflections using a ray/beam tracing-based approach and uses conventional 3D sound APIs for sound spatialization and rendering (Wand and Straßer, 2004; Neumann, 2004; Jedrzejewski, 2004).

To the author's knowledge, none of the existing ray-based sound simulations were so far examined regarding an application of global and local illumination models towards ray/energy acoustics in greater detail. The next sections therefore provide an in-depth analysis of computer graphics and ray tracing techniques, and concentrate especially on building a foundation for ray/energy acoustics by extending models used in computer graphics towards a time- and frequency dependent acoustic energy propagation.

8.4.1 *Acoustic Energy Propagation*

Sound is defined as mechanical energy and propagates through pressure variations within a participating media, and can be described by attributes such as frequency, wavelength and speed of propagation. Light on the other hand is an electromagnetic radiation that is characterized by similar, however, largely different quantities. In order to study and describe the propagation of sound waves using ray tracing techniques, an adequate propagation model that includes time- and frequency dependencies needs to be defined. Such a model can be developed in analogy to the physics of light transportation by using and extending the tools of radiometry to also include spectral and temporal propagation effects (Dutre et al., 2003; Beranek, 1986).

Acoustic energy is described as the amount of pressure variations per unit volume and time, or, more precisely, by the changes in velocity of air particles contained in a volume element per unit time. Acoustic energy is quantitatively measured in $Watt$ or $Joule/sec$ and described as radiant power Φ or flux (Dutre et al., 2003). The intensity is thereby defined as the amount of acoustic energy that flows from/to/through a surface element per unit time:

$$I(t) = \frac{d\Phi}{dA} dt \tag{8.12}$$

The energy density in the medium of propagation is defined as the sum of its kinetic and potential energy per unit volume dV and time: $E(t) = E_{kin}(t) + E_{pot}(t)$ (Beranek, 1986). The kinetic energy density, or sound wave pressure, is therefore described as:

$$E_{kin}(t) = \frac{1}{2}\frac{Mc^2}{V_0} dt = \frac{1}{2}\rho_0 c^2 dt \tag{8.13}$$

with c being the average velocity of air particles, ρ_0 the average media density and $\frac{M}{V_0}$ the mass per unit volume. The potential energy density can be derived from the gas law as:

$$E_{pot}(t) = \frac{\int p dp}{c^2 \rho_0} dt = \frac{1}{2}\frac{p^2}{c^2 \rho_0} dt \tag{8.14}$$

with p as the pressure of the sound wave and c the speed of sound in this medium. The total amount of acoustic energy density is therefore described as (Beranek, 1986):

$$E(t) = E_{kin}(t) + E_{pot}(t) = \frac{1}{2}(\rho_0 c^2 + \frac{p^2}{c^2 \rho_0}) dt \tag{8.15}$$

With Equation 8.15 being sound at any position and time within the virtual auditory environment, it serves as basis for defining an acoustic energy propagation model. In order to quantitatively measure flux per unit projected surface area and per unit angle, radiance is introduced with (Dutre et al., 2003):

$$L(x, \Theta) = \frac{d^2\Phi}{d\omega dA \cos\theta} \tag{8.16}$$

and varies with position x and the ray's direction Θ. By incorporating the wavelength λ_j of the frequency bands used (refer to Table 7), Equation 8.16 changes to:

$$L(x, \Theta, f_j) = \int_{A_s} L(x, \Theta, f_j) d\lambda \tag{8.17}$$

The acoustic energy that interacts with a surface element is further differentiated in incident E_i (incoming) and exitant E_e (outgoing) energy:

$$E_i = \frac{d\Phi}{dA}, E_e = kE_i \tag{8.18}$$

The scalar k (defined over $[0, 1]$) hereby describes the reflectivity of a surface element with $E_{surface} = E_i - E_e$, and is affected by the surface material definitions. Using a lossless participating media, the exitant radiance at one point $L(x_1 \rightarrow \Theta)$ is exactly the same as the incident radiance at another point receiving this amount of energy $L(x_2 \leftarrow \Theta)$ (Dutre et al., 2003). This approach of reciprocity is also sound in real world acoustics and employed in highly efficient *inverse* HRIR measurements (Li et al., 2004). Using a density function and volume elements, $p(x)dV$ defines the physical number of *sound particles* that are carrying an *acoustic energy quant*. If moved in time dt across a differential surface area dA, and by using the direction ω and speed of propagation c:

$$N = p(x, \omega, f_j)\ c\ dtdA\ \cos\theta\ d\omega d\lambda \tag{8.19}$$

describes the number of particles flowing through this surface element. The radiance per unit volume is accordingly redefined to:

$$L(x, \Theta, f_j) = \int_{A_s} \int p(x, \omega, f_j) h \frac{c}{\lambda_j} d\lambda \tag{8.20}$$

In this model, acoustic energy radiates from a sound source (speaker) using a certain emittance pattern. This pattern can be homogenous in any direction (eg. spherically), or direction dependent (eg. cone shaped). Similar to light, acoustic energy also attenuates with distance using the familiar inverse square law and through atmospheric absorption effects. For small enclosures, this factor can safely be ignored, but becomes more prominent with increasing distances. To sample the acoustic energy present at a certain location, an observer, or listener is required. This listener does not interfere or participate in the energy propagation, but, if required, such as in a binaural listening simulation, additional geometry can be used to emulate head-shadowing effects.

8.4.2 *Local acoustic Energy Exchange*

The most interesting part in a ray-based acoustic simulation is the interaction and exchange of acoustic energy with objects and surface elements. Depending on the objects size and the acoustic material parameters specified, some of the incoming energy might get absorbed, reflected, refracted or transmitted, with the total amount of energy, according to Equation 8.18, being constant. Figure 62 shows a schematic of the local acoustic energy exchange. Every ray that is cast into the scene contains, depending on the sound source emittance, energy from all frequency bands. The contribution of each ray is evaluated at the point of intersection with the surface element using the ray's length, its frequency spectrum, as well as the surface material properties defined.

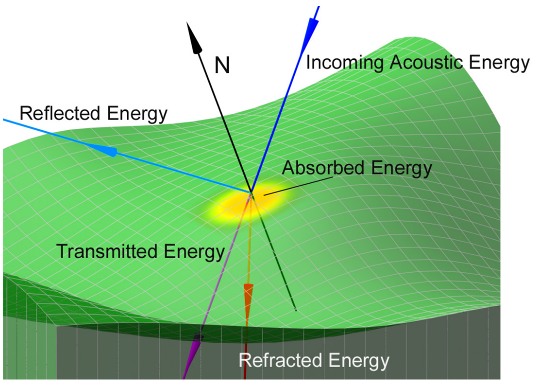
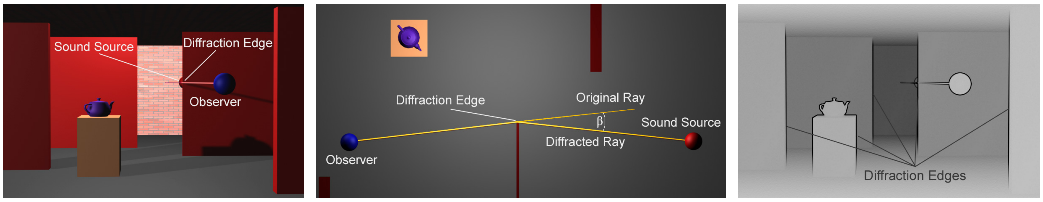


Figure 62: Acoustic Energy Exchange.

Parts of the incident energy are usually absorbed and *removed* from the system. The absorption is frequency dependent and characterized through a coefficient per frequency band α_{f_j}:

$$L_{e_{absorbed}}(x \leftarrow \Theta) = \sum_{j=0}^{9} E_{i_j} \alpha_{f_j} \tag{8.21}$$

The diffraction of sound waves is also frequency dependent. Sound waves are thereby simply bend around objects smaller than their wavelength, but continue unchanged otherwise. In this case, transmission is redefined to describe the amount of energy



(a) Normal Scene Rendering. (b) Top View – original and diffracted Ray. (c) Combined Depth/Edge Map.

Figure 63: Ray Acoustic Diffraction Simulation.

that passes *through* an object unaltered and without refraction. A frequency dependent modeling can be implemented similar to Equation 8.21, which describe the transmission of acoustic energy whose wavelength is equal or above a certain cutoff wavelength set by the objects bounding box:

$$L_{e_{transmitted}}(x \rightarrow (\pi + \Theta)) = \sum_{j=0}^{9} E_{i_j} \tau_{f_j}. \tag{8.22}$$

Here $L_{e_{transmitted}}(x \rightarrow (\pi + \Theta))$ describes the amount of exitant energy per ray for all bands, which simply passes along the direction opposite to the incoming ray, ie. the ray's original direction. The term τ_{f_j} is used for a finer modeling and a frequency-weighting of the transmission effects.

Reflection and diffuse scattering are probably the two most important qualities in an acoustic ray tracing simulation system, and can be very well described using bidirectional reflection distribution functions (BRDF) (Dutre et al., 2003). A BRDF is defined for a point x as the ratio of the differential radiance reflected in an exitant direction Θ_e and the differential irradiance incident through an incoming angle Θ_i:

$$brdf_{reflected}(x, \Theta_i \rightarrow \Theta_e) = \frac{dL(x \rightarrow \Theta_e)}{dE(x \leftarrow \Theta_i)} \tag{8.23}$$

The BRDF is thereby frequency dependent, but direction independent, eg. $f_r(x, \Theta_i \rightarrow \Theta_e) = f_r(x, \Theta_e \rightarrow \Theta_i)$ (Dutre et al., 2003; Li et al., 2004). Diffuse scattering hereby uniformly reflects the incoming acoustic energy in all directions. In acoustics, this behavior is largely influenced by the surface roughness μ, which can be used to derive a specular reflection coefficient that describes the ratio between specular and diffuse reflections. A frequency dependent BRDF for acoustic ray tracing includes all frequency bands, and can be described through:

$$L_{e_{reflected}}(x \leftarrow \Theta_i) = \sum_{j=0}^{9} E_{i_j} \upsilon_{f_j} clamp(\frac{\mu}{f_j}, 0, 1) \tag{8.24}$$

in which υ_{f_j} is an additional weighting factor per frequency band f_j. True refraction effects can be computed similarly to Equation 8.23, in which the outgoing angle Φ of the refracted ray can be determined using *Snell's Law*. A frequency band weighted refraction can be defined in a similar way to Equation 8.24.

8.4.3 *Acoustic Ray Tracing and Diffraction Modeling*

Figure 64 shows the auralization pipeline of the graphics-based ray acoustics system. The 3D scene geometry is converted into a uniform grid structure in a pre-

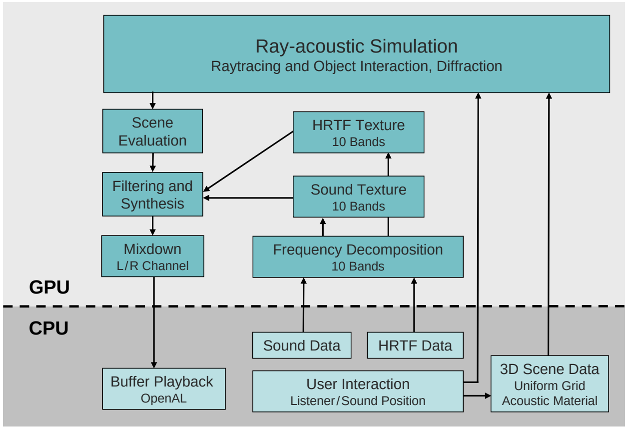


Figure 64: Auralization Pipeline.

processing step. It subdivides 3D space and groups neighboring triangles in an axis aligned uniform voxel-based topology (Purcell et al., 2002; Purcell, 2004; Moreno-Fortuny, 2004). This re-structuring is necessary for a more efficient ray/object intersection testing, as now only triangles from the same voxel element have to be evaluated. In a second step, footage sounds and HRIR filters are loaded as floating point textures into graphics memory and are decomposed using windowed sinc filters with their cutoff frequencies set to the bands respective border frequencies, refer also to Section 8.2.2 and Table 7. Furthermore, sound data is assigned to virtual speakers as well as a position and an emittance pattern.

The actual casting of rays and the energy accumulation is performed using cube maps that are centered around the listeners position (Kaminski, 2007), refer to Figure 65. These cubemaps sample the scene with one ray cast per cubemap texel. Each ray is thereby traced through the virtual scene and its acoustic energy is accumulated and stored per frequency band within the cubemap texel. At points of ray/object intersection, the local surface acoustic energy exchange is evaluated according to Section 8.4.2. Secondary rays, emanating from points of refraction, transmission and/or reflection, are further traced until their energy contribution falls below a certain threshold ϵ. The final sound mixdown is performed binaurally using the previously decomposed sound data and HRIR filters. According to the accumulated energy spectrum in the cubemap texture, the sound data is weighted and delayed per frequency band. The frequency bands of all contributing rays are accumulated and stored in a two channel texture (binaural sound buffer), refer to Figure 64. This sound data is streamed back to the CPU and fills a native OpenAL stereo buffer for playback. All convolutions, simulations, ray tracing, mixdown and sound rendering are performed using fragment shaders in graphics hardware, with a single shader for each task. Several of these shaders and additional code examples are discussed in more detail in Section A.3.

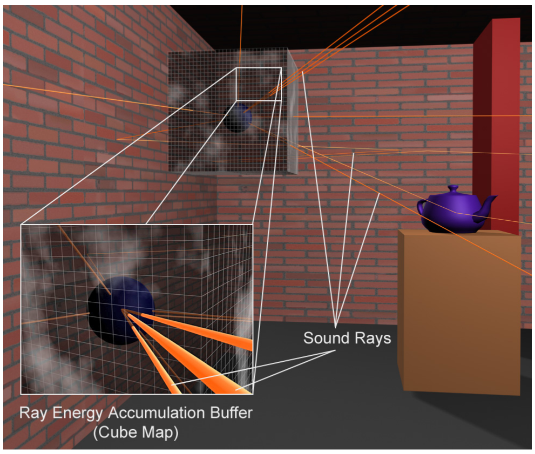


Figure 65: Cubemap ray tracing/Sampling.

Diffraction and interference are both two phenomena that can be well modeled using wave-based simulation techniques, but are difficult to capture using ray-based approaches. Frequency-dependent diffraction effects, however, can be approximated using a ray bending technique, in which the outgoing ray is bent according to the ray's associated frequency band f_j. Here, lower frequencies diffract stronger than higher frequency bands.

		Size of Cubemap				
		8×8	16×16	32×32	64×64	128×128
Possible Number of Directions		384	1,536	6,144	24,576	98,304
Simple Box (80 pol.)	ray tracing only	75.3 fps	72.5 fps	70.2 fps	68.4 fps	63.2 fps
	with auralization	23.1 fps	9.7 fps	3.4 fps	0.97 fps	0.27 fps
Church (800 pol.)	ray tracing only	55.1 fps	44.3 fps	37.8 fps	24.9 fps	18.9 fps
	with auralization	12.8 fps	6.2 fps	2.6 fps	0.77 fps	0.24 fps
Apartment (1,400 pol.)	ray tracing only	43.1 fps	42.1 fps	42.5 fps	34.2 fps	26.8 fps
	with auralization	15.8 fps	7.5 fps	3.1 fps	0.88 fps	0.25 fps
KEMAR (5,500 pol.)	ray tracing only	16.8 fps	16.4 fps	16.4 fps	16.4 fps	16.0 fps
	with auralization	11.2 fps	6.7 fps	3.0 fps	0.89 fps	0.25 fps
Large Hall (37,000 pol.)	ray tracing only	4.2 fps	4.2 fps	4.2 fps	4.2 fps	4.1 fps
	with auralization	3.7 fps	3.0 fps	1.9 fps	0.76 fps	0.23 fps

Table 9: Ray Acoustics Efficiency (fps per 44.1kHz).

The amount of diffracted energy is determined individually per frequency band and depends on the band's maximum diffraction angle. Figure 63 exemplifies the concept and shows a virtual scene from the listener's perspective (Figure 63a), the constructed diffraction/edge map (Figure 63c) and the by β diffracted ray from the listener to a sound source (Figure 63b). The edge map in Figure 63c is constructed by using the scene's depth buffer and an image based edge detection algorithm, in which for each edge additional rays are cast into the scene to perform the diffraction simulation.

Demo Reflection.

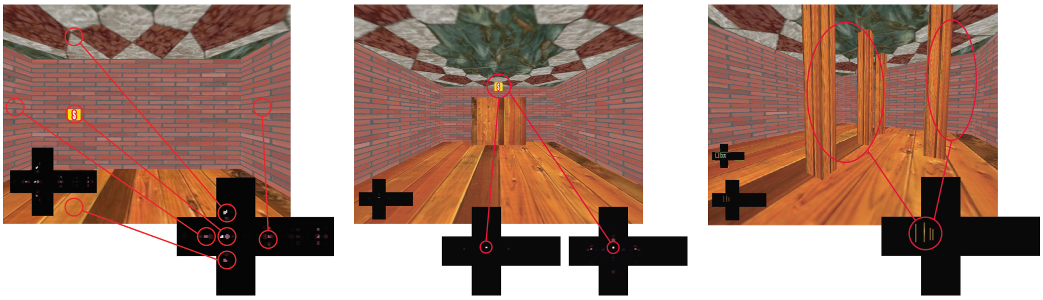
Demo Refraction.

Demo Diffraction.

Interference describes the superposition of two or more sound waves and the subsequent changes in amplitude. By employing a ray-based acoustic simulation, interference effects can only roughly be approximated using the ray's length and its center wavelength λ_{center_j} (Table 7). Although such a system only allows coarse approximations, the results can clearly enhance the simulation and provide a more realistic virtual auditory environment. More and finer subdivided frequency bands will, nevertheless, improve both techniques and also allow a finer modeling of frequency-dependent material definitions.

8.4.4 *Results and Experiments*

Several experiments and tests have been performed to evaluate the quality and the efficiency of this approach. A more detailed discussion of examples can be found in Section 9.7, which discusses and compares several room acoustics and virtual HRIR simulations. KEMAR is a dummy head model that is generally used in acoustics to measure head-related impulse responses. The results of this simulation can be found in Section 9.7.2. All evaluations have been performed on the same computer that was used for assessing the efficiency of the 3D waveguide mesh implementation. Table 9 shows the evaluation results with the number of frames per second (fps) for five different environments and five cubemap sizes, as well as with and without auralization applied.



(a) Reflections (Direct, 1st and 2nd). (b) Different Material Settings. (c) Diffraction around Pillars.

Figure 66: Modeling of Sound Wave Propagation Effects.

In acoustics, the term *realtime* is defined as an update rate of 10Hz or more (Funkhouser et al., 1999a), which the system currently achieves for meshes of up to 15,000 polygons. A cubemap with a size of 32×32 thereby allows for more than 6,000 possible ray directions. With the simulation running for all 10 frequency bands, this results in over 60k *convolutions* per simulation step, refer to Section 8.2.2. If a direction's acoustic energy is below a certain threshold, it does not contribute to the final auralization, and therefore, the real number of convolutions is vastly reduced. The acoustic quality of smaller cubemaps is comparable with higher resolutions, even though, some details are only audible using larger cubemap sizes, especially diffraction effects. The decomposition of sound data and publicly available HRIR filters was performed using windowed sinc filters with a length of 512 samples (Gardner and Martin, 2000). As expected, the performance decreases with the size of the scenario, as more triangles have to be checked for intersection. An update frequency of at least 10Hz is reached easily for smaller scenarios, but fails for the large hall. A better acceleration structure with a non-uniform subdivision, such as kD trees, might solve this problem.

Original Sound.

Lowpass Filtered.

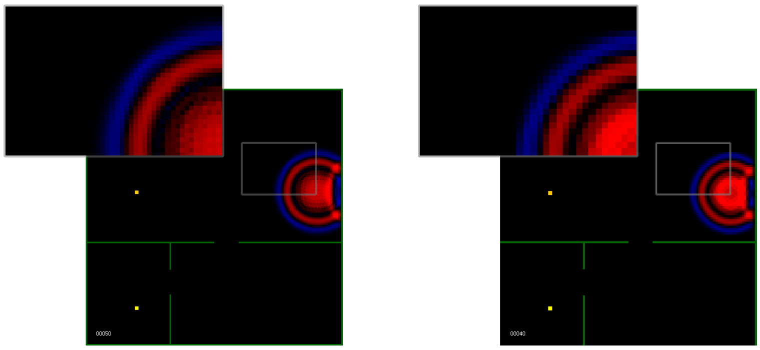
CC Simulation.

BCC Simulation.

Figure 66 displays three visualizations of sound wave propagation effects from the ray acoustics system. Here Figure 66a and Figure 66b display direct, 1^{st} and 2^{nd} order reflections with different material settings applied, whereas Figure 66c shows diffraction effects around several wooden pillars. The simulation results are visualized in the unfolded cubemaps and display the frequency range of 22Hz - 320Hz (Figure 66c). The red/brown shifting in color in all cubemaps denotes a stronger transmission/diffraction in the lower frequency end.

8.5 ANALYSIS AND DISCUSSION OF THE RESULTS

The goal of this research was not to develop a new and more efficient library for 3D sound rendering and simulation, but to discuss these simulations in the context of the special requirements for an auralization of 3D virtual auditory environments. As these environments entirely rely on sound and acoustics to convey abstract data and information, the techniques employed have to fulfill certain requirements: This is on one hand a more accurate and efficient 3D sound rendering and simulation of room acoustics, but also a non-realistic design of the auditory environment and an exaggeration of certain sound propagation effects. As currently available sound rendering APIs only emulate real sound wave propagation to a certain degree, this chapter examined the possibilities for a more efficient and especially for a higher quality 3D sound rendering and room acoustics simulation. Candidates were found in the realm of 3D computer graphics through the exploitation of programmable graphics hardware. The results from Section 8.2, Section 8.3 and Section 8.4 clearly show the advantages of a graphics-based implementation. Besides an improved computational efficiency, also the quality of the examples – see here also Section 9.7 – are very promising and strongly encourage a further research in this direction. Using the results and example implementations in this chapter, it was shown that even the highest demands of 3D virtual auditory environments regarding sound rendering and simulation can be fulfilled.



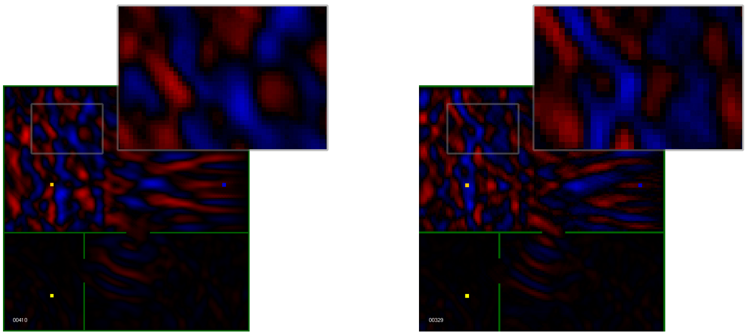
(a) Rectilinear Grid Frame 50. (b) BCC Grid Frame 40.

Figure 67: 3D Waveguide Meshes – Wavefront at $t = 50$.

The sound examples on the left allow to listen to several results of a simulation using the Cartesian and the BCC lattice. The examples provide the original sound and a low-pass filtered version, as well as the convolved results of both lattices. The implementation of the 3D waveguide technique in Section 8.3 was realized using two different mesh topologies. Figure 67 and Figure 68 visually compare the differences of both lattices using the example from Figure 61. The visualization of the BCC lattice is not aligned with an axis of propagation, but displayed as a rectilinear grid and sliced along the z axis. The visualizations clearly show a very similar wave propagation at the beginning, which, however, diverges as the animation continues. This is due to the differences in the mesh topologies, as well as the wave propagation itself. The BCC lattice has a smaller dispersion error and propagates the *pressure* along 4 instead of just 3 axes. This can be very well seen in Figure 67, which shows a much *smoother* wave front for the BCC lattice. The coarser resolution is due to the optimal sampling scheme, which requires only 70% of the sampling points to represent the same information, refer Section 8.3.2. The simulation runs approximately $\sqrt{2}$ faster than the implementation of the rectilinear grid.

8.5.1 *Combining Ray- and Wave-based Techniques*

The discussed ray- and wave-based approaches have their respective advantages and drawbacks and are both only applicable to certain parts of the audible frequency range. A combination of the two techniques would allow each method to perform at its peak efficiency and both simulations to complement each other. Depending on the rooms size and the objects therein, a certain threshold (overlapping frequency part) needs to be defined where both techniques are still applicable. The propagation of lower frequencies is simulated using 3D waveguide meshes, while the middle and higher frequencies are approximated using ray acoustics and the here described ray tracing techniques. Alternatively, both techniques can also be combined in a different way. The waveguide technique would have to be changed towards a boundary element method (BEM) and



(a) Rectilinear Grid Frame 400.

(b) BCC Grid Frame 325.

Figure 68: 3D Waveguide Meshes – Wavefront at $t = 400$.

would now only be applied in the direct vicinity of objects, sound sources and listeners. This would allow a disposal of all waveguide nodes positioned in free space and those which are far away from any object or sound source. At the same time, a decrease of the internodal sampling distance can be performed to achieve a higher sampling frequency for the remaining waveguide meshes. Ray tracing techniques can be used to connect the individual meshes and to transmit and propagate the acoustic energy over larger distances. Ideal for a realization would be an SLI[2]-based graphics system, in which each graphics board is assigned one of the simulation techniques and later both results are combined into a single impulse response. In order to be employed for a continuous auralization, the update frequencies of both techniques have to be aligned.

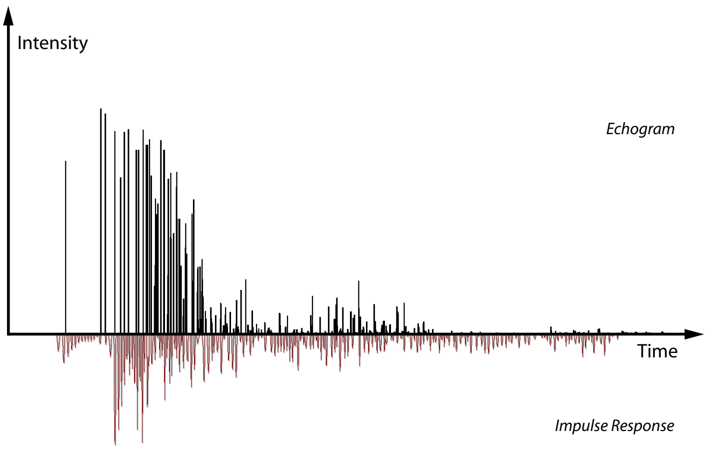


Figure 69: Comparison of Ray- and Wave-based Acoustics Simulations.

Figure 69 shows a direct comparison of both implementations. It displays an echogram (top) of the ray acoustics system and a measured room impulse response (bottom) using 3D waveguide meshes. The room depicted is the same as displayed in Figure 67, and was initialized in both simulations with the same parameters. Even though both techniques are designed for different frequency ranges, the major features of the frequency responses are clearly visible in both results. As this additionally proves the viability of both approaches, it also demonstrates the great potential for a combination of both techniques.

2 Scalable Link Interface refer http://developer.nvidia.com

9 CASE STUDIES

I'm in favor of an art that does something other than just sit on its ass in a museum. — *Claes Oldenburg*

SOUND and music are the two primary ingredients for the design of interactive 3D auditory environments. Depending on the actual sounds used, the application can vary from an interactive audiogame to a guiding system for the visually impaired. Sounds and auditory display systems are often utilized by artists and the performing arts to convey abstract ideas and meanings, thereby creating an interactive form of auditory art, different to any conventional exhibition and museal experience.

The last four chapters examined the topic of 3D VIRTUAL AUDITORY ENVIRONMENTS in great detail, thereby concentrating on an auditory interface design, intuitive task-related sonification and interaction techniques, issues for an efficient authoring and design, as well as on 3D sound rendering and simulation techniques on a very technical level. The following sections present several applications and case studies that implement many of the previously discussed approaches and emphasize on the respective ideas. Some of the examples are new, while others were introduced and outlined in earlier chapters. Opposite to the last chapters, the following sections allow a chapter-crossing perspective, and a presentation of the most significant results from various for an INTERACTION WITH SOUND.

9.1 OVERVIEW AND METHODOLOGY

The chapter is divided into several sections, with each focussing on a specific area of application. The majority of examples and applications presented is examined and evaluated using a short user analysis. Additionally, several audio and video examples are provided to demonstrate the implementations and the results achieved.

The following list provides a short overview of the examples and case studies that are discussed in this chapter:

- Sonification techniques for 2D and 3D data sets (Section 9.2).
- Sonification and interaction techniques for user navigation and orientation tasks in 3D virtual auditory environments (Section 9.3).
- An evaluation of several audio-only computer games (Section 9.4).
- A study on auditory perception using bone-conducting headphones, as well as an evaluation of an augmented audio reality system along two main examples (Section 9.5).
- An introduction to *Interactive Audiobooks*, along a discussion of several example story implementations (Section 9.6).
- An analysis of the implemented sound rendering and simulation techniques with applications to virtual room acoustics and HRIR simulations (Section 9.7).

The discussion of all examples and applications within one single chapter possesses several advantages, and allows an easy comparison of techniques, implementations, and the presentation of content. A few examples in this chapter also share a similar setting. The story of *The hidden Secret* is implemented as 3D audio-only computer game

(Section 9.4), as an augmented audio reality scenario (Section 9.5), as well as in the form of an interactive audiobook (Section 9.6). This not only allows to compare various technical realizations, but to also assess the immersion and the perception of the story using different presentations.

An often employed method to assess the quality and the design of a new system is the performance of usability tests, in which a selection of participants has to complete a certain pre-defined task. The performance of the participants is thereby observed and recorded, as well as questionnaires are handed out that have to be completed before and after an evaluation of the interface or application. Additional information can be derived from an event-logging procedure within the application, which allows a one-to-one reconstruction of a specific user's performance. All results, measurements and questionnaires are interpreted and evaluated by professionals, who are able to assess the user's performance, and therefrom conclude on the efficiency of the tested systems. The two scales that are used in the questionnaires in this chapter are yes/no and a weighting function ranging from 1 (poor/low) to 5 (high/great). Several examples presented in this chapter were examined within a multi-user evaluation, see the *General Questionnaire* on the left.

Questionnaire "General and demographic Information".

The requirements for an evaluation of auditory displays are similar to those of visually-centered user interfaces. Added specifics due to an auditory presentation are a silent environment, high quality headphones, as well as the use of carefully designed high quality audio samples. All evaluations were performed in a way that the setting as well as the time allotted to complete each task was the same for all participants. Prior to each evaluation, a demonstration was provided to familiarize everyone with the controls and the user interface of the example application. The results of the evaluations are displayed as frequencies using either percentages, or a mean value that ranges between 1 (poor/low) and 5 (high/great). To assess the scattering of the data, as well as to provide a confidence interval of the mean value, the standard deviation, as well as the standard error of mean are provided for all measurements. However, the discussions of the user evaluations are summarized and discuss the most significant results only. More details can be found in Appendix B, which provides an additional analysis for each evaluation, as well as references the questionnaires and some of the evaluations. The examples, as well as the SPSS data files can be found on the DVD and in Section C.4 (SPSS Inc., 2008).

9.2 2D/3D DATA AND IMAGE SONIFICATION

Although the focus of this research is the development of 3D virtual auditory environments, several of the therein employed 3D scene sonification and interaction techniques are directly related and based on methods for data, image and volume sonification (Stockmann, 2008; Stockmann et al., 2008). These methods have been introduced and discussed in Section 5.3. During this discussion, several techniques were advanced and further developed to improve the perception and to allow a more intuitive data sonification. The goal of this section is to assess the applicability of these techniques for an auditory display of stock market data and 2D shapes, but also for a sonification of 3D objects and 3D data volumes. The developed techniques were examined using a user evaluation, in which participants had to fulfill certain tasks. Four hypotheses have been postulated in advance to focus on and examine specific parts in greater detail:

- Sonification techniques are sufficient to acoustically display simple 2D/3D data sets
- Some techniques (sound spatialization, rhythm, melodies) improve the perception and allow a finer stream segregation

- A combined audio/visual examination of data sets is more efficient and thorough than a clean graphics-based data visualization
- Spatial interaction techniques thereby greatly improve the understanding of the data set and its topology

The questionnaire to analyze and proove/disproove these hypotheses was grouped into five sections, in which each section was further divided into individual tasks:

Questionnaire "Data and Volume Sonification".

- Classification of melody and rhythm
- Stock data sonification
- 2D Shape sonification
- 3D Object sonification
- 3D Volumetric data sonification

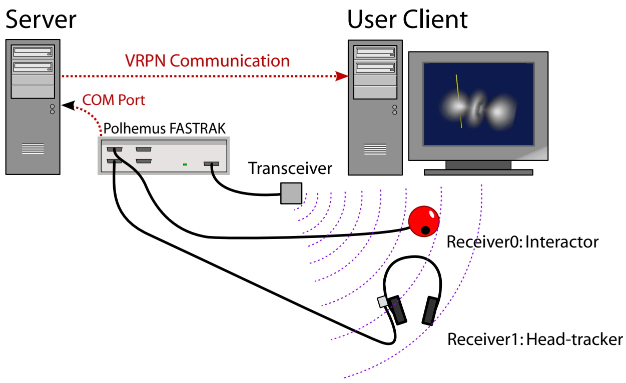
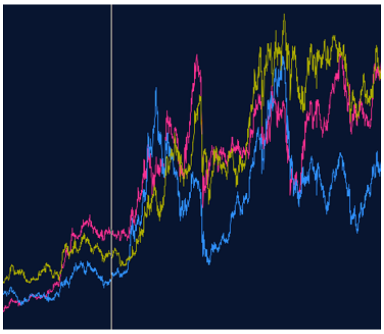


Figure 70: User Evaluation Setup.

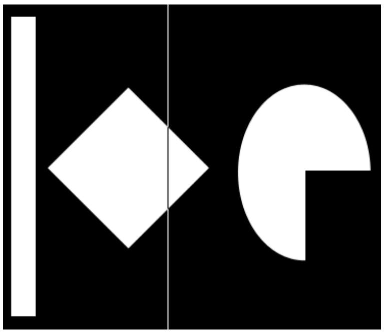
The tasks for this evaluation are discussed and explained in the following two sections. The setup for this evaluation can be seen in Figure 70. Overall, the following equipment was employed in the evaluation:

- Three desktop computer systems:
  - One computer for the evaluation of the 2D sonification techniques
  - One computer for the evaluation of the 3D sonification techniques
  - One control computer for the tracking system
- Two regular HiFi headphone systems
- One tracking system (Polhemus FASTRAK plus the 3Ball sensor)

In this setting, one PC was employed as the host for the FASTRAK tracking system, which used the local network to communicate the tracking device signals using the VRPN library (Taylor II et al., 2001, 2008). Two sensors were attached to the tracking system: One that allowed a measurement of head-rotations, ie. head-tracking, while another was employed as 3D interaction device to rotate and interact with the 3D objects and 3D volumetric data sets (Polhemus 3Ball). The interaction with the 1D and 2D sonification techniques were performed using a regular computer mouse. Standard HiFi headphones have been used as auditory display in all experiments.



(a) Stock Market Data Sonification.



(b) 2D Shapes Sonification.

Figure 71: 1D/2D Data Sonification.

9.2.1 *Stock Market Data and 2D Shapes*

For several data sets and applications, sound and acoustics provide a more intuitive way to represent and understand the inherent information. An example that was employed in previous discussions is the sonification of stock market data, see also Figure 71a and the demonstration on the left. While the sonification of a 1D data set is relatively easy to implement using a direct auralization approach, a parallel sonification of several 1D data sets as well as the sonification of higher dimensional data, such as 2D shapes and images, is more difficult.

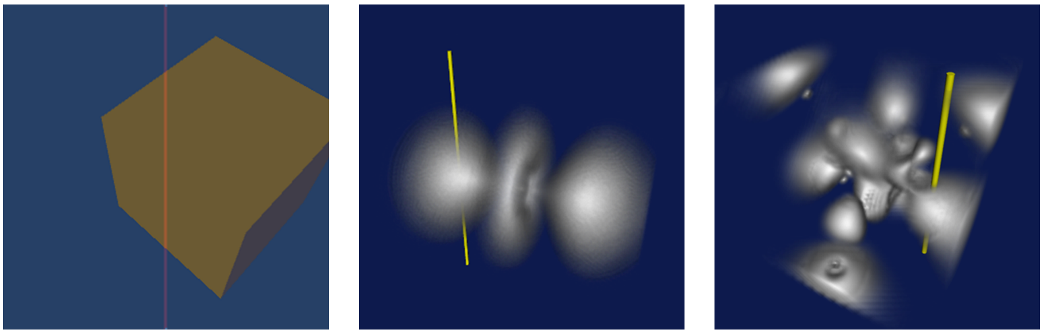
Stock Data Sonification.

The first assignment in this evaluation presented four different melodies. The task of the participants was to describe and classify these melodies and to rank their level of *ascendence*, ie. to describe the melody's direction (up/down). The gathered data was used to determine of how one perceives, appreciates and interprets melodic rhythms. This information was employed in the analysis of the other data sonification tasks.

The second and the third assignment in this evaluation were the sonification of stock market data, as well as an acoustic representation of several 2D shapes. Two examples that were used can be seen in Figure 71. Figure 71a shows an example of the stock market data, while Figure 71b shows several 2D shapes that had to be identified acoustically. The sonification of the stock market data was performed using three different techniques. The first technique employed a simple auralization of three stocks, in which each curve was represented by a different instrument. The height of the curve was mapped to timbre and a computer mouse was employed to *scroll* through the data. The task of the participants was to identify the number of curves (ie. three), as well as to sketch their individual characteristics (ie. shape). The second sonification employed the same technique, but additionally used sound spatialization to disperse the stock sonifications around the listener's head. The third technique further extended this approach and used sound spatialization, as well as a rhythmic sequencing. The number of curves in the last experiment was raised to four, and despite these difficulties, the overall perception using this method was best. The performance of the participants increased gradually and with each technique.

2D Shape Sonification.

The third assignment was concerned with the identification of several 2D shapes using the acoustic scanline technique that was introduced in Section 5.3.1. Figure 71b shows an example of the 2D shapes to be identified, which almost all participants answered correctly, see also the demonstration on the left. Overall, three tasks had to be completed, in which the shape and the number of several different 2D objects had to be named.



(a) Cube. (b) Hydrogen Atom. (c) Protein Molecule.

Figure 72: Sonification of 3D Objects and Data Volumes.

9.2.2 *3D Objects and Data Volumes*

Section 5.3.1 discussed several possibilities for a sonification of 3D objects and volumetric data sets, and advanced the 2D scanline sonification technique for an application on 3D data. As an acoustic display to perform this task is very complex, it relies on adequate sonification and interaction techniques, as well as requires a well designed user interface. The setup for this experiment is depicted in Figure 70, in which the interaction device (sphere) was used to rotate and orient the data, while the object was scanned through head movements (head-tracking).

3D Object Sonification.

3D Volumetric Data Sonification.

The forth and the fifth assignment in this evaluation were the sonification of 3D objects and 3D volumetric data sets. Examples can be seen in Figure 72. Figure 72a shows a 3D cube along the center sonification scanline, while Figure 72b and Figure 72c display two 3D volume data sets with the interactive chimes in their center, see also both demonstrations on the right. The sonification of 3D objects is a direct adaptation of the sonification technique used for 2D shapes. The participants task was to examine three different 3D object using the sonification technique described, as well as to examine three volumetric data sets using the interactive chimes. In the volume data sonification, the participants were asked in the first task to identify the data set they have heard using a visual representation, and in a second task to also sketch the data sets characteristics (ie. its density distribution). A final experiment evaluated an audio/visual examination of a 3D data set, from which one could only derive additional information by using the acoustics cues. The performance was generally good and also exceeded the anticipations. However, two of the 3D objects had a very similar resemblance and were sometimes confused with each other, refer to Section B.1.

9.2.3 *Discussion*

In this evaluation, a total number of 15 participants (14 male, 1 female) were involved, of which two had a vision impairment and one a slight hearing deficiency. The range in age was between 20 and 39, and none of the participants had major knowledge and experiences with auditory displays or audio-only computer games, see also Section B.1.

The results in Table 10 show a weighted analysis of the participants performances for the tasks described above. A mean value of 3 thereby expresses an average performance, while 2 discloses problems of several participants, and a value of 4 displays a good to very good performance of the majority of participants. The results for the stock

Technique/Task	Performance (Mean) (poor/low (1) – great/high (5))	Std. Deviation	Std. Error of Mean
Stock Sonification (Timbre)	3.67	1.11	0.25
Stock Sonification (Timbre + 3D)	3.80	1.32	0.34
Stock Sonification (Timbre + 3D + Rhythm)	3.87	1.19	0.31
2D Shapes	3.95	0.73	0.19
3D Objects	2.55	1.13	0.29
3D Data Volumes	3.38	1.04	0.27

Table 10: Data and Volume Sonification Results.

data sonifications were assessed through the participants performance (ie. the number of identified stocks and the quality of stock curves drawn). Additionally, each user's performance was assessed by two observers. The evaluation of the stock market data sonifications clearly show a better performance through an added spatialization, especially when combined with an additional rhythmic sequencing. Although this has been anticipated, it was assumed that the performance of the last stock sonification, due to the increased number of parallel stocks, would only perform as equal, but not better than the others. Interesting to note is also that the quality of the shape drawings increased, and performed best at the last sonification technique. The sonification of 2D shapes and 3D objects performed well as well, although some shapes/objects had a similar auditory resemblance and were sometimes misinterpreted (eg. sphere/cylinder). The sonification of volumetric data sets proved that even more difficult volumes can here be identified correctly. An added spatial sonification and exploration allows thereby a good understanding of the data's inherent topology. A combined data sonification/visualization achieved overall the best performance. Adding to the results of Table 10, $\frac{2}{3}$ of the participants reported that a (multivariate) audio/visual sonification of volume data is more efficient and that they also gained further knowledge through the added sonification. Also to note is that two participants scored 100 percent in all tests, meaning that all data and the information therein have been identified correctly. Overall, the results clearly show that not only an acoustically enhanced, but also an audio-only sonification of 2D and 3D data sets is possible by untrained ears. The conclusions to be drawn are that all sonification techniques performed even better than anticipated, and that all initial hypotheses could be confirmed.

9.3 SONIFICATION AND INTERACTION WITH 3D ENVIRONMENTS

Using the results of the last section, as well as the research from Section 5.3.2 and Section 5.4, several 3D scene sonification and interaction techniques were developed for an intuitive display and interaction with 3D virtual auditory environments. This section discusses and evaluates these techniques and explores their applicability to specific problems. For an evaluation of these techniques, two user studies have been performed (Röber and Masuch, 2006). The goal of these evaluations is to assess the functionality and applicability of the devised techniques, and to examine the performance of users that are exploring 3D virtual auditory environments. Postulated hypotheses for this evaluation are:

- An orientation, navigation and exploration in 3D virtual auditory environments is easily possible with adequate 3D scene sonification and interaction techniques
- A selective listening (auditory lens) allows a better perception and understanding of the environment
- The soundpipes approach improves the orientation and navigation within an auditory environment
- Head-tracking and sound spatialization improve perception and navigation
- Speech analysis and synthesis are both only partially applicable
- The interaction with a 3D ring based menu system can be performed trough
  - Earcons and/or speech for information sonification
  - 3D Gestures and standard (gamepad) interactions

The 3D Sound Stage.

Although sonification and interaction were discussed separately in Chapter 5, they are evaluated together in this section. In these evaluations, the participants deploy dedicated techniques of *Interaction* to input information into the virtual environment, which on its behalf employs methods of *Scene Sonification* to convey and display information to the user. The techniques examined in this section have varying applications and are partially based on different approaches.

The evaluation of most examples is conducted using a *virtual sound stage*, as it is depicted in Figure 73, see also the demonstration on the right. This basic setting allows an easy examination of the developed techniques and is also very easy to maintain and adjust (Miede and Futterlieb, 2005). The stage, as it is shown here, consists of two rooms with varying sizes and a different number of 3D sound sources (cylinders). The setup of this evaluation is similar to Figure 70, and based on a FASTRAK system to perform a user head-tracking and to permit 3D spatial interaction techniques. The overall requirements for this evaluation are:

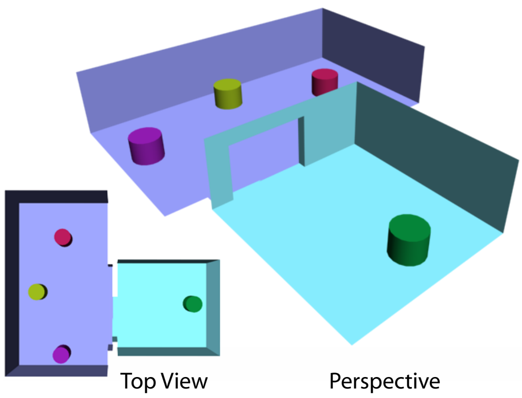
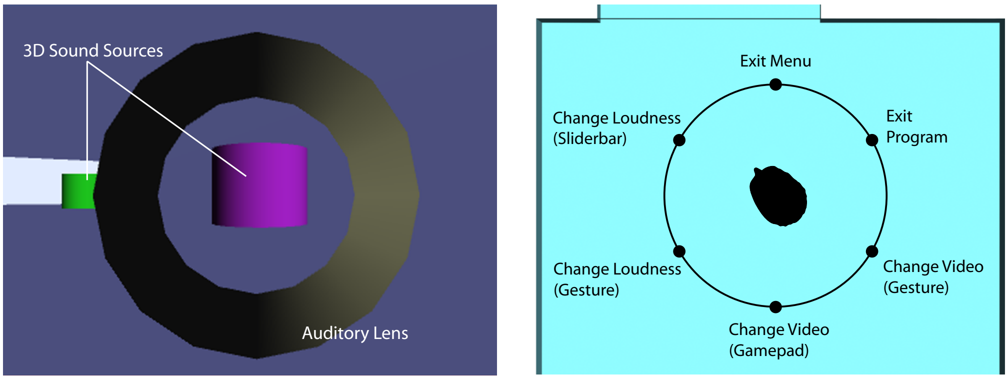


Figure 73: The Sound Stage.

- Two desktop computer systems:
  - One computer for the evaluation of techniques
  - One control computer for the tracking system
- One HiFi headphone system with a microphone for speech input
- One tracking system (Polhemus FASTRAK plus the Stylus sensor)
- One gamepad for regular interactions



(a) Variable Auditory Lens. (b) Menu Interaction (Setup).

Figure 74: 3D Scene and Menu Interaction.

A regular gamepad is used for the control of the environment and to input various information. The additional microphone is employed in one of the experiments for a speech-based interaction, while a regular HiFi headphone system is used as auditory display in all experiments.

The following three sections focus on and evaluate different aspects of 3D scene sonification and interaction. Section 9.3.1 discusses the more general techniques, which were analyzed using a detailed user evaluation. Section 9.3.2 discusses the soundpipes approach to improve the user's navigation and orientation, while the Section 9.3.3 examines the possibilities of a speech-based interface to control 3D auditory environments.

9.3.1 *3D Scene Sonification and Interaction*

Questionnaire "3D Scene Sonification".

A first evaluation was concerned with an analysis of the primary 3D scene sonification and interaction techniques, see also the questionnaire on the left:

- 3D Scene – navigation and orientation
- Selective 3D scene sonification – examination of the auditory lens
- Navigation and pathfinding through a complex 3D auditory environment
- Speech-based 3D scene interaction and sonification
- Interaction with a 3D auditory ring menu system

The evaluation was performed using 14 participants (13 male, 1 female) with an age ranging between 20 and 39. Two of the participants had a visual impairment, while one had a slight hearing insufficiency. The majority of the participants ($\geqslant 70\%$) had a high familiarity with 3D interaction techniques, with a few users also being experienced with auditory displays and audio-only computer games ($\leqslant 20\%$). The results of this evaluation are summarized in Table 11.

Two examples from this evaluation are depicted in Figure 74 and Figure 75. Figure 74a displays a visualization of the earlier discussed auditory lens, while Figure 74b shows a schematic of the ring-based auditory menu system. The setup in this evaluation is exactly as described above and employs a tracking system to measure user orientation and to allow 3D interaction. Additionally, a gamepad was employed for a variety of tasks, to control the auditory lens and the auditory menu system, as well as for navigation and orientation within the 3D scene.

The first experiment was based on the layout of the sound stage as it is depicted in Figure 73. The participants task was to explore a 3D auditory environment with the techniques provided to find and activate four different 3D sound sources (cylinders in Figure 73). Additionally, an overview of the perceived scene topology with the location of all sound sources found had to be drawn. The interaction with the environment was based on 3D head-tracking and a navigation/orientation using a regular gamepad. The performance of all participants was good, although two candidates (which did not complete the evaluation) had large difficulties in 3D sound perception, possibly due to the generalized HRTFs used. Also interesting to note is that several participants, but especially one, simply walked in this experiment from source to source in a matter of seconds, only utilizing the cues of sound spatialization and 3D head-tracking. These candidates constantly rotated and turned their heads, which allowed them a more efficient use of the time differences encoded in the binaural signal to localize the 3D sources.

The auditory Lens.

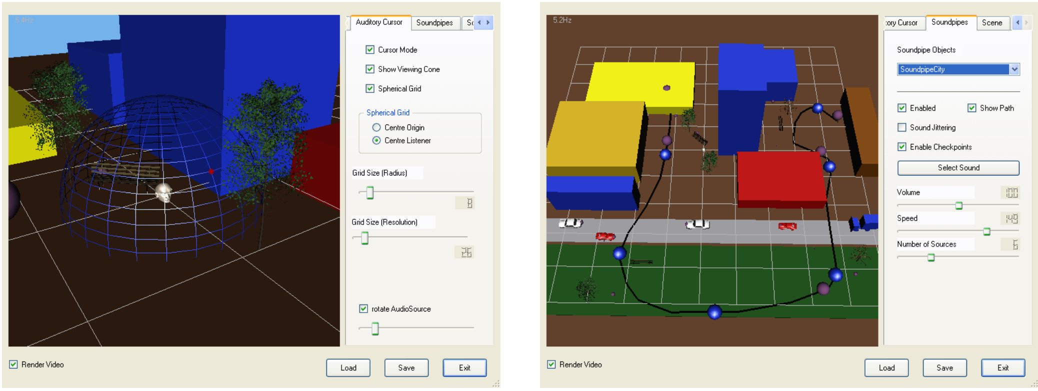
The second task included the exploration of differently complex scenes with and without the assistance of the auditory lens system, refer to Figure 74a. Out of 14 participants, 10 declared that both, orientation and navigation were easier to perform with the auditory lens system, while 4 stated that the level of difficulty was equal. The parameters to vary the lens' depth, radius and source selection were only used occasionally ($\leqslant 5$), which is probably a result of an unfamiliarity with the interface itself. The majority of participants, however, reported that a differentiation of source types in ambient and object sounds is helpful ($\geqslant 9$) and permits a more intuitive perception of the auditory scene, especially within more complex environments.

Navigation through a complex auditory Environment.

The third experiment included the navigation through a complex environment with 12 sound sources which were playing constantly. The task in this experiment was to reach a distant signal in the fastest way, but without colliding with other sound objects. All but one participant performed this task very good, and reached the goal without, or only a very few collisions. 12 participants reported that the head-tracking greatly improved their performance and allowed a detection of sound sources more easily.

Technique/Task	Performance (Mean) (poor/low (1) – great/high (5))	Std. Deviation	Std. Error of Mean
Orientation in 3D Scene	3.87	0.46	0.12
Orientation/Navigation with Hear-Frustum	3.68	0.82	0.22
Head-tracking (Function)	3.93	0.75	0.20
Speech Control	2.87	1.14	0.30
Speech Synthesis	3.36	1.33	0.35
Gamepad Control	4.64	0.49	0.13
3D Source Localization	4.53	0.57	0.15
Menu Interaction	3.64	1.15	0.30
Gesture Interaction	3.46	1.02	0.27
Earcon Design	3.64	0.93	0.25

Table 11: 3D Scene Sonification and Interaction Results.



(a) Auditory Cursor.

(b) Soundpipes Path Sonification.

Figure 75: 3D Scene Sonification and Interaction.

Ring-based Menu System.

Another experiment was concerned with the exploration of a virtual auditory ring-based menu system, as it is depicted in Figure 74b. The task for each participant was to activate the menu and to interact with it to gather an experience and an understanding of its individual functions. The available techniques for interaction were based on the regular head-tracking, a magic wand device (Polhemus Stylus) and a gamepad. The individual menu items were sonified through either speech or a descriptive auditory hearcon. Using the scale of Table 11, all items were easy to localize (4.0) and the general interaction with the menu was reported as good (3.64). The performance of spatial interactions using the stylus were intuitive (3.57), although the majority preferred an interaction using a classic gamepad (4.64), possibly due to a greater familiarity with such an interface.

Table 11 summarizes the results of this evaluation. The data values in Table 11 are a combination of the answers in the questionnaires and an assessment of the participants performance by two observers. Three results that require further attention are the *Orientation in the 3D Scene*, as well as the functionality of the employed *Head-tracking* and the high efficiency of *3D Source Localization*. These three findings not only underline the applicability, but show the use of a head-tracking system and 3D spatialized sound sources as an imperative requirement for an interaction with 3D virtual auditory environments. A fifth experiment, which was also part of this evaluation, examined a speech-based interface and is discussed in Section 9.3.3.

9.3.2 *Soundpipes Path Sonification*

An exploration of an unfamiliar auditory environment is a very complex and difficult task. Even more difficult are the tasks of navigation and wayfinding in such an environment. Auditory landmarks can support these tasks to a certain degree, but only if the landmarks are known, or if a map is available for additional referencing. To improve navigation and orientation, Section 5.3.2 introduced the soundpipes approach, a technique that assists the user in traveling through an auditory environment by providing an auditory pathway sonification.

Figure 75b shows an overview of an earlier evaluation environment that was used to implement and examine certain 3D scene sonification and interaction techniques. The system displayed was implemented using C ++ and uses Qt for the graphical user interface. OpenAL is employed for sound rendering and sound spatialization, and a

Polhemus FASTRAK and a regular gamepad are used for 3D scene interaction.Figure 75a shows here an evaluation of the implemented auditory cursor, while Figure 75b displays a visualization of the soundpipes approach, see also the demonstration on the right.

Soundpipes and auditory Cursor.

The displayed *city demo* in Figure 75b includes several buildings that are acoustically represented through a descriptive hearcon. An added soundpipe moves through the environment and along with it 6 moving sound sources. In an informal evaluation, several participants were asked to explore the applicability of this system by following the soundpipe and by moving from a certain start to another predefined end position (Röber and Masuch, 2006). As part of this evaluation, three tasks had to be performed:

- Follow the path from point A to B,
- Follow the path from point A and find the right exit C,
- Find the soundpipe and follow it in the right direction to B,

in which A, B and C were predefined positions along the soundpipe that highlighted objects of interest. Six users participated in an informal evaluation of this approach. None of the participants had any prior experiences with 3D auditory display systems or audiogames of any genre. However, the concept was understood quickly and everyone accomplished the tasks as required. The soundpipes approach was perceived as very helpful to navigate through 3D auditory environments. The integration of a 3D head-tracking thereby greatly improved the performance of all participants. Using this technique, it was easy to determine the position of the soundpipe and the direction of its movement. Some difficulties were introduced by the use of generalized HRTFs, which in certain cases impeded a fast source localization. However, due to the dynamic listening cues provided through the head-tracking system, nearly all of these situations could be resolved. In conclusion, the soundpipes approach is valid and can be easily employed also by unfamiliar users.

Figure 75a shows the auditory cursor within the same evaluation environment (Röber and Masuch, 2004b). An auditory cursor is an extension of a regular computer cursor and is based on a 3D pointing technique. Thereby several spheres are centered around the listener's head, on which the auditory cursor is placed and sonified using a 3D hearcon. The interaction is performed using an additional 3D sensor (Polhemus FASTRAK Stylus or 3Ball), which is used to select the direction of the cursor as well as to specify its depth, refer to Section 5.4. An auditory cursor can be employed in a number of 3D scene interaction tasks, such as for object and menu selection/interaction, but was, however, not examined in the form of a user evaluation.

9.3.3 *Speech-based Interaction*

Several of today's applications employ speech perception and synthesis, especially systems and assistance devices developed for the visually impaired. The control of 3D auditory environments using speech seems therefore to be a logical choice, but it might also hinder an efficient and effective interaction. A speech-based control is appropriate for situations in which speech is generally used, eg. for communication, or when a control of the system is otherwise not possible, eg. no hands free. The synthesis of speech in its current development is relative mature, well to use and often good to understand. Speech perception, however, still has several difficulties and limitations, as it requires a trained speaker and a low-noise environment, and still does not perform in a way that all words are sufficiently identified.

During this research, speech perception and synthesis were both implemented and evaluated in two example applications. The first one was the control and interaction with the above described sound stage using speech commands and the display of feedback information using speech synthesis (Miede and Futterlieb, 2005). For this task, the free available Microsoft Speech SDK was employed. Both tasks, speech synthesis and speech recognition, did not perform very well, refer to Table 11. The speech synthesized sounded very computerized and was in cases difficult to understand. Even more difficult was the speech-based interaction, in which movement and interaction commands were mapped to speech input. Although the speech recognition was implemented to classify similar sounding words as correct, eg. *walk* and *hawk*, the overall interaction was relatively poor, refer to Table 11. A second implementation employed speech perception for the play of the classic adventure game *"Day of the Tentacle"* (Lucas Arts, 1993; Malyszczyk and Mewes, 2005), see also Figure 76. The control of this game using speech worked relatively well, but required a time-consuming training of the speech perception software. Nevertheless, a speech-based control can efficiently be employed for the interaction with 3D virtual auditory environments, but a few guidelines have to be obeyed: Speech is best suited for communication purposes and should only be used if other forms of interaction are not available. However, to reduce occurring user annoyances, the perception system must either be trained well enough, or be adjusted in a way that also false detections are *interpreted*.

Speech-based Interface of the "Day of the Tentacle".

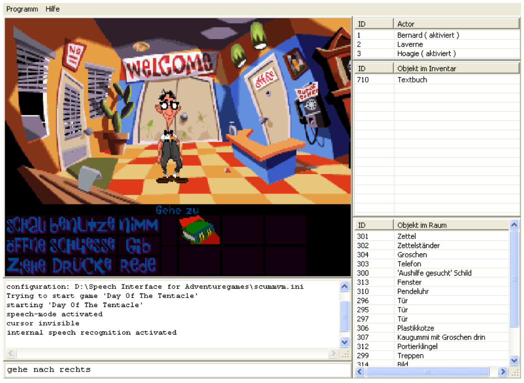


Figure 76: Speech-based Gameplay.

9.3.4 *Discussion*

This section discussed several 3D scene sonification and interaction techniques, of which the majority were examined using user evaluations. The results of this examination are very promising and underline the validity of the designed techniques. Two methods proven to be essential for an interaction with 3D virtual auditory environments are 3D head-tracking and sound spatialization. Although only generalized HRTFs could be employed for the spatialization of sound in these experiments, the majority of participants had no difficulties with sound localization and perception. However, two out of 36 participants were not able to localize virtual 3D sound sources, which further emphasizes the importance of the research of Chapter 8 and the discussion of personalized HRIR simulations in Section 9.7.2. The devised 3D scene sonification techniques, such as the auditory lens and cursor, as well as the soundpipes navigation approach, could be examined and have been confirmed to improve the perception and navigation of a 3D auditory scene. Techniques employing speech recognition were only found to be partially helpful, as deficiencies still exist in the accuracy of currently available systems.

9.4 AUDIO-ONLY COMPUTER GAMES

Audiogames have been introduced in Section 3.4 as a special form of auditory display system. These games exist in a large variety of genres and differ from their audio/visual counterparts in many aspects. The information perceived is generally less detailed and more difficult to interpret. However, this difficulty turns in a second view into one of the great advantages of audio-only computer games, as it allows – through a careful design – the creation of highly immersive auditory environments that are shaped by the player's own imagination. Unfortunately, the majority of audiogames are still developed for and played on the PC platform only, although several efforts exist to move this genre to portable devices and to use a more audio-centered game design (Röber and Masuch, 2005a; Huber et al., 2007).

The audio framework that was developed in Chapter 5 was designed with an implementation of audiogames already in mind. The key components for developing 3D interactive audiogames are:

- A 3D virtual auditory environment focussing on an audio-centered gameplay
- A non-realistic design of the auditory environment
- Intuitive sonification and 3D interaction techniques for exploration, navigation and user interaction
- Possibilities to implement a narrative concept that focuses on an auditory narration

To evaluate the framework, as well as the applicability of the previously discussed sonification and interaction techniques, three basic action games, as well as one auditory adventure game were devised and implemented:

- Mosquito,
- MatrixShot,
- AudioFrogger, and
- The hidden Secret.

All four games are implemented using the audio framework discussed and employ 3D head-tracking, as well as spatial interaction techniques. Figure 77 shows four control screenshots of the implemented audiogames. These visual representations are used to test and analyze the gameplay, while the games itself are played audio-only without any visual feedback. The following two sections discuss the implementation of these audiogames, as well as compare them with other audiogame examples that are reviewed regarding their approach, level of difficulty, acoustics, design, as well as their fun and novelty. The goal of this evaluation is to explore the potential of an audio-centered gameplay and to assess the applicability of the previously evaluated 3D scene sonification and interaction techniques towards an employment in audio-only computer games. The hypotheses for this evaluation are:

- An audio-centered gameplay is more enjoyable than an adaptation of a visual genre
- Spatial interactions and 3D head-tracking improve the perception and the playability of a 3D audiogame
- Efficient and high-quality 3D sound spatializations are required
- An audio-only gameplay is highly immersive
- Audiogames can be played and enjoyed by unexperienced and sighted users as well

The questionnaire employed in this evaluation was divided into two sections, a general classification of the game and its genre, as well as an assessment of the interaction and sonification techniques used, refer also to Section B.7. 13 users (12 male, 1 female) participated in this evaluation, of which two had a slight visual and one a slight hearing impairment. Two participants had no prior experiences with computer games at all, while three users were also familiar with an auditory gameplay. The participants played and evaluated the following six games:

Questionnaire "Audiogames".

- Mosquito (played 11 times)
- Audio Frogger (played 9 times)
- The hidden Secret (played 5 times)
- Der Tag wird zur Nacht (played 5 times) (Dannecker et al., 2003)
- Terraformer (played 3 times) (Pin Interactive, 2003)
- Shades of Doom (played 3 times) (GMA Games, 2001)

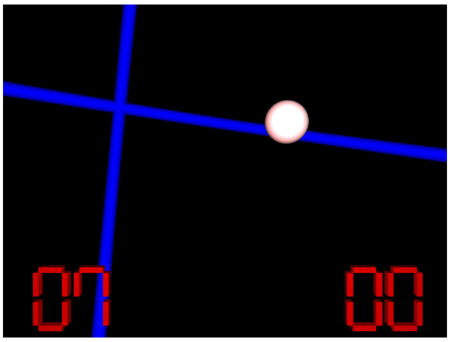
The results of this evaluation are discussed throughout the following two sections and are summarized in Table 12 and in Section B.3. A play of the regular audiogames requires a standard PC only, while the four games that are based on the audio framework demand additional hardware to perform the spatial interactions. The setup is therefore similar to Figure 73 and based on:

- Three desktop computer systems:
  - One computer for the evaluation of regular audiogames
  - One computer for the evaluation of the audiogames that utilize 3D head-tracking and spatial interaction
  - One control computer for the tracking system
- Two regular HiFi headphone systems
- One tracking system (Polhemus FASTRAK with the Stylus sensor)
- Two gamepads for regular interaction

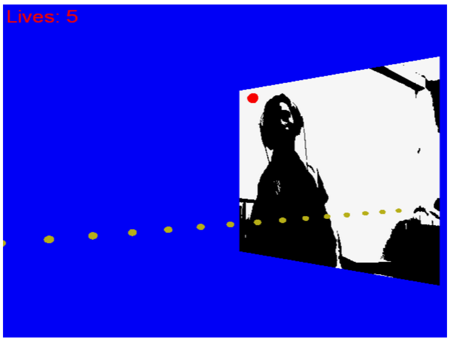
9.4.1 *Auditory Action Games*

Action games challenge the player in speed, reaction and situational analysis, and often employ tactical conflicts such as in first-person shooter games. The genre of auditory action games acoustically displays pieces of information, which the player has to interpret correctly and to which he has to react as quickly as possible (van Tol and Huiberts, 2006). Examples of this category are racing and shooter games, as well as certain arcade and sports games. The target audience of these games are in many cases the visually impaired only, which often results in game complexities that only reaches a fraction of that of conventional audio/visual computer games.

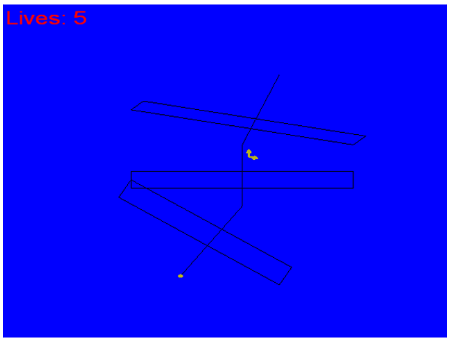
The majority of audiogames available are played on the PC platform and use the keyboard as main interaction device. The primary goal was therefore to devise a new development for action-centered audiogames that explicitly focusses on an auditory perception and gameplay. To exemplify this new approach, three small action games were developed, based on the audio framework discussed, and utilize techniques of 3D scene sonification, 3D head-tracking and spatial interaction.



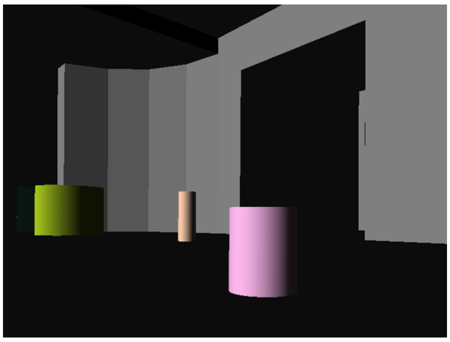
(a) Mosquito.



(b) MatrixShot.



(c) Audio Frogger.



(d) The hidden Secret.

Figure 77: Playing Audio-only Computer Games.

The audiogames that are evaluated in this analysis are:

- Audio Quake (Atkinson and Gucukoglu, 2008),
- Shades of Doom (GMA Games, 2001),
- Mosquito (Figure 77a),
- Matrix Shot (Figure 77b), and
- Audio Frogger (Figure 77c).

Audio Quake and *Shades of Doom* are both adaptations of classic audio/visual first-person shooter games (GMA Games, 2001; Atkinson and Gucukoglu, 2008). Both games are enriched by speech synthesis, but employ a similar gameplay as the original visual implementation and also use the same mapping and the same sounds. The interaction is performed using the computer keyboard, and is, especially for *Audio Quake*, very complex and difficult to understand.

Shades of Doom Demo.

AudioFrogger is an acoustic adaptation of the classic *Frogger* game developed in 1983 (Sierra On-Line, 1983). Similar to the original implementation, the player has to cross several streets without getting involved in a traffic accident. The difficulty increases with the number of lanes and the moving objects to keep track off, refer to Figure 77c. The avatar is controlled with a gamepad interface and 3D head-tracking is used to improve the perception of the auditory scene.

Video of "AudioFrogger".

Mosquito is played in real 3D space and with a possible 360° radius of interaction, see also Figure 77a. The scope of this game is to repel up to three attacking mosquitos and to

Video of "Mosquito".

Video of "Matrix Shot".

avoid being stung. The virtual mosquitos circle the player, who can focus on individual mosquitos using 3D head-tracking and kill them by using a virtual fly swatter. This is implemented by using either a gamepad interface, or an additional 3D sensor from the Polhemus system.

The last of the games is *Matrix Shot*, which was inspired by the first Matrix movie and the artistic movements of certain characters to evade enemy bullets (Wachowski and Wachowski, 1999). The goal of *Matrix Shot* is to detect the approach of virtual acoustic bullets and to evade them in the most acrobatic style possible. The bullets itself are moving 3D sound sources that approach the player from the front, while an additional USB camera that is mounted in front of the player is employed to determine whether the player was hit or not, refer to Figure 77b. Using head rotations, the player detects the direction and distance of the virtual bullets and can evade them, refer to the examples on the left.

The majority of existing audiogames aims on adaptation of a visually-based gameplay towards an auditory perception. As this is not only difficult to achieve, it also fails to take advantage of a real auditory gameplay. In contrast to *Audio Quake* and *Shades of Doom*, the last three examples discussed explicitly concentrate on an auditory gameplay and implement this using an orientation-based head-tracking, as well as by employing a real 360° wide interaction. The player has to estimate the location and the direction of movement for several 3D sound sources using natural listening cues. This facilitates a higher involvement and a better perception of the auditory scene (Röber and Masuch, 2004b, 2005b). The utilization of head-tracking, as well as the audio-centered gameplay of *Mosquito* and *Audio Frogger*, clearly enhances the fun and enjoyment, but also improves the display and the perception of game information. During the evaluation, the simple concept and the intuitive gameplay of *Mosquito* and *Audio Frogger* were praised by all participants. Other games, such as *Audio Quake*, were described by most players as too difficult to play. *Audio Quake* conveys the majority of information using a very bad speech synthesis that is combined with a very complex and difficult interface based on various keyboard shortcuts (Atkinson and Gucukoglu, 2008).

9.4.2 *Auditory Adventure Games*

Several sections in this research already discussed the suitability of auditory environments for a presentation of narrative content. Classic narrative computer games are associated with the adventure genre, in which a player follows a storyline and thereby unravels several mysteries and puzzles along the way. The adventure genre was introduced in the 1980s as classic text-based computer games and later transitioned to a graphics-based gameplay in the 1990s.

The evaluation of audio-based adventure games was arranged around the questions of user immersion, the perception of 3D space, as well as how to interact with an interactive auditory storyline. The games included in this evaluation are:

- Der Tag wird zur Nacht (Dannecker et al., 2003),
- Terraformers (Pin Interactive, 2003),
- Seusse Crane: Detective for Hire (Destiny Media, 1999), and
- The hidden Secret (Figure 77d).

Der Tag wird zu Nacht was developed as a student project in 2003 and is realized in Flash (Dannecker et al., 2003). The game is set in antique Pompeii during an eruption of the Vesuvius and the task of the player is to find an exit out of the city. For this, the player has to search the different rooms and navigate the virtual avatar to safety. Although the mission and the control of the game are easy, the task itself is very difficult to accomplish as not enough feedback sounds are provided and the player is simply lost in darkness.

Terraformers Demo.

Terraformer is a so called hybrid/acessible game that can be played by sighted and blind players together. The story is set on a distant planet and the task of the player is to regain control of the terraforming process, which has been shut down by revolting robots (Pin Interactive, 2003). Although the game is innovative and has received a lot of attention during its initial release, the gameplay itself is very difficult and not intuitive. The scene sonification is based on 3D sounds and the user can utilize a sonar-like technique for exploring the virtual 3D environment. Different to the majority of audiogames, *Terraformer* is quite complex and contains a large game world for play. However, a play and interaction solely using sound and acoustics is still difficult.

Seuss Crane Detective for Hire.

Seuss Crane is a classic audio-based adventure game and a technology demo by Destiny Media. Within the setting of the game, one plays a detective that has to unveil a murder mystery (Destiny Media, 1999). It is based on a radio play, in which the player chooses the locations to investigate and after a while, the player has to accuse someone for murder. The game has an interesting story and is played by professional voices, but the user interface is in the form of a simple hypertext-like menu from which one can choose the next location. A large drawback is the predefined sequence of the storyline, which does not permit deviations in order to receive points and to solve the game. Nevertheless, a play of *Seuss Crane* is enjoyable, and its realization classifies it as distant related work to the later introduced interactive audiobooks.

Video of "The hidden Secret".

The story of the last game – *The hidden Secret* – evolves around a tourist visiting the city of Magdeburg and his adventures in the city's main cathedral (Huber, 2004). Thereby one discovers and unveils several mysteries and puzzles to gain the long lost cathedral's treasure. The story is loosely constructed around several real myths and sagas found in an old book about the Cathedral of Magdeburg (Leinung and Stumvoll, 1904). The game is realized as a plain 3D auditory adventure and is implemented using the audio framework described earlier. It uses spatialized sound sources and 3D head-tracking to enhance the orientation and navigation in the virtual game world, refer to Figure 77d.

Although the implementation of *The hidden Secret* provides a rich acoustic atmosphere and an intuitive interaction, difficulties occurred within the determination of the player's position in the game environment and the estimation of distances and directions. A direct result was that several players got lost and did not find *back* into the game. With a concentrated listening and by using the 3D head-tracking technique, the game is easy to play and the strong immersion and involvement that was anticipated earlier, can be experienced. Overall, the audiogame genre is very well suited for presenting narrative content and for a design of adventure-based audiogames.

9.4.3 *Rethinking Audiogames*

Audiogames represent a relatively young genre compared to other computer games, but have received a high level of attention over the recent years and are continuing to grow in terms of quality, complexity and variety. So far, the majority of audiogames are still played in front of a regular PC and often remain auditory adaptations of successful audio/visual computer games. A rethinking of audiogames with a stronger focus on an auditory design and gameplay will hopefully advance the genre to the next level. This section further explores this development and provides guidelines for the authoring and design of audiogames, as well as examines the applicability of certain auditory

Audiogame	Mission	Difficulty	Acoustics	Design	Fun	Novelty
	Rating from (poor/low (1) – great/high (5))					
Mosquito	4.9	3.6	3.7	3.7	4.2	4.3
Audio Frogger	4.2	3.3	3.6	3.6	3.5	3.7
The hidden Secret	4.0	3.1	4.2	3.8	4.0	4.3
Tag wird zur Nacht	4.8	2.7	3.0	3.6	3.5	3.8
Terraformer	2.7	2.7	2.9	2.4	2.4	2.8
Shades of Doom	4.7	3.4	3.7	4.0	3.1	3.2

Table 12: Audiogames Evaluation Results.

display and 3D interaction techniques. Although this section concludes the discussion of audiogames within this research, the following sections continue to investigate their applicability towards the design of a mobile and location-aware gameplay (Section 9.5), as well as further discuss the narrative benefits of an auditory presentation (Section 9.6).

The design of audiogames requires special attention in order to develop an application that is fun to interact with, but which also provides sufficient information for the gameplay itself. The most important objective in designing audiogames is to immerse the player in a high quality virtual auditory world, and to utilize techniques that foster and enhance this experience. The design of the user interface and its integration into the game requires thereby special attention as well. The access to the menu, as well as the alteration of parameters has to be performed using the same techniques that are employed for playing the game. A difficulty that often occurs is the estimation of distances and the mapping of sounds to specific events. Certain methods for interaction and 3D scene sonification, such as head-tracking, spatial interaction and the use of a radar/sonar technique, have proven to be helpful (Röber and Masuch, 2004b, 2005b). Important is also to not clutter the auditory display with too much information, but rather to design it in a way that keeps an adequate balance between aesthetics and function. The quality of the sounds and music used is of high importance as well, as a poor sound design can easily ruin an otherwise well designed game.

Table 12 displays the results of several audiogames that were examined in the previously discussed evaluation. The analysis of the audiogames differentiates between *Mission, Difficulty, Acoustics, Design, Fun* and *Novelty*. *Mission* describes how easy the task/goal of the game was understood, while *Difficulty* measures the challenge that the game exhibited to the participants. *Acoustics* and *Design* are parameters for assessing the acoustic quality, as well as the overall design of the game, while *Fun* and *Novelty* measure the enjoyment during the gameplay and how the participants rank the game's idea. The values are derived from an analysis of the questionnaires, as well as through a visual observation of the participants during their game evaluation, ie. play. Table 12 discloses some partially large discrepancies between *Mission* and *Difficulty*, which also display themselves in other parameters. Interesting to note is that a very simple, yet audio-centered gameplay, such as for the game *Mosquito*, results in the highest enjoyment and rank. This emphasizes again the high importance of a well designed and intuitive gameplay. The game *Audio Quake*, which was also included in this evaluation, has here been omitted as all participants had huge difficulties playing this game due to a too complex user interface. Audiogames with a strong narrative component, such as *The hidden Secret* and *Der Tag wird zu Nacht*, also received a high ranking, but show that an interaction using head-tracking and sound spatialization provides a better understanding and gameplay (*The hidden Secret*). The final conclusion of this short evaluation is that the previously designed 3D scene sonification and interaction techniques together with the audio framework developed can very well be applied to the authoring and play of audio-only computer games. An audio-centered gameplay design, as well as the utilization of sound spatialization and user head-tracking improve and enhance the gameplay, as well as increase the player's immersion into an auditory game world.

The next section continues with an evaluation of an augmented and location-aware gameplay, and thereby (not only) evolves the story of *The hidden Secret* towards an interactive, augmented audio reality experience.

9.5 AUGMENTED AUDIO REALITY APPLICATIONS

An extension of the audio framework designed in Chapter 5 towards an augmented audio reality experience has been discussed and laid out in Chapter 6. This discussion included a review of the necessary requirements, both hardware and software, as well as an outline of possible areas of application. For a more detailed evaluation of the developed system, two applications have been discussed and are presented in more detail in this section.

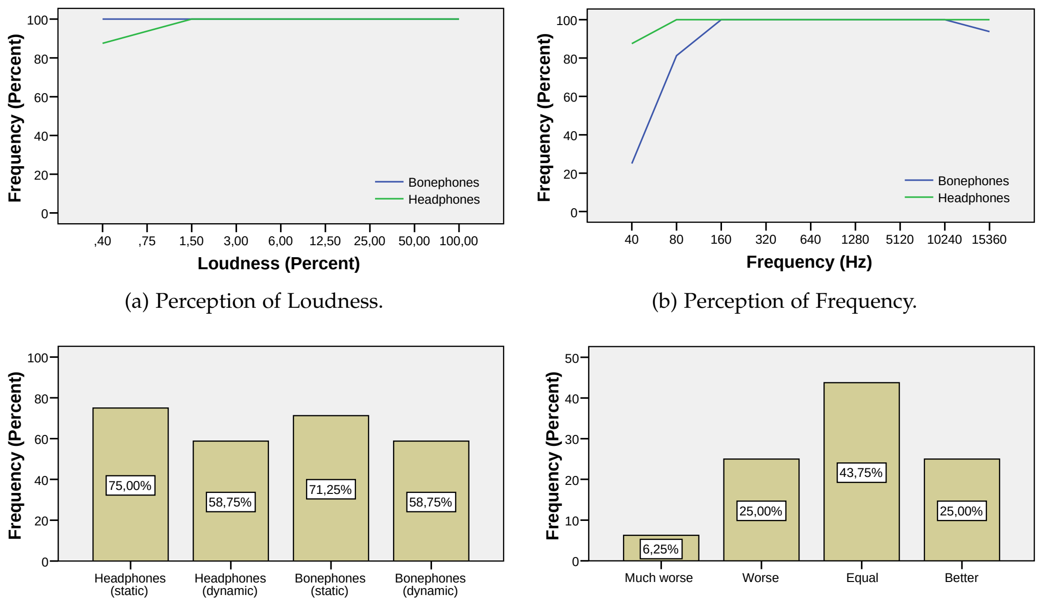
A large portion of the discussions in Chapter 6 was centered around suitable techniques and technology for a combined presentation of real and virtual auditory environments. The focus was centered around a presentation that allows the perception of both environments as one, and described the use of so called bone-conducting headphones for the display of the virtual acoustics. The first section in this evaluation is therefore dedicated towards a closer analysis of bone-conducting headphones in terms of quality and 3D sound perception, to examine their applicability within an augmented audio reality system.

9.5.1 *Sound Perception using Bonephones*

An inherent component of the augmented audio reality system devised in Chapter 6 is the presentation of the virtual auditory environment using so called *Bone-conducting Headphones*. Such bonephones have multiple advantages, but the perception of sound via skin and bone also has several drawbacks. Bonephones are relatively new on the market, and employed in areas that require a presentation of artificial sounds while one still needs to be able to listen to a natural environment. The headphones that are used in this analysis are developed by the *Vonia Corporation*[1] and are displayed in Figure 78. The application and employment of bone-conducting headphones makes them ideal candidates for an augmented audio reality system. The questions that arise are whether the perception using these special headphones permits a good understanding of speech and music, and especially, if it allows an interpretation of virtual acoustics and the localization of 3D virtual sound sources. Two initial evaluations of (different) bone-conducting headphones for

Figure 78: Bone-conducting Headphones *EZ-80P/S20*[1](Vonia Corporation, 2008).

1 http://www.dowumi.com



(a) Perception of Loudness.

(b) Perception of Frequency.

(c) 3D Source Localization Accuracy.

(d) Sound Perception using Bonephones is ...

Figure 79: Acoustic Perception – Headphones vs. Bonephones

an audio-based navigation were performed by Walker and Stanley and Lindeman et al. (Walker and Stanley, 2005; Lindeman et al., 2008). Walker and Stanley evaluated an earlier implementation of a bonephone system within an auditory VR environment and compared them to the perception with regular headphones (Walker and Stanley, 2005; Walker and Lindsay, 2006). The results were encouraging, though not overwhelming, as most participants performed better using the regular headphones. Lindeman et al. later evaluated in an empirical study the performances of a speaker array, regular headphones and a bonephone system for the auditory display of virtual acoustics within an augmented audio reality setting (Lindeman et al., 2008). The results showed that there were no difference in accuracy between the speaker array and the bone-conduction system for moving sound sources. The conclusion of this study is that both systems, including the bonephones, clearly outperformed a perception using standard headphones. Both evaluations, and especially the one from Lindeman et al., are very encouraging, although in both cases very different bone-conduction systems were used. Therefore, an additional evaluation of the bonephones that are employed in this augmented audio reality system had to be performed.

In order to evaluate the applicability of the bonephone system, several hypotheses were formulated in advance to describe the results anticipated:

- Sound perception using bone-conducting headphone systems probably causes partial impairments at certain loudness levels and frequency ranges, with
  - Expected difficulties for low loudness levels, and
  - Expected difficulties at very low, and very high frequency ranges
- Bone-conducting headphones can be employed for the perception of environmental acoustics and to localize virtual 3D sound sources
- Bone-conducting headphones perform qualitatively similar in the perception of speech, but overall less for music and high-quality acoustics

Based on these hypotheses, an evaluation was designed. The goal of this evaluation was to compare the acoustic quality of a regular headphone set (Hear0999 Audiosphere) with the Vonia bonephone system (AKG Acoustics GmbH, 2008; Vonia Corporation, 2008). Through this evaluation, the question whether or not these bonephones are applicable within an augmented audio reality system should be answered. Therefore, an evaluation based on four tasks was devised:

- Perception of varying levels of loudness
- Perception of varying frequency ranges
- Perception and quality assessment of different speech, music and environmental acoustic samples
- Source localization of stationary and dynamic 3D virtual sound sources

The evaluation itself was performed using Powerpoint slides, which described each task and presented the various sound files. Each participant performed the evaluation once for each headphone system, but using two different sets. 16 users (13 male, 3 female) participated in this evaluation, with two persons having a slight visual, as well as one a slight hearing impairment. The results of this evaluation are listed in Figure 79 and Table 13, while the auditory examples used in this comparison can be found on the right, as well as in Appendix C.

Questionnaire "Bonephone Evaluation".

Headphone Evaluation Tests.

Figure 79a and Figure 79b compare the perception of loudness and frequency using both systems and clearly show that the bone-conducting headphones lack a perception of frequencies below 100 Hz. As auditory cues for the perception and localization of 3D sound sources are encoded in the middle and higher frequency ranges, this auditory perception is highly sufficient. Figure 79c displays the correctly identified stationary and dynamic 3D sound sources for both headphones. The test data used was a helicopter sound with a close resemblance to white noise, which was spatialized using the AM:3D API (AM3D A/S, 2008). The results of Figure 79c reveal that nearly the same accuracy was achieved for both headphone systems, although, surprisingly, dynamic sound sources were less often identified correctly than stationary 3D sounds. This might be due to the use of sounds that were relatively difficult to identify, refer also to Section B.4 and Appendix C. Better results could have been clearly achieved using an additional head-tracking that allows a much more precise 3D source localization. An overall comparison of both headphones is displayed in Figure 79d, which shows that the participants rated the perception using bonephones as equally good. More results can be extracted by analyzing the questionnaires, which, interestingly enough, show that most participants rated the perception of dynamic sound sources as better, despite the opposite detection accuracy, compare Table 13 with Figure 79c. The perception of speech, music and acoustics is generally rated higher using the HiFi headphone system, which is clearly due to the missing lower frequencies in the auditory presentation of the bone-conducting headphones. Overall, the results in Figure 79 and Table 13 confirm the hypotheses and the anticipated findings.

Concluding this evaluation is the assumption that the bone-conducting headphones used can be very well employed in an augmented audio reality system as long as the application does not require the presentation of high-quality music and acoustic samples. Despite the fact that the detection accuracy for 3D spatialized sound sources was more or less equal, the development of specialized HRTFs tailored to a bone-conducting perception are an essential next step. These *bone-conducting transfer functions (BRTF)* can account for the differences in frequency perception and enable an even better 3D sound localization.

Perception Task	Performance (Mean) (poor/low (1) – great/high (5))	Std. Deviation	Std. Error of Mean
3D Sounds (static)	3.81	0.65	0.43
3D Sounds (dynamic)	3.94	0.93	0.86
Speech	4.38	0.72	0.52
Music	4.31	0.70	0.45
Acoustics	4.56	0.51	0.26

(a) Regular HiFi Headphones (AKG Acoustics GmbH, 2008).

Perception Task	Performance (Mean) (poor/low (1) – great/high (5))	Std. Deviation	Std. Error of Mean
3D Sounds (static)	3.19	1.22	1.50
3D Sounds (dynamic)	3.44	0.96	0.90
Speech	3.88	1.09	1.18
Music	3.69	1.19	1.43
Acoustics	3.56	1.36	1.86

(b) Bone-conducting Headphones (Vonia Corporation, 2008).

Table 13: Sound Perception with Bone-conducting and normal Headphones.

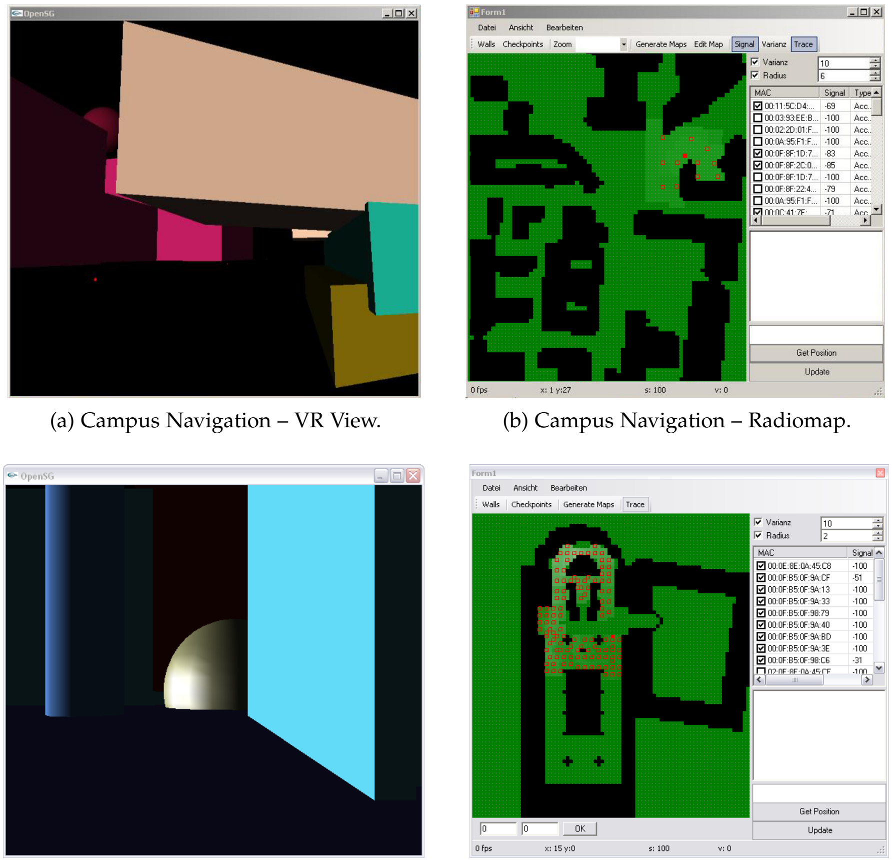
9.5.2 *Examples and Applications*

Augmented audio reality can be employed in a variety of applications and tasks. This includes augmented audio reality games, but also more serious applications, such as guiding and training simulations for tourists and the visually impaired. The examples that are discussed in this section were authored using the system presented in Section 7.3 and evaluated using the augmented audio reality system developed in Section 6.3.

Figure 80 provides an overview of both example scenarios. The first example represents an assistance and training simulation for the visually impaired and is located on the campus of the University of Magdeburg, while the second example is an augmented audio reality implementation of the story of *"The hidden Secret"* and is located in the Cathedral of Magdeburg. Here Figure 80a and Figure 80c show a *visual impression* of the system during the evaluation, while Figure 80b and Figure 80d display the radiomaps employed for the WiFi-based user positioning. The VR views show the virtual environment from the listener's current position and orientation and are employed for test supervision only.

Campus Navigation

The first example represents a training simulation for the visually impaired that is set on the campus of the University of Magdeburg. The major objective was to devise an application that allows an acoustic enhancement of the University campus to improve the navigation and orientation of visually impaired users (Röber et al., 2006a; Deutschmann, 2006). Integrated in this application is a training component that allowed users, prior to the use of the actual system, a familiarization with the augmented environment in an off-line training simulation using a desktop-based computer system. The interaction with this training component is performed using a keyboard/mouse combination, but can also



(a) Campus Navigation – VR View. (b) Campus Navigation – Radiomap.

(c) The hidden Secret – VR View. (d) The hidden Secret – Radiomap.

Figure 80: Augmented Audio Example Scenarios

employ a regular gamepad. Within the augmented audio reality system, this interaction is replaced by user tracking and positioning techniques. A previous familiarity with the environment and the auditory icons assigned allows thereby a better orientation and navigation in the final application. The modeling of the 3D environment for this training simulation was performed using 3DStudioMAX, refer also to Section 7.3. Descriptive sounds for each building's identification were taken from an available sound pool, but also created from recordings on the campus itself. The sounds applied are auditory icons and represent, for example, a rattling of plates and cutlery to denote the cafeteria, a rustling of pages and books for the library and space/future-like sounds to acoustically represent the department of computing science. During the authoring, these sounds were assigned to each building along a definition of several object and positional dependencies.

Example Campus Navigation.

Unlike the second example discussed in the next section, this scenario and the augmented audio reality system were tested by sighted users only. The results achieved were good, nevertheless exhibited several points for improvement. The example was developed during an earlier stage of the augmented audio reality system and experienced difficulties with the user head-tracking and positioning. This first implementation of the system employed a larger and far less efficient digital compass for the head-tracking, which resulted in large latencies and even the omittance of values. Also the user positioning showed several problems, as only a few WiFi access points were available and were scattered over large distances. Adding to this problem was that the signal strength of the

(a) Setting up the Gear.

(b) During the Evaluation.

Figure 81: Evaluation of the AAR Game *"The hidden Secret"*

WiFi access points dissipated slowly in this outdoor application, which resulted in a very homogenous radiomap, compare with Figure 80b. Advantageous for an outdoor implementation of augmented audio reality is an existing vivid ambient sound environment, which directly improves the user's orientation by providing natural auditory landmarks. The authoring of outdoor AAR applications is less complex as many sound sources are already present. Additionally, the distribution of event locations is scattered over a larger area outdoors than indoors, which allows a better positioning as overlapping effects from position dependencies are easy to avoid.

"The hidden Secret"

The example of *"The hidden Secret"* is an augmented audio game and set in the Cathedral of Magdeburg. It unfolds a storyline similar to the previously discussed auditory adventure game with the same name, refer also to Section 9.4. The augmented audio reality system employed in this implementation is the one described in Chapter 6, and uses a new, more efficient compass for user head-tracking. As the prototype and the example discussed in the last section were evaluated using an informal testing only, the new system had to be evaluated and examined more thoroughly.

The main research questions and hypotheses for this evaluation were:

- Evaluation and assessment of the systems overall performance
  - Positioning accuracy of the WiFi-based user tracking
  - Orientation accuracy of the 3D head-tracking
  - Efficiency and accuracy of the 3D pointing and selection
- Perception and experience of the auditory overlay, ie. how well are both, the virtual and the artificial, environments perceived as one?
- Expressivity, effectiveness and performance of the employed sonification and interaction techniques
- Presentation and perception of the storyline (immersion)

For the evaluation of this system, two example scenarios have been designed using the 3D authoring environment discussed in Chapter 7. Both scenarios required a setup and evaluation of the system within the Cathedral of Magdeburg, refer also to Section 9.5.2. The evaluation itself, as well as the questionnaires used used are divided and grouped into three main sections:

Questionnaire "Augmented Audio Reality".

- General perception and classification of the augmented audio reality system
- Scenario 1 – Path tracking and path following
- Scenario 2 – A play and interaction with the augmented audio reality game *"The hidden Secret"*

The first section thereby assessed the general perception and performance of the augmented audio reality system, while the second and third part explicitly focussed on an evaluation of the two examples. Scenario 1 required the tracking and following of a virtual auditory pathway, while Scenario 2 allowed a partial interaction with the story of *"The hidden Secret"*, this time in its augmented audio reality implementation and played *on location*.

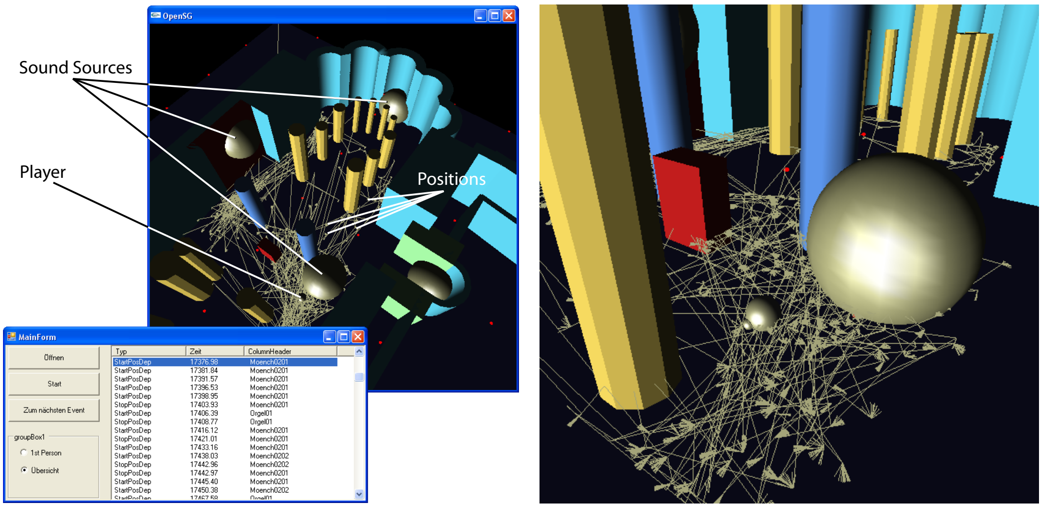
The evaluation of the AAR system required the largest and most complex setup. The radiomap that was employed for the WiFi-based user positioning was measured in advance to allow a quick start of the actual evaluation. The hardware that was employed for this evaluation was:

- One wearable computer system (laptop), equipped with
  - One set of bone-conducting headphones for sound presentation
  - One gamepad for regular interaction
  - One gyro mouse for 3D pointing and picking
  - One digital compass employed for 3D user head-tracking
  - One WiFi computer card equipped with an external antenna for user-positioning
- Nine portable WiFi access points

Impressions from the evaluation can be seen in Figure 81, which shows a setup of the gear and a participant wearing the system during the evaluation. Figure 80d displays the radiomap that was employed for the user positioning and shows the distribution of measurement points through the entire venue. Unfortunately, a large area in the center of the location could not be mapped and used for the WiFi-based user positioning due to an ongoing archeological excavation, refer to Figure 80d. This introduced partially large errors for the user positioning, which also affected the accuracy and the perception of other components. However, for about half the participants the positioning worked quite well, as they described an experiencing of both environments as one, as well as felt immersed in the story. These persons moved relatively slowly through the environment, which seemed to provide the positioning system with enough time to adapt.

Demo of "The hidden Secret".

A total number of 13 users (10 male, 3 female) participated in this evaluation, of which three users were completely blind, as well as one had a slight visual impairment. The range in age was between 20 and 59 and the majority of users had no or limited experiences with 3D interactions and auditory display systems. The following section summarizes and discusses the results of this evaluation, as well as emphasizes the problems and difficulties experienced.



(a) Evaluation System. (b) Detail with Player at Start Position.

Figure 82: Augmented Audio – System Analysis.

9.5.3 *Discussion*

Table 14 shows a weighted analysis of the evaluation of the augmented audio reality system using the example discussed in the last section. Key aspects in the evaluation are the accuracy of the user positioning, the functionality of the head-tracking and 3D interaction, as well as the system's latency and the mapping of the virtual auditory environment onto a real location. Several parameters are thereby dependent and influence each other. One of the major difficulties encountered is the low accuracy of the user positioning system, which is also visible in Table 14. The problems in this area are a result of the system's higher latency, as well as difficulties with the radiomap measurement itself. Some areas in the environment could not be mapped, which resulted in larger positioning inaccuracies in neighboring areas. Additionally, the positioning algorithm required up to three seconds for the determination of a location due to an interpolation of several measurement cycles. These inaccuracies in the user positioning, however, also influenced the entire perception and performance of the augmented audio reality system, and resulted in a less accurate perception of the head-tracking component. The assessment of the bone-conducting headphones is concurrent to an evaluation in an earlier section. The auditory design and quality of the scenario itself was praised. The implementation of the 3D pointing and object selection using the gyro mouse worked well as well. During the evaluation, several 3D objects, such as the organ, could be selected and were emphasized acoustically using an auditory icon.

Figure 82 displays a screenshot of an additional evaluation tool that visualizes the user's position and orientation, as well as the interaction using the 3D interactor and the sounds and objects activated. It clearly displays the varying accuracy of the positioning system, as some areas exhibit an accuracy of +/- 2 meter, while others contain outliers of up to 12 meter. Additionally, a clean function of the head-tracking component and the 3D interaction can be observed and confirmed using this application. Their accuracy, however, has been perceived less due to the poor performance of the positioning system. This also impaired the mapping between the virtual and the real environment and reduced the overall perception of the entire system, see also the demo on the previous page.

Technique/Task	Performance (Mean) (poor/low (1) – great/high (5))	Std. Deviation	Std. Error of Mean
System Handling	3.46	0.88	0.24
Task/Mission (Clearness)	3.46	1.25	0.35
Difficulties	2.87	0.71	0.19
System Latency	3.46	0.83	0.23
Quality Headphones	4.00	0.91	0.25
Real-World Mapping	3.23	0.90	0.25
Acoustic Design	4.12	0.78	0.22
Positioning (Accuracy)	2.67	0.93	0.26
Orientation (Accuracy)	3.27	1.09	0.30
Source Differentiability	3.75	0.82	0.23
Head-tracking (Function)	3.54	1.26	0.36
3D Interactor (Function)	3.94	0.52	0.16
Fun & Enjoyment	4.46	0.72	0.20

Table 14: Augmented Audio Reality – System and Application.

Although the devised system had difficulties with the accuracy of the user positioning, the concept itself appears to be valid. Improvements to the positioning accuracy can be made through an employment of additional proximity-aware technology (eg. Bluetooth), and the use of a larger number of of WiFi access points (Otto and Kurth, 2008). A narrative presentation that involves the display of images, or which takes place in a real environment, *reduces* the setting to what is displayed and seen. An audio-only presentation exhibits here a much more immersive presentation and achieves a higher involvement through the absence of visual cues. Therefore, the next section explicitly concentrates on the narrative component of auditory environments and devises the application of *Interactive Audiobooks*. One of the example implementations discussed also employs the story of *"The hidden Secret"*, which can now be examined from a third – more narrative – perspective.

9.6 INTERACTIVE AUDIOBOOKS

Over the last years, audiobooks and radio plays have enjoyed a constant increase in popularity that is still on the rise. One reason is their convenient usability, which results in a use of audiobooks by people who are committed to another – possibly boring – task, like driving or ironing, that requires visual, but no continuous auditory attention. The lack of visual information requires an active participation and a focused attention of the audience in order to reconstruct the fictional story universe. Auditory presentations are therefore, compared to visual depictions of the same content, considered to be much more stimulant and immersive. A drawback of audiobooks and radio plays is their linear storyline, at which, if heard once, the user knows the story as well as its ending. This changes with the introduction of *Interactive Audiobooks*, in which the listener/player may and can intervene with the story at predefined and user-selected points using an auditory user interface. One of the major differences of interactive audiobooks compared to the

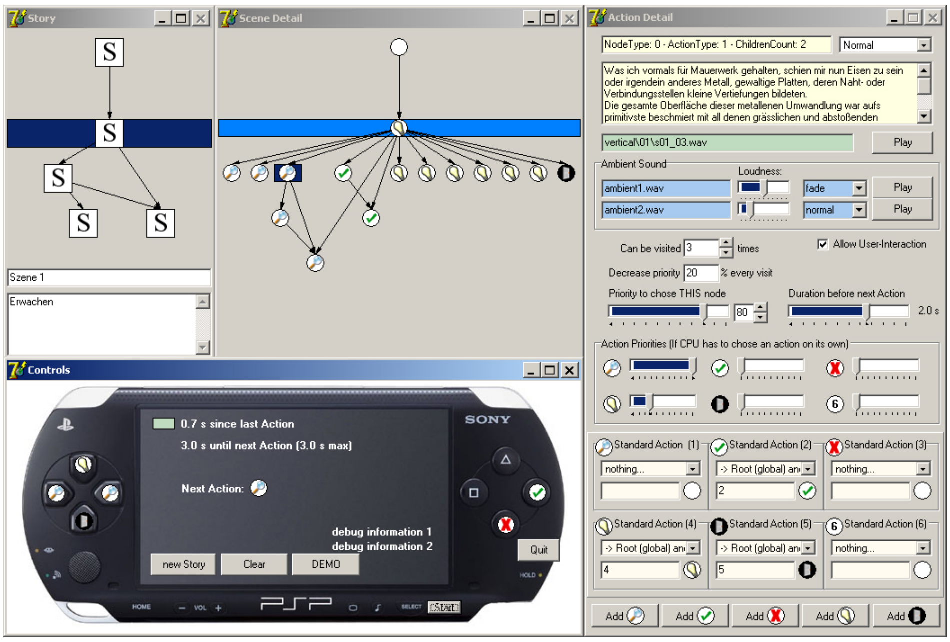


Figure 83: Interactive audiobook authoring environment.

applications discussed thus far is an environmental representation that is not bound by a spatial (3D) setting. The interaction and navigation in interactive audiobooks is performed along points in the storyline, and not along spatial dimensions.

The concept of interactive audiobooks aims to combine complex narratives with game elements from adventure- and audio-only computer games. Using story-dependent interactions, players can influence the development of the plot and steer it in their own direction, thus combining the advantages of an auditory storytelling with the benefits of interaction and the experience of a more personal storyline. One requirement for this is the creation of a story structure that permits a non-linear storytelling, as well as an integration of interaction and action elements. Story-graphs exhibit here an interesting alternative and are often employed by game designers to provide alternative plot and game endings that are consistent with the player's performance. To maintain the plot's consistency, functional dependencies between plot elements have to be considered (Murray, 1998; Hartmann et al., 2005b). Therefore, payoff elements may trigger the inclusion of setup elements. Also additional story correction techniques have to be specified and laid out by the game designer, eg. the player should not be able to kill one of the main characters if this is not part of the storyline.

9.6.1 *Narration and Interaction*

To combine interaction and narration into one structure, a story-graph system can be used as a basis, which is later extended by *interaction nodes*, refer Figure 84. These interaction nodes contain dialogs and story-dependent minigames and allow the user to control the plot within a predefined range. The story branches at previously defined points, at which decisions and challenges in the form of interactive components are placed. The final path through the story-graph depends on the player's actions and decisions, but also on the main character's conduct that is defined by the story-engine. The amount of interaction is thereby adjustable and can be varied smoothly throughout the storyline. This allows the user to start the story as a regular audiobook that runs autonomously, and later change it into an *interactive* audiobook with additional game- and interaction components. This varying interactivity is possible through the story-graph structure, which allows an automatic branching through the story parameters available, see also (Röber et al., 2006b). Playing and *winning* interactive parts results in rewards and benefits for the storyline. If a minigame is lost, however, the story branches into a different path and *penalizes* the player. A visualization of a simplified story-graph structure can be seen in Figure 84.

Narration Nodes

Narration nodes are the non-interactive parts of an interactive audiobook and represent the basic narrative elements of the story. They contain the majority of narrative information, eg. the narrators voice, (internal) monologues, non-interactive dialogs, and ambient and environmental sound effects. The narrator of the story introduces the initial setting of the fictional universe, as well as advances and controls the storyline. The main character's monologues provide additional hints and guidance for the player, but must not dominate his decisions. Ambient and environmental sound effects, as well as background music contain no narrative information, but intensify the atmosphere and deepen the player's immersion into the virtual environment. Additionally, they can provide information that can not be conveyed using regular narration nodes.

Interaction Nodes

Interaction nodes are placed in between and sometimes in exchange of narration nodes. They comprise story-related minigames, dialogs and techniques to influence the storyline and the main character's behavior. For the input of the interaction and for a play of the minigames, a regular gamepad is employed. Within small games, players can re-enact certain story events and therefore add personal experiences to the storyline. These games thereby focus either on action and a fast user reaction (arcade-style), or on a precise listening using a 3D interface to search for various items and hints. Interactive dialogs are designed with predefined answers, which also consider the mood of the main character to determine the right selection, refer (Huber et al., 2007).

A first approach employed only a small set of interaction primitives (eg. *think, look, do it, do not*, and *exit*), while the second implementation was based on an indirect control using a behavior classification scheme (Huber et al., 2007; Sasse, 2007). Actions and interaction are now customized to the current situation and position in the storyline. Not all decisions may lead in a different conclusion, but the selections made influence the outcome.

9.6.2 *Implementation and Design*

The developed system is divided into two parts and consists of an authoring and a runtime component. The authoring itself is further divided into the motif authoring, which is used to lay out and construct the story-graph, and an interaction authoring component (Röber et al., 2006b; Hartmann et al., 2005b). The system is based on the PC platform and uses OpenAL/EFX for sound rendering and for the acoustics simulations. An initial design goal was to develop a runtime component for the PlayStation Portable (PSP), therefore the evaluation component within the authoring environment, compare with Figure 83, still displays a PSP in its center (Huber, 2006; Sasse, 2007).

Figure 83 shows a screenshot of the interaction authoring environment. It visualizes in the top left view the motif-graph of the entire story and provides in the middle section a more detailed overview of the current selected scenes along their narrative and interaction nodes. The larger window on the right hand side presents all information associated with the currently selected node, and authors can intuitively adjust and specify the acoustic parameters for narrative nodes and define the systems behavior at points of interaction.

The design and adaptation of existing stories is not too difficult, but possesses several challenges. One of the largest is the design of a convincing storyline that permits an interesting game play, as well as allows the integration of interaction nodes and minigames into this story arc. Two interactive audiobooks have been created and evaluated, one as an adaptation of a short story from Edgar Allan Poe, and another with the familiar story about a tourist visiting the cathedral of Magdeburg, refer to Section 9.4. Both were further examined and evaluated through user studies. Table 15 shows the results of an evaluation for the most recent system with the story of *"The hidden Secret"*.

"The Pit And The Pendulum"

The story of *"The Pit And The Pendulum"* is one of the most popular works by Edgar Allen Poe and contains elements of the grotesque and arabesque. The plot is staged in a dark and wet dungeon during the time of the Spanish Inquisition. The description of the trial and death sentence of its protagonist is very short and only vaguely outlined, with the story itself centered around his endeavors to explore and escape his prison. One of the most scary parts in the story is as the hero awakens under a large pendulum with a razor sharp blade that slowly moves towards his chest. After several deadly situations, the hero is safely rescued in the last minute by French forces under the command of General Lasalle (Poe, 1843). The plot as it appears in the original text is entirely contained in the interactive audiobook and can be experienced in the very same way. Adding to this, the story has now three different endings, additional narrative to support a broader and more complex storytelling, three dialogs and several minigames. Depending on the user's selections and interaction, the story experienced can be the same as the original one, or be completely different. A simplified story-graph of the interactive audiobook is presented in Figure 84. It highlights the key narrative components, the added narrative and the dialogs, as well as shows several mini-games (Huber, 2006; Röber et al., 2006b).

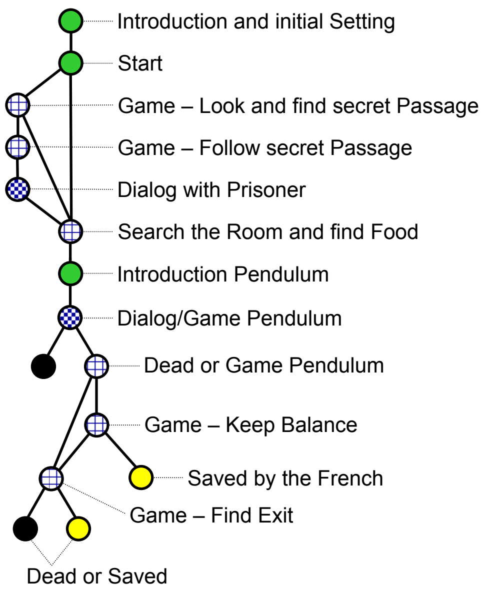


Figure 84: Simplified Story-graph.

Example of "The Pit And The Pendulum" (Narration).

Example of "The Pit And The Pendulum" (Minigame).

Per default, the story traverses as it appears in the original text, and without interaction, the story ends the same way. This first implementation started with a tutoring level that teaches the user interface along the gameplay, as well as the interaction techniques available. The duration depends on the degree of interaction and varies between 20 and 30 minutes. A short evaluation with 17 participants between age 20 and 59 revealed a recognition of the innovative concept of interactive audiobooks. About two third found the user interface intuitive, while several experienced difficulties to find the minigames and the interactive parts. This proves that the concept itself is valid, yet some aspects

Technique/Task	Performance (Mean) (poor/low (1) – great/high (5))	Std. Deviation	Std. Error of Mean
Game/Play Principle	4.00	0.57	0.22
Immersion	4.57	0.56	0.20
Story non-linearity	3.71	1.11	0.42
Minigames (Difficulty)	4.44	1.50	0.75
Minigames (Enjoyment)	4.03	0.53	0.26
Navigation & Control	3.38	0.75	0.28
Implementation	4.33	0.27	0.10

Table 15: Interactive Audiobooks.

had to be resolved. Therefore, a second version of the interface has been devised, and evaluated using the familiar story of *"The hidden Secret"*.

"The hidden Secret"

As the evaluation of the initial user interface exposed a few problems, several alternative approaches were evaluated that led to a modified interface and interaction design (Huber et al., 2007; Sasse, 2007). The main character/storyline is now affected and indirectly controlled using four types of interaction: *thinking*, *aggressive acting* and *defensive acting*, as well as *passive waiting*. The interaction still uses a gamepad, but with several added functionalities. The concept of interactive audiobooks moved a little closer to an audio-only adventure game and features more of their characteristics. Some of the new functions allow now a more detailed exploration of the local *auditory scene*, but the story-graph structure and the point-based exploration of the *3D* environment were retained.

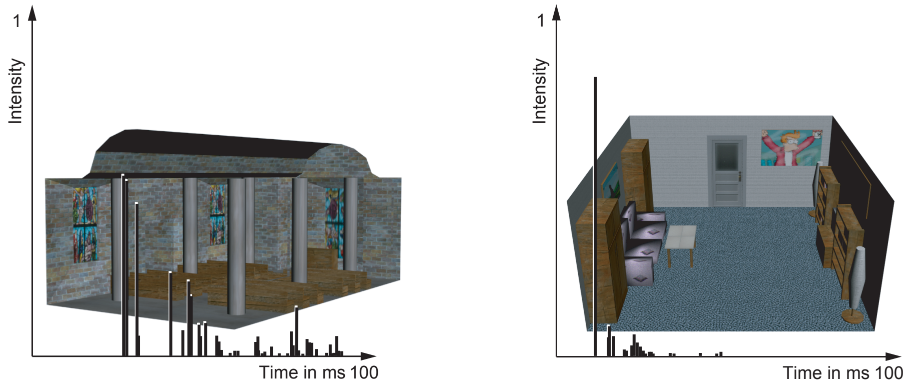
The story in this new example is very similar to the story employed in two previous examples, refer to Section 9.4 and Section 9.5. However, the here employed story arc contains additional narratives that is not available in the other implementations. A comparison with the other forms of presentation regarding immersion, interaction and presence is, nevertheless, still possible.

Questionnaire "Interactive Audiobooks".

Example of "The hidden Secret".

The story itself includes three minigames. Each has a different focus and requires alternative techniques to solve them. The first one is a small action game based on fast reactions, while the second game concentrates on the users puzzle solving skills. The last one is a quite difficult auditory puzzle, in which the player has to find a hidden entrance in order to find the cathedral's long lost treasure. Additional hints are provided throughout these minigames, if the user appears to have difficulties. Table 15 displays the results of an evaluation for the story of *"The hidden Secret"*, which clearly shows the very high acceptance of the concept. Interaction, control, implementation and immersion all received very high scores, while it is surprising that the system is not perceived as non-linear as it actually is. This might, however, result from the fact that most users only played it once, and did not explore the possibilities of choosing an alternative path in the storyline.

Overall, the concept of interactive audiobooks seems to be very successful and valid, although some minor problems still exist. One of the biggest challenges is the development of an intuitive user interface that immediately absorbs the player into the story arc, and which does not present itself as such. A second issue is an implementation on mobile, and thereby less efficient hardware. This requires a strong and versatile sound API,



(a) Model and Echogram Church. (b) Model and Echogram Living Room.

Figure 85: Ray-based Room Acoustics – Example Scenarios.

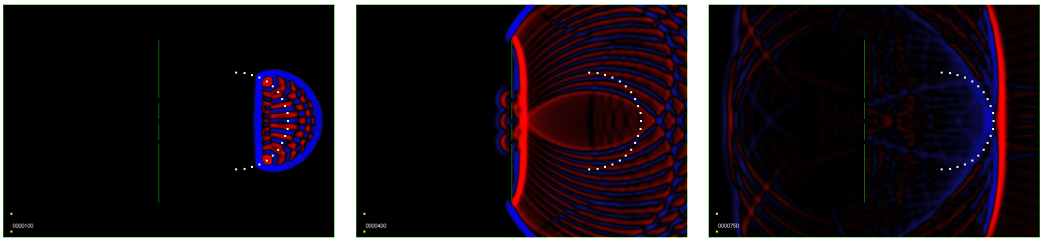
which is already available as the *PAudioDSP* API developed by Stockmann (Stockmann, 2007). An interesting project was here described in 2008 by Brode et al., in which listeners experience the narration of fictional stories at real locations in Berlin (Brode et al., 2008). Their system is based on a GPS localization, and once the listener reaches a certain point, the storytelling starts.

9.7 SOUND RENDERING AND SIMULATION

The components that are of the highest importance to all of these applications are the techniques for 3D sound rendering and room acoustics simulation. Chapter 8 examined the requirements for an acoustic rendering and an auditory display of 3D virtual auditory environments in more detail. It further motivated and presented several graphics-based sound rendering and simulation techniques, whose results are discussed and analyzed in this section. The simulations have been performed on a standard PC equipped with a P4 3GHz processor, 1GB of main memory and an *n*Vidia GeForce 8800GTX graphics accelerator. Chapter 8 thereby concentrated especially on the development of techniques for:

- A GPU-based sound signal processing (Table 6),
- An efficient implementation of 3D waveguide meshes (Table 8), as well as
- On a ray/energy acoustics simulation approach (Table 9).

The two most important applications for the techniques developed are room acoustics simulations, as well as the simulation of virtual HRIR measurements. The two following sections extend the discussions of results that was started in Chapter 8. In this section, the focus lies on a more detailed analysis of an application of both techniques, eg. wave- and ray-based acoustics, towards the simulation of room acoustics and the measurement of room- (RIR) and head-related impulse responses (HRIR). Both are essential ingredients for an auditory display of 3D virtual auditory environments, as one provides environmental, and the other directional and distance information of 3D sound sources. Additional details and results can be found in (Andres, 2005; Röber et al., 2006,c, 2007; Kaminski, 2007).



(a) Wavefront at $t = 100$. (b) Wavefront at $t = 400$. (c) Wavefront at $t = 750$.

Figure 86: 2D Wavefield Synthesis.

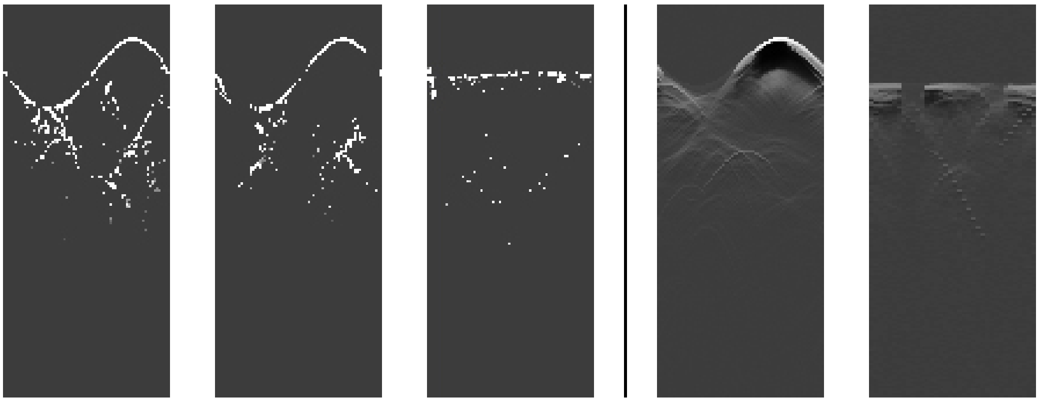
9.7.1 *Room Acoustic Simulations*

Room acoustics simulations are generally concerned with the measurement of room impulse responses (RIR) to derive acoustic qualities and to determine a room's *acoustics*. The characteristics of a specific room are thereby imprinted into the response, and if later convolved with a dry sound file, it will sound as if played within the original environment. A physically correct and efficient technique to perform such virtual room impulse response measurements is thereby of great importance, as the acoustic qualities of yet-to-be-constructed buildings can be verified and adjusted in advance.

Two simulation systems have been implemented and evaluated: One focussing on a more physically correct modeling using 3D waveguide meshes (Section 8.3), while the other approximates sound waves using directional rays and ray tracing techniques (Section 8.4). Section 8.5 already discussed the differences of both techniques, but also showed several ways to combine the two methods. Both implementations can be employed for virtual RIR measurements, as well as for a direct sound rendering approach, eg. for a continuous auralization of the acoustics. Figure 85 shows two examples from the ray-based acoustics simulation system. The echogram of the church in Figure 85a shows strong late reverberation effects, while the echogram of Figure 85b shows that nearly all acoustic energy has been absorbed by the walls and furniture in this room. The authoring of both 3D scenes could be conveniently performed using 3DStudioMAX, for which a custom-built plug-in was used to assign an acoustic material to each object. The properties for these acoustic material definitions were taken from the CARA database (ELAC Technische Software, 2008). These material definitions also include a surface roughness factor and a material density, as well as wavelength specific coefficients for sound wave absorption, reflection, refraction and transmission. Another example can be found in Figure 66 in Section 8.4, which is very interesting, as it visually verifies several of the implemented sound propagation effects.

Wavefield Synthesis.

An example animation from the wave-based sound simulation system that visualizes the propagation of sound waves can be seen in Figure 61 in Section 8.3. It shows four time frames of an animation and highlights especially the occurring interference and diffraction effects. The red and blue waves denote positive and negative acoustic pressure, while the sound source is marked by a blue and the two microphones by yellow dots. The walls (green) and ceilings reflect the sound waves in a phase-reversing manner and a short sine pulse was used to excite the mesh at the speaker's position (blue mark). Another application for this technique can be seen in Figure 86, which shows an example of wavefield synthesis. This experiment is based on Huygens principle, and aims at the generation of large wavefronts by combining the acoustic energy of several smaller waves (Boone, 2001). If arranged in the form of a circle, speaker arrays can be used to synthesize



(a) Horizontal. (b) Frontal. (c) Sagital. (d) Horizontal. (e) Sagital.

Figure 87: Virtual HRIR Simulations using a KEMAR Head Model.

large parallel wavefronts to simulate and study wave-based phenomena in greater detail, see here Figure 86a. The example shows a large wavefront as a result of 21 sound sources that excite the mesh time controlled using a sine wave with a frequency depending on the sound source separation. The animation shows diffraction and interference effects, as well as the modeling of anechoic outer walls and the phase-reversing reflections from the inner walls. Wavefield synthesis also plays a huge role in creating ultra-realistic auditory environments, such as 3D audio theaters for which this system can be employed as an evaluation system, as well as for the actual simulation (Boone, 2001; Gräfe et al., 2007).

9.7.2 *HRIR Simulations*

Head-related transfer functions are an essential ingredient for the spatialization of monaural sounds. Unfortunately, these HRTF vary from person to person and the differences can be quite substantial, refer Section 3.2 and Section 8.1. Therefore, one part of the research on acoustic rendering was concerned with the idea to employ ray-based sound simulation techniques to simulate virtual HRIR measurements, thus creating personalized HRTFs for a better perception of 3D sound spatializations. Two different approaches have been explored, one with a slower, but more accurate offline simulation, and a second one that employed an implementation using efficient graphics hardware (Andres, 2005; Kaminski, 2007).

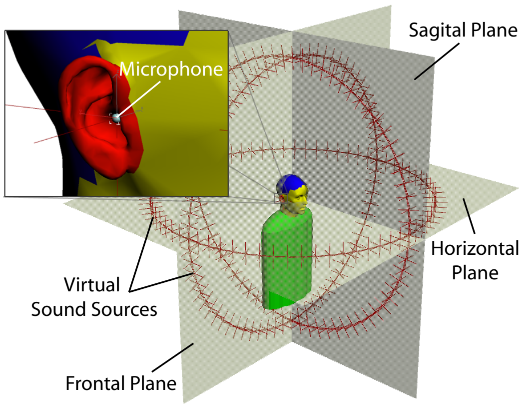
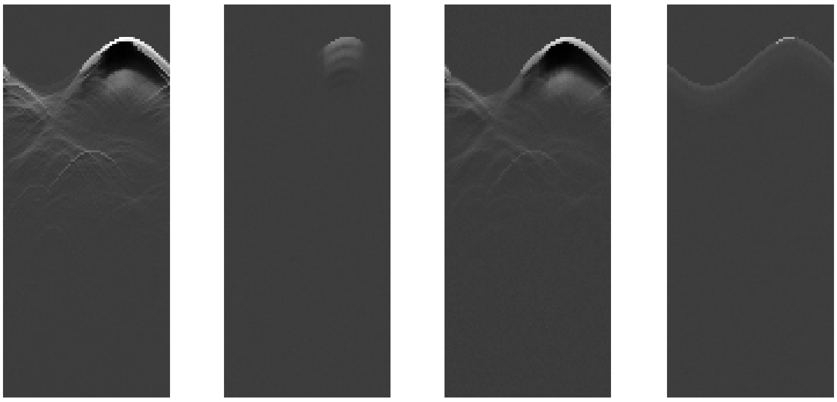


Figure 88: HRIR Simulation System.

The experiments for the virtual HRIR measurements were performed using a 3D model of the KEMAR mannequin. A total of 72 virtual sound sources per plane were thereby arranged in a circular array around the listener, each 1.2m apart at a 5° interval, refer to Figure 88. The size of



(a) Complete. (b) Direct. (c) Diffract./Reflect. (d) Transmiss.

Figure 89: Virtual HRIR Simulation of the Horizontal Plane (Propagation Effects).

the original 3D mesh was reduced for the simulation from 23k polygons down to only 5,500 polygons, leaving the shape of the ear unaltered, although one of the ears was removed for efficiency reasons. Four different materials have been defined and are specified to describe: *Skin, Body, Hair* and *Ear*, see also Figure 88.

Figure 87 displays five HRIR simulation results that were achieved using these two systems. The first three (ie. Figure 87a through Figure 87c) represent HRIRs of the horizontal, frontal and sagital plane – compare with Figure 88 – and were generated using the ray acoustics system described in Section 8.4.

HRIR Simulation – Horizontal Plane.

HRIR Simulation – Median Plane.

In each of these simulations, roughly 18 million rays were traced per sound source with a cube map resolution of 512×512, resulting in a simulation time of only 3 seconds per sound source, refer to Table 9. Although the most important features are clearly present, some effects are still missing, compare with Figure 87d and Figure 87e. This partially results from the fact that currently only one diffraction per ray is considered, therefore not allowing multiple diffractions per ray. Additionally, a more detailed fine tuning of the parameters along the material definitions of the head, torso and ear would clearly yield better results. The final goal of this application would be to combine individualized HRTF simulations with room acoustics simulations, to achieve a realtime personalized binaural room simulation.

A similar acoustic ray tracing system was developed earlier, based on the offline renderer POVray[2] (Andres, 2005; Röber et al., 2006). This system was entirely dedicated to the simulation of virtual HRIR measurements, of which some results are depicted in Figure 87 and Figure 89. Opposite to the other implementation, this simulation allows multiple diffractions per ray and uses an overall even higher number of rays to accumulate the acoustic energy. Therefore, all important pinna and shoulder echoes are clearly visible in the examples (Huang and Benesty, 2004), compare with Figure 87d and Figure 87e.

The results depicted in Figure 89 are also very insightful, as they visualize the importance of the individual propagation effects. It shows – of-course – the largest contribution originating in the diffraction/reflection system, although a lot of acoustic energy is also related to transmission effects, especially in the lower frequency ranges. These results proof the overall significance of the concept and show that virtual HRIR simulations are possible and valid.

2 http://www.povray.org/

9.8 CONCLUSIONS AND DISCUSSION

The last sections explored several areas of application for 3D virtual auditory environments, and have shown the high applicability of the research conducted in this thesis. Example applications and evaluation scenarios have been developed for several areas, including:

- The sonification of abstract 2D/3D data sets,
- The sonification of and interaction with 3D virtual auditory scenes,
- Audio-only computer games,
- Augmented audio reality and applications,
- Interactive audiobooks, as well as
- 3D Sound rendering and simulation techniques.

The majority of the examples were based on the audio framework that was devised in Section 5.5.1 and later extended in Section 6.3, as well as were authored and designed using the 3D environment developed in Section 7.3. Throughout this chapter, several of the proposed sonification and interaction techniques, as well as authoring and design guidelines have been applied to specific problems, and were studied and evaluated in close detail. Although the vast majority thereby performed as expected and even beyond, also some problems and difficulties emerged. These problems have been discussed and analyzed in detail, and solutions were developed to diminish these issues in future improvements.

The main goal of this research was to design and develop intuitive sonification and interaction techniques for the exploration of 3D virtual auditory environments, which are centered around an auditory perception and interaction. An analysis of related and similar work revealed that many audio-based applications, such as audio-only computer games, are still played and accomplished in front of a view screen. A constant motivation throughout this research was therefore to *Leave the Screen* and to devise techniques for an intuitive and natural interaction with auditory environments. This rethinking of auditory displays must continue and be further pursued.

The next chapter finally summarizes the research presented in this thesis and discusses issues of future investigations and improvements. Several propositions are thereby developed to summarize the topic and the research accomplished.

10 CONCLUDING REMARKS

After this, we will all have to live a little differently. — Rainer Maria Rilke[1]

IN this thesis, a number of ideas for an INTERACTION WITH SOUND were discussed, analyzed and illuminated from various perspectives. Several examples have been provided to emphasize the discussions, as well as to evaluate the results achieved. Concluding this analysis, this last chapter serves as a summary for the work, discusses open problems and current limitations, as well as provides several possibilities for future improvements. After the research in this thesis, the exploration of 3D virtual auditory environments will continue, but with new directions set and additional goals provided.

10.1 SUMMARY

The research in this thesis examined 3D virtual auditory environments and explored several associated areas of application. After a short introduction and motivation of the topic in Chapter 1, Chapter 2 started with an in-depth analysis of the subject matter and discussed the related areas of research. In a first step, several hypotheses were devised and a schedule of the research was developed. Both Chapter 3 and Chapter 4 discussed several required fundamentals, as well as related and existing research in the areas of sound & acoustics and auditory display systems. Chapter 3 provided a broad perspective on the entire area and discussed topics ranging from sound synthesis, propagation and perception, towards an employment of sound in entertainment computing. The succeeding Chapter 4 continued this discussion, but with a more focussed perspective on auditory display systems, and here especially on 3D spatial auditory displays along the established standards and applications.

Several conclusions towards the research goal could already be drawn from these initial discussions, and led towards a definition and design of 3D virtual auditory environments in Chapter 5. This chapter exclusively focussed on the modeling of 3D auditory environments, as well as on the techniques required for a sonification of 3D scene information and spatial interaction. Starting out with a research objective in the design of a 3D auditory display system that supports an efficient and intuitive perception, quickly methods for a non-realistic auditory display of 3D auditory environments moved into focus.

Proposition 1 *A **Non-realistic auditory Display** is essential for an efficient and intuitive auditory presentation of a 3D virtual auditory environment. For that, the display is altered towards a non-physically based acoustic representation of a 3D scene and the objects therein. This is achieved by integrating additional virtual sound objects, by an exaggeration and/or reduction of certain physical parameters/laws, as well as through the use of situation-based auditory display styles for object sonification.*

Despite this non-realistic approach in the auditory display, the majority of the applications discussed are based on and exploit a physically correct 3D sound spatialization, as well as 3D head-tracking techniques for an intuitive perception and 3D scene interaction. 3D sound spatialization is essential for the display of 3D virtual auditory environments, as it provides directional and distance cues for localized sources. Additionally, it enhances the segregation between several sound sources and streams, and thereby improves the

1 upon seeing Constantin Brancusi's 1919 sculpture *Bird in Space*

overall perception of the 3D environment. Head-tracking mimics as a technique a natural human listening behavior, which further improves the perception and evaluation of 3D sound sources.

Proposition 2 ***3D Sound Spatialization and Head-Tracking*** *are key elements for the display and the interaction with 3D virtual auditory environments. Both concepts enhance the perception of 3D auditory spaces and represent imperative techniques that are required for an adequate interaction/sonification of 3D virtual auditory environments.*

A large portion of Chapter 5 was dedicated to the exploration of suitable sonification and interaction techniques to convey information from a virtual auditory scene to the user, as well as to input information. Chapter 5 started with an examination of abstract data sonification techniques and devised methods for a global and local 3D scene sonification. Sonification and interaction are bonded and require each other in order to derive/input information from/into a 3D scene.

Proposition 3 *A task-dependent* ***Sonification and Interaction Design****, that is customized to the specific requirements of an application is required for all applications. Techniques of 3D scene sonification are thereby used to convey abstract information of a virtual 3D scene and the objects therein, while interaction techniques allow an input of information into the virtual environment. The techniques employed must enable an adequate interaction, orientation, navigation and wayfinding, and thereby convey local and global information of the auditory environment. The sonification techniques should be based on a non-realistic auditory design and aim at the most intuitive display of an environment's semantics.*

Utilizing this information, the last section of Chapter 5 was dedicated to the design and development of an audio framework, suitable for an evaluation of this research. The design of this framework was based on the sonification and spatial interaction techniques discussed, as well as employed 3D sound sources and user tracking technology. This section also discussed several areas of applications, of which some examples have been implemented and examined in Chapter 9.

Chapter 6 extended these concepts and techniques towards a design of an augmented audio reality framework. The discussion started with the requirements for such a system, and later explored spatial interaction techniques, as well as methods for a combined display of a real-world location that is augmented by a 3D virtual auditory environment.

Proposition 4 ***Augmented Audio Reality*** *describes a system and techniques that support an extension of a real-world environment using additional auditory information. The underlying 3D virtual auditory environment must be synchronized with the real location in terms of position, orientation and time. Further requirements for such a system are a non-realistic 3D auditory display, efficient techniques for user-orientation and -positioning, as well as methods for spatial interaction. The system itself is realized as a wearable computer that requires an implementation on mobile and lightweight hardware.*

A spatial interaction design allows an intuitive interaction with virtual environments based on a natural real-world interaction behavior. 3D sound spatialization and spatial interaction techniques are both required for an adequate representation of 3D auditory environments. Chapter 5 introduced the basic concepts, which can, due to a direct applicable spatial mapping, be used for an interaction with augmented auditory environments as well.

Proposition 5 *A* ***Spatial Interaction Design*** *mimics a real-world interaction behavior and provides a more intuitive interface for an interaction with 3D auditory environments. The* *techniques require additional user-tracking equipment, which measure the user's orientation and position and translates this information into virtual 3D scene interaction techniques. A variety of spatial interaction designs can be employed, ranging from 3D gestures and 3D pointing towards real object interactions. The techniques are related to the application's task and the tracking technology available.*

Several possibilities for an efficient and low-cost design of user-orientation and positioning techniques have been discussed and implemented. The augmented audio reality system developed is based on a WiFi-enabled user-positioning, a digital compass for 3D head-tracking and gyro-technology to implement various 3D interaction designs. Using this system, two examples have been prototypically implemented and were examined in more detail in Chapter 9.

Besides the actual techniques for an interaction with 3D virtual/augmented auditory environments, also authoring and design are of high importance and were discussed in Chapter 7. This chapter first explored common principles for the design of both 3D virtual and augmented auditory environments, and derived therefrom several authoring guidelines and design principles. These guidelines have been implemented within an additional 3D authoring environment, which was also deployed as authoring system in a variety of example applications.

Proposition 6 *Techniques of **Authoring and Design** assist the user in the design and setup of 3D virtual auditory environments. The authoring process includes the creation of 3D geometry, the design of sound, speech and music samples, the definition of dependencies and a selection of suitable sonification and interaction techniques, as well as the actual auditory scene authoring through the definition and setup of auditory textures. During the authoring and auditory scene design, one must adhere to an appropriate balance between a display's functionality and its aesthetic appearance.*

Both Chapter 5 and Chapter 6 expressed the importance of high-quality, yet efficient methods and techniques for 3D sound spatialization and a simulation of environmental acoustics. Using the current state of the art audio APIs, these requirements can only partially be fulfilled. Therefore, Chapter 8 examined sound and light wave propagation principles, and discussed the possibilities of using efficient computer graphics rendering techniques and hardware to aid 3D sound rendering and simulation. The chapter especially concentrated on the development of a GPU-based sound rendering for both, room acoustic simulations and 3D sound spatialization. A second focus of Chapter 8 was centered around the development of personalized HRIR filters to remedy several artifacts introduced by standard HRIR filters. Although the proposed techniques were not fully integrated into the 3D audio framework, it could be shown in the concluding analyses of both Chapter 8 and Chapter 9 that the devised concepts and techniques indeed largely enhance the quality and efficiency for 3D sound rendering and simulation.

Proposition 7 ***Graphics-based acoustic Simulations*** *greatly improve both quality and efficiency of 3D sound rendering and acoustic simulation. Several similarities between sound and light propagation can be exploited to exploit dedicated computer graphics hardware for an expedited and improved sound simulation. Graphics-based sound simulations can also be used for the simulation and measurement of individual HRIR filters to improve the perception of 3D sound sources and 3D virtual auditory environments.*

A number of applications and example scenarios were discussed and analyzed throughout the research of this thesis. Chapter 9 summarized the majority of these examples and presented them in a chapter- and topic-overlapping manner. Chapter 9 evaluated techniques for 2D and 3D data sonification, analyzed methods to interact with and display information within 3D virtual auditory environments, as well as examined several very specific applications, such as an augmented audio reality system, and audio-only computer games with an introduction of the concept of interactive audiobooks.

A central role in this thesis research is the evaluation of 3D auditory environments for entertainment and edutainment purposes. As a result, the audio framework developed was primarily designed with entertainment and edutainment applications in mind. A large section of Chapter 9 was therefore dedicated to the analysis of audio-only computer games. This section compared four own developments with existing audiogames in terms of playability, immersion, complexity and enjoyment. The audiogames developed were implemented on top of the audio framework and utilized the techniques for 3D scene sonification and spatial interaction.

Proposition 8 *An **Audio-centered Gameplay** concentrates on the specifics of an auditory perception and proposes a new way of play and interaction with audio-only computer games. Using adequate scene sonification and interaction techniques, the content and the story of an audiogame can be perceived with a high level of immersion. These characteristics make this form of presentation very suitable for a display of narrative content, ie. adventure-based computer games. Audiogames benefit effectively from the use of 3D spatial interaction techniques, as well as from a general employment of 3D sound spatialization in combination with user head-tracking. Similar to audio/visual computer games, also audiogames benefit from a simple and clear design, which in this respect focusses on an auditory perception and an audio-centered gameplay.*

Another very interesting application that was introduced in Chapter 9 are so called *Interactive Audiobooks*. They combine the immersive advantages of an auditory storytelling (ie. based on audiobooks and radio plays) with interactive elements from adventure based computer games. Further advantages of interactive audiobooks are a non-linear storytelling, as well as a varying degree of interaction. This allows to reexperience the same or a similar storyline, which can be either perceived passively as a regular audiobook, or actively as an audio adventure game. This becomes possible through the use of a story-graph structure that consists of narration and interaction nodes, and allows the construction of several different storylines with various endings.

Proposition 9 ***Interactive Audiobooks** combine the advantages of an immersive, non-linear storytelling with interactive elements from adventure-based audio-only computer games. The result is a highly immersive presentation of narrative content, which becomes even more effective through the added interaction. The level of interaction can be varied smoothly, which allows a free blending between a passive audiobook and an interactive adventure computer game. Interactive audiobooks can be played and controlled using only few commands that influence the main characters conduct, which is also responsible for controlling the audiobook in a non-interactive mode.*

The last section of Chapter 9 evaluated and analyzed the results from the graphics-based acoustic simulations that were developed in Chapter 8. An emphasis in this analysis was an application of these techniques for general room acoustic simulations, as well as especially a discussion of virtual HRIR simulations using ray-based sound simulations.

10.2 DISCUSSION

After this summary of the thesis and its research, the following section discusses the contributions of this work in respect to the individual research areas, as well as critically reflects the results achieved.

10.2.1 *Contributions*

The research that has been presented in this thesis has contributed to a variety of scientific areas and was published in a number of conference articles. The two major contributions are a redefinition and extensive analysis of 3D virtual auditory environments, as well as the research for a more efficient, graphics-inspired sound rendering and simulation.

2D/3D DATA SONIFICATION — The area of 2D/3D data, image and volume sonification has been advanced through a refinement of existing sonification techniques, as well as through an added spatial sonification for the exploration of 3D objects and volumetric data sets. It could be shown that an added spatialization and rhythmic sequencing improved the parallel perception of linear data (stocks), and that 3D interaction improved the understanding of 3D shapes and the topology of 3D volumetric data sets (Stockmann et al., 2008).

3D VIRTUAL AUDITORY ENVIRONMENTS — Due to ambiguities in its definition, the term *3D virtual auditory Environment* has been redefined in the confines of virtual reality and 3D auditory display systems using an abstract definition of VR/MR environments. The new definition focusses on an audio-centered design, as well as employs a non-realistic auditory scene description. A number of 3D scene sonification and 3D spatial interaction techniques were devised, implemented and their applicability also evaluated, including the concepts of an auditory cursor, -guides, -landmarks, -lens, sonar/radar and soundpipes system. Furthermore, the concepts of dependency modeling and auditory textures were devised and implemented to allow a broad spectrum for an interaction design (Röber and Masuch, 2004b, 2005b,a, 2006).

INTERACTIVE AUDIOBOOKS — The concept of *Interactive Audiobooks*, which combines the advantages of a non-linear narration with interactive computer game elements, has been developed and implemented. In two user evaluations, the concepts functionality could be confirmed and a high level of immersion was shown (Röber et al., 2006b; Huber et al., 2007).

AUGMENTED AUDIO REALITY — The concept of augmented audio reality has been advanced in terms of 3D spatial interaction, as well as through an evaluation of new areas of application. A low-cost, yet efficient system has been devised and implemented, and was employed in a user-guidance scenario targeting the visually impaired, as well as in a narrative augmented audio reality game (Röber et al., 2006a).

AUTHORING AND 3D SCENE DESIGN — Further contributions were made through an analysis of 3D virtual/augmented auditory environments and the development of 3D scene authoring and design techniques. An additionally implemented 3D authoring framework demonstrated the developed guidelines and principles in practice (Röber and Masuch, 2004a).

3D SCENE AURALIZATION AND SOUND RENDERING — Due to the high demands in acoustic realism, additional research was spent in the exploration of more efficient 3D sound rendering and simulation techniques. A number of graphics-inspired methods were discussed and implemented, including a technique for general GPU (sound) signal processing, as well as two systems for a GPU-centered ray- and wave-based sound simulation. Furthermore, advancements in the direction of virtual HRIR simulations for the creation of personalized HRTFs were discussed and developed (Röber et al., 2006,c, 2007; Röber et al., submitted).

10.2.2 *Critical Reflections*

One of the major differences between the research on auditory display systems and visual (computer graphics) display techniques is the availability of related work and the amount of research conducted. As we primarily live in a visually-centered environment, more research has been accomplished in the visual domain. Although the area of auditory display has left its infancy many years ago, the awareness of its potential, also in respect to new areas of application, is still underdeveloped. The research that was conducted in this thesis explicitly focussed in this direction and discussed and advanced 3D auditory display systems, both in technology, as well as in areas of application. Although the techniques that were developed to perform a 3D scene sonification and interaction are applicable to a broad variety of applications, a slight focus was placed in this research on entertainment and edutainment tasks.

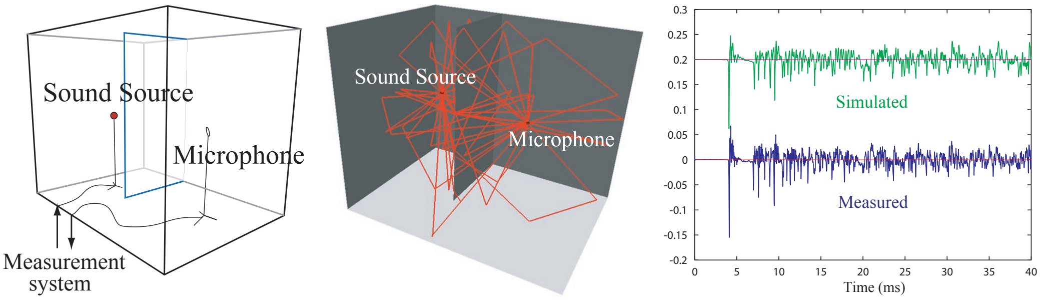
Chapter 9 discussed a broad spectrum of different applications that were prototypically implemented and evaluated. Selected applications and proposed techniques have been analyzed in more detail using user evaluations. Although the main goal of the research was not specifically the development of techniques to aid the visually impaired, the techniques discussed are, nevertheless, very applicable and useful in this domain. The focus on this specific group of users could have been stronger in specific evaluations, see also Appendix B. As this would also have shifted the center of this research, these analyses are left for future development.

During the development and implementation of the techniques and applications, several approximations had to be applied. These approximations include the realism of the sound rendering and synthesis techniques used, as well as in certain cases the design and the presentation of the auditory display systems themselves. The chosen applications and example scenarios focus on a single user presentation and interaction. The binaural sound rendering that is employed is – due to the use of 3D head-tracking – generally only applicable for a single person perception. Although a multi-user presentation and interaction environment can also be realized using the audio framework developed, but this would require major modifications in terms of sound rendering and especially the display of content. The spatial interaction techniques, however, are directly applicable in such a setting as well.

The audio framework that is used as basis in the majority of applications was developed based on OpenAL/EFX for sound spatialization and rendering. This API, however, is limited to standardized HRTF filters only, as well as uses several severe approximations for the simulation of room and environmental acoustics. Within the conducted user evaluations, several participants claimed to have difficulties in the localization of virtual 3D sound sources, which, no doubt, resulted from the use of generalized HRTFs. Chapter 8 examined therefore the possibilities of using computer graphics hardware and rendering techniques to improve the sound rendering process in terms of quality and efficiency. Very promising results were achieved, but could not yet be integrated into the audio framework itself. This future development could, through an increased perception of the 3D auditory scene, shed new light on the performances of the developed 3D scene sonification and interaction techniques as well.

10.3 FUTURE IMPROVEMENTS

A thorough research poses often more (*new*) questions than it is able to answer. The research within this thesis has been extensive, yet several areas and ideas remain for future development and improvement. Specific possibilities for future improvements



(a) System Setup. (b) Simulation. (c) Comparison.

Figure 90: Evaluation of Acoustic Rendering Techniques using the *Bell Labs Box* (Tsingos et al., 2002).

have already been outlined and discussed previously in the respective chapters. This final section summarizes the most interesting and promising areas of future research, and provides directions for further improvements. The discussion is thereby divided into aspects of interface design, technical improvements and further areas of application.

Interface Design

The last section already started a discussion on important interface related issues, such as multi-user presentations and interactions, but also a stronger focus on the requirements of the visually impaired. This not only requires additional research regarding sound rendering and the auditory display of 3D virtual environments for group perception, but also includes different and modified applications to enable group interaction. An interesting direction is here, which is also applicable to single-user systems, a perceptual presentation. This technique, similar to the MP3 music format, only considers those sources and acoustical effects that are audible in the final presentation. Moeck et al. have studied a first approach for a progressive perceptual audio rendering of large and complex scenes (Moeck et al., 2007). This approach is highly applicable to the sonification of 3D auditory environments as well, and could also reduce the requirements and complexity of acoustic rendering.

But also *classic* future improvements of interface design are important, such as a detailed analysis of the individual techniques, especially in combination with the enhanced graphics-based sound simulations, as well as a generally improved user interface and a further exploration of advanced spatial interaction techniques.

Acoustic Rendering

At the time of writing, the enhanced sound rendering and simulation techniques are not yet fully integrated into the audio framework developed. Although their function and general applicability has been confirmed using several tests and evaluations within this thesis, a real-world comparison would finally reassure their performance. One possibility is here to use the so called *Bell Labs Box* as is depicted in Figure 90. This system has been used by Tsingos et al. to study virtual sound simulations and to compare them with real-world measurements (Tsingos et al., 2002). A similar approach could be used to further evaluate the graphics-based wave- and ray-acoustic simulation techniques that were developed in this thesis.

Another interesting direction of research is here the perception of sound and which propagation effects are most significant. This is especially important for the advancement of virtual HRIR simulations, and could enhance the overall 3D sound perception.

A third technique that is potentially interesting for a graphics-accelerated implementation is ambisonics. Ambisonics can be realized in graphics hardware using an implementation of spherical harmonics, which are currently applied in rare cases for global illumination effects (Dempinski and Viale, 2005).

Physics Simulations

In 2006, Ageia introduced a physics processing unit (PPU) as additional hardware for the PC system, dedicated to the simulation of real-world physics within computer games (AGEIA Corp., 2006). Although this hardware seemed potentially interesting to perform sound simulations as well, a first attempt of implementing ray- and wave-based sound simulations failed due to a less efficient hardware design and a game-centered API that only permits an implementation of *game physics* (Hugenberg, 2007). However, the availability of dedicated physics hardware is intriguing and the hardware might evolve in the same direction as computer graphics hardware has a decade ago. Until then, advantages of employing physics hardware in sound-based simulations do not exist.

Augmented Audio Reality

A large part of the research was directed to an implementation of 3D spatial interaction techniques and the design of an augmented audio reality system. An implementation of augmented reality is only possible using additional, generally very expensive, hardware for user-tracking and positioning. The proposed system that was conceptualized within this research is based on a low-cost approach and only employs commodity hardware that is broadly available. An evaluation revealed that the developed system worked very well, but the WiFi-based user positioning turned out to be a weak point of the implementation. Without the need to resort to expensive hardware, which no doubt would remedy all problems, additional sensors and technologies, such as Bluetooth, can be integrated to advance the system.

More and more everyday devices, such as mobile game consoles, PDAs and telephones, feature additional sensors that can be easily employed in the direction of an audio-centered ubiquitous computing and used in an ambient intelligence design. Auditory user interfaces and auditory display systems will no doubt play a larger role in the near future. A lot of the information that is conveyed visually today might be perceived and displayed acoustically tomorrow.

And Beyond...

Beyond all these technical questions, the perspectives and future possibilities of the applications and user scenarios discussed are very intriguing. In a web article, Kennedy wrote about the *"Bubbles of Sound in Public Space"* (Kennedy, 2007), see also Figure 91. These bubbles emerge through the large availability of mobile MP3 players, which immerse their listeners into their own personal auditory environment:

> *"... the music they're listening to (...) becomes a kind of background soundtrack to the experience of public space. Public space, or the urban environment, has become a kind of background scenery to the music we listen to rather than the primary focus of our experience." (Mosco, 2005)*

This alludes to a discussion of the social effects of music. People who are shutting each other off by plugging their ears in public spaces are achieving the opposite of what music was initially created for – a means of communication and to bring people together. Therefore, a more philosophical challenge is the question of how to unite and connect different people over sound and music together, rather than immersing themselves in their own secluded worlds. A huge influence is here the application itself and the interaction it allows, as well as the content that is displayed.

Leaving this last statement unanswered, this concludes this chapter and the research in this thesis that hopefully provides a strong and firm basis for further research and future developments.

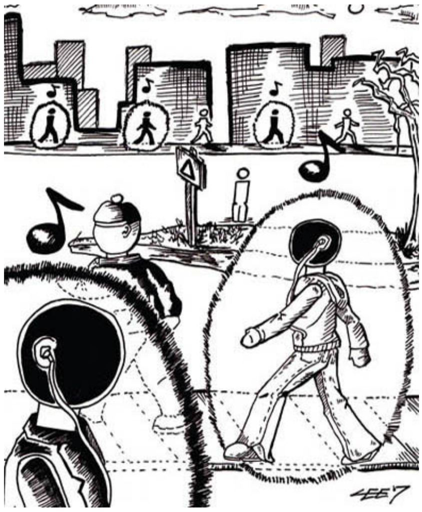


Figure 91: *"Bubbles of Sound in Public Space"* by Dave Lee (Kennedy, 2007).

APPENDIX

A
SHADER CODE EXAMPLES

The examples that are listed in this section emphasize the in Chapter 8 discussed techniques for a more efficient and GPU-based sound rendering and acoustic simulation. The here presented examples are, however, simplified and represent only the *core* of the techniques and algorithms discussed.

A.1 GPU-BASED SOUND SIGNAL PROCESSING

This section explains and discusses several example shaders that are related to Section 8.2, which examined the possibilities for a GPU-accelerated (sound) signal processing. The first example in Listing A.1 shows a resampling routine for sound signals, which also allows to perform a time-scaling of the input data. The signal is stored within a 2D texture with *texSize* being the dimension of this texture, which is passed in as an external parameter. The texture itself is a 32 bit floating point texture and contains the uncompressed sound data. Using a *scale* factor, the new coordinates for texture access are determined and returned: *return coords*, and employed in a second step to perform the actual *sampling*.

```
static const float step     = 1.0 / (texSize - 1);
static const float rowSize  = texSize * step;
static const float pix_size = 1.0 / texSize;

float2 sample(float2 coords, float index)
{
  float scale = 1.0;
  float scaleY = floor(coords.y / pix_size);
  coords.x += scaleY + (index * pix_size);
  coords.x *= scale;
  coords.y = 0.0;

  if (coords.x > 1.0) {
    int up = floor(coords.x);
    coords.x = coords.x - float(up);
    coords.y = coords.y + (up * pix_size);
  }
  if (coords.x < 0.0) {
    int down = abs(floor(coords.x));
    coords.x = coords.x + float(down);
    coords.y = coords.y - (down * pix_size);
  }
  return coords;
}
```

Listing A.1: GPU Signal Processing (Sampling).

A similar sampling is also employed in the example shown in Listing A.2. It displays a small code fragment for creating a chorus-like sound effect. In this example, the original sample is blended with two time-shifted copies of the original signal. To create a chorus effect, an additional *sin* is used depending on the sample's position. This *sweep* factor is

```
float4 chorus(float2 coords : TEX0, uniform sampler2D texture) : COLOR
{
  float sweep = sin(((coords.x/pix_size) + (coords.y*texSize)) / 12000.0);

  float4 s1 = tex2D(texture, coords);
  float4 s2 = tex2D(texture, sample(coords, (-440 - (150 * sweep))));
  float4 s3 = tex2D(texture, sample(coords, (-500 - (200 * sweep))));

  return ((s1+s2+s3)/3.0);
}
```

Listing A.2: GPU Signal Processing (Chorus).

then used to determine the texture coordinates for the actual sampling and the final blending of all three signals: *((s1+s2+s3)/3.0)*.

A last example for a GPU-based signal processing is shown in Listing A.3, and shows an implementation of an equalizer using an additional set of pre-computed bandpass filters. Depending on the filter's center frequency, a certain frequency range of the signal is enhanced or suppressed by the factor *gain*. Although all these examples demonstrate basic signal processing techniques only, more complex filter kernels that support a gathering and scattering of (sound) signal data can easily be realized (Harris, 2005).

```
const uniform float gain;
const uniform float f_center;
const uniform float k_length;

float4 eq_filter(float2 coords : TEX0, uniform sampler2D texture : TEXUNIT0, uniform
     samplerRECT filter :  TEXUNIT1) : COLOR
{
  float4 s1 = tex2D(texture, coords);
  float tmp = 0.0;
  float2 f_coords;
  f_coords.y = f_center;

  for (int i=-k_length ; i<k_length ; i++) {
    f_coords.x = (i+k_length);
    tmp += (texRECT(filter, f_coords)) * (0.5-tex2D(texture, sample(coords, i)).r);
  }

  s1.r = ((s1.r + (gain*tmp)));
  return s1;
}
```

Listing A.3: GPU Signal Processing (Equalizer).

A.2 GPU-BASED 3D WAVEGUIDE MESHES

Section 8.3 discussed 3D waveguide meshes and their application for acoustic simulations. Shader code that implements this technique efficiently in graphics hardware is shown in Listing A.4. It displays the fragment shader for the BCC lattice with phase-reversing reflections enabled at walls and ceilings. It also shows the implementation of a *sound source* in lines 39 – 43, which excites the mesh time-controlled using a given impulse. The first lines 1 – 5 setup several different variables and provide access to the 3D texture, while lines 8 – 11 determine the position of the current node within the texture. The computations of the base grid are shown in lines 14 – 32, while the computations of the offset grid at the cell center are omitted. Their implementation is almost identical to the ones shown for the base grid. Lines 39 – 43 show the excitation of a node using a provided

```
1  uniform float      layer;
2  uniform vec3       step;
3  uniform vec4       impulse;
4  uniform sampler3D  myTexture;
5
6  void main (void)
7  {
8    vec3 stepX = vec3(step.x,0,0);
9    vec3 stepY = vec3(0,step.y,0);
10   vec3 stepZ = vec3(0,0,step.z);
11   vec3 pos   = vec3(gl_TexCoord[0].xy, layer);
12
13   // --- base grid ---- with red=t and green=(t-1) ------------------
14   vec3 posLeft     = pos - stepX;
15   vec3 posLeftDown = posLeft - stepY;
16   vec3 posDown     = pos - stepY;
17
18   vec4 center1   = texture3D(myTexture, pos);
19   vec4 left1     = texture3D(myTexture, posLeft);
20   vec4 leftDown1 = texture3D(myTexture, posLeftDown);
21   vec4 down1     = texture3D(myTexture, posDown);
22
23   vec3 pos0      = pos - stepZ;
24   vec4 center0   = texture3D(myTexture, pos0);
25   vec4 left0     = texture3D(myTexture, posLeft-stepZ);
26   vec4 leftDown0 = texture3D(myTexture, posLeftDown-stepZ);
27   vec4 down0     = texture3D(myTexture, posDown-stepZ);
28
29   float baseGrid = center0.b + left0.b + leftDown0.b + down0.b;
30   baseGrid += center1.b + left1.b + leftDown1.b + down1.b;
31   baseGrid *= 0.25;
32   baseGrid -= center1.g;
33
34   // --- phase shifted grid ---- with blue=t and alpha=(t-1) --------
35
36   shiftGrid = ...
37
38   if (abs(pos.x-impulse.x)<step.x/2.0 &&
39       abs(pos.y-impulse.y)<step.y/2.0 &&
40       abs(pos.z-impulse.z)<step.z/2.0)
41         center1.r += impulse.a;
42
43   if (wall.g >= 8.0)
44     gl_FragColor = vec4(0.0, center1.r, 0.0, center1.b);
45   else
46     gl_FragColor = vec4(baseGrid, center1.r, shiftGrid, center1.b);
47 }
```

Listing A.4: Waveguide Fragment Shader (BCC Lattice).

impulse, while lines 45 – 50 perform the final computations and update the different time frames of the waveguide data according to the principles discussed, refer to Figure 60.

A.3 GPU-BASED RAY ACOUSTIC SIMULATIONS

Additionally, a graphics-based implementation of an acoustic ray tracing system was introduced and discussed in Section 8.4. Listing A.5 shows here the main fragment shader that controls the acoustic ray tracing. First, a sound source is initialized, approximated as a sphere and positioned. Using the current direction and position, several rays are initialized with the actual ray tracing being performed individually and depending on the ray's type. The ray tracing itself along with the intersection tests and the computation of

```
fragOut main( float2 texCoord : TEXCOORD0,
              uniform float3 direction,
              uniform float3 position,
              uniform float3 right,
              uniform float3 soundPos,
              uniform float type )
{
  // init sphere
  sphere s;
  s.position = soundPos;
  s.radius = distance( position, soundPos ) / 16;

  // init ray
  ray r;
  r.origin = position;
  r.len = 0.0;
  r.direction = float3( texCoord.xy, -1.0 );
  r.direction = normalize( r.direction );
  r.direction = mul( getModelViewMatrix( direction, right ), r.direction );
  r.color = createWhiteSpectrum();

  // perform  ray tracing
  fragOut retValue;
  if ( type == REFLECTION )
    retValue = reflection( r, s );
  else if ( type == REFRACTION )
    retValue = refraction( r, s );
  else if ( type == DIFFRACTION )
    retValue = diffraction( r, direction, right, s );

  return retValue;
}
```

Listing A.5: Ray Acoustics Fragment Shader.

the final acoustic energy is performed in separate shader files. At the end, all acoustic energy is accumulated and stored within a 2D texture with two buffers, which is read back to main memory and played back as a native binaural OpenAL stereo sound buffer.

B USER EVALUATIONS AND QUESTIONNAIRES

One of the research's main concern was an evaluation of the developed sonification and interaction techniques to explore and interact with 3D virtual auditory environments. Throughout this research, several tools, techniques and applications have been prototypically implemented and tested. The evaluation of the developed techniques were performed with user studies to examine the implementations more closely. In these evaluations, participants were asked to complete certain predefined tasks. These tests were accompanied by a set of research questionnaires, which had to be answered and filled out before and after the tests.

Chapter 9 discussed all evaluations, as well as presented their major conclusions. The following section summarizes each evaluation in more detail and discusses their initial hypotheses as well as their final conclusions. The evaluations were performed within two user studies, a combined multi-evaluation study and a separate analysis of the augmented audio reality system:

- Combined evaluation study
  - 2D/3D Data- and volume sonification techniques
  - 3D Scene sonification and interaction techniques
  - Evaluation and comparison of audio-only computer games
  - Evaluation of bone-conducting headphones
  - Evaluation of interactive audiobooks
- Augmented audio reality – system and application
  - Pathfinding and -following
  - Augmented audio entertainment (AAR game)

The first five studies were evaluated together as part of a multi user-evaluation, while the last two were setup in the Cathedral of Magdeburg to examine two location-based augmented audio reality implementations. All questionnaires, as well as the SPSS data files can be found on the accompanying DVD, see also Section B.7. As the evaluations were performed in German, the questionnaires and the SPSS evaluations are in German as well. This section, however, translates the most important findings and results to make them more accessible to a broader audience.

Questionnaire "General and demographic Information".

The multi evaluation study started with a general questionnaire to gather common user information and to collect demographic data. A total number of 26 users (23 male, 3 female) participated in these studies, of which 3 had a visual impairment and one user had a slight hearing disability. The distribution in age ranged from 20 to 49, and all participants had at least a high school degree (Abitur).

Analysis of Questionnaires.

Almost all participants had no prior experiences with auditory display systems or audio-only computer games. This makes these evaluations even more interesting, as it allows an assessment of the developed techniques using an unbiased audience. However, around 60% of the participants had a fundamental knowledge of 3D interaction and are playing computer games on a regular basis. This knowledge makes an interaction with 3D virtual auditory environments, as well as the performance of the required tasks much more easy.

B.1 2D/3D DATA- AND VOLUME SONIFICATION TECHNIQUES

This first section discusses the evaluation of the in Section 5.3.1 developed 2D/3D data and volume sonification techniques. The main results of this user evaluation were presented in Section 9.2 and have been discussed together with several examples. The goal of this evaluation was to assess the potential and the functionality of the developed techniques, as well as to determine the efficiency of their application. Prior to the user analysis, several hypotheses have been postulated to focus on and examine specific parts more closely:

- Sonification techniques are sufficient to acoustically display simple 2D/3D data sets
- Some techniques (sound spatialization, rhythm, melodies etc.) enhance the perception and allow a finer stream segregation
- A combined audio/visual examination of data sets is more efficient and thorough than a clean graphics-based data visualization
- Spatial interaction techniques thereby greatly improve the understanding of the data set and its topology

Questionnaire "Data and Volume Sonification".

Analysis of Questionnaires.

The questionnaire to analyze and proove/disproove these hypotheses was grouped into five sections, in which each section was further divided into individual tasks:

- Classification of melody and rhythm
- Evaluation of stock data sonification techniques
- 2D Shape sonification
- 3D Object sonification
- 3D Volumetric data sonification

The first task was based on a personal rating of four different melodies to gather information of how one perceives, appreciates and interprets different melodic rhythms. The second task included a listening to three different stock sonifications, in which the participants had to identify the correct number of parallel stock data sonifications, as well as had to draw an estimate of the ascending and descending stock graph figures. Although the last example that was sonified was also the most complex, yet it utilized two techniques that allowed a better segregation of parallel sound streams. Therefore, most participants performed here at

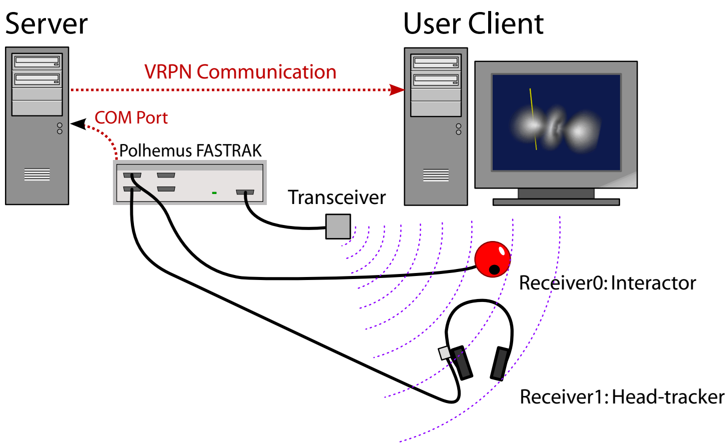


Figure 92: User Evaluation Setup.

best. The third and the fourth task were centered around the sonification of various 2D shapes and 3D objects that the participants had to identify correctly. The last task comprised the sonification of three volumetric data sets, in which the listeners had to visually identify the data set, draw an estimate of its density distribution, as well as assess the benefits of a combined audio/visual data visualization and sonification.

The setup for this evaluation was as follows, and is illustrated in Figure 92:

- Three desktop computer systems:
  - One computer for the evaluation of the 2D sonification techniques
  - One computer for the evaluation of the 3D sonification techniques
  - One control computer for the tracking system
- Two regular HiFi headphone systems
- One tracking system (Polhemus FASTRAK plus the 3Ball sensor)

The conclusions to be drawn from this evaluation are that all sonification techniques performed even better than anticipated, and that all initial hypotheses could be confirmed. A total of 15 participants (14 male, 1 female) were involved in this evaluation, of which two had a vision impairment and one a slight hearing deficiency. More details can be found in the analysis of the questionnaires. The evaluation of the stock market data sonifications clearly show a much better performance of the participants through the added sound spatialization, especially when combined with an additional rhythmic sequencing. The sonification of 2D shapes and 3D objects performed well as well, although some shapes/objects had a similar auditory resemblance and were sometimes misinterpreted (eg. sphere/cylinder). The sonification of volumetric data sets proved that even more difficult volumes can here be identified correctly. An added spatial sonification and exploration allows thereby a better understanding of the data's inherent topology. A combined data sonification/visualization achieved overall the best performance.

B.2 3D SCENE SONIFICATION AND INTERACTION

The second user study examined several of the in Section 5.3.2 and Section 5.4 developed 3D scene sonification and interaction techniques. The evaluation was previously discussed in Section 9.3, which also presented the major results and conclusions. The goal was to assess the functionality and applicability of the devised techniques and to examine the performance of users in an exploration of 3D virtual auditory environments. The postulated hypotheses for this evaluation are:

- An orientation, navigation and exploration in 3D virtual auditory environments is easily possible with adequate 3D scene sonification and interaction techniques
- A selective listening (auditory lens) allows a better perception and understanding of the environment
- Head-tracking and sound spatialization improve perception and navigation
- Speech analysis and synthesis are both only partially applicable
- The interaction with a 3D ring based menu system can be performed trough
  - Earcons and/or speech for information sonification
  - 3D Gestures and standard (gamepad) interactions

The evaluation and questionnaire to examine these questions were grouped into five sections:

- 3D Scene – navigation and orientation
- Selective 3D scene sonification – examination of the auditory lens
- Navigation and pathfinding through a complex 3D auditory environment
- Speech-based 3D scene interaction and sonification
- Interaction with a 3D auditory ring menu system

Questionnaire "3D Scene Sonification".

Analysis of Questionnaires.

Each section thereby concentrated on a specific task that was explained and demonstrated using a short example prior to the evaluation. The first task utilized the sound stage, as is depicted in Figure 93, in which the participants had to explore a 3D auditory environment with the techniques provided to find and activate four different sound sources. Additionally, an overview of the perceived scene topology had to be drawn. The interaction was based on 3D head-tracking and a navigation/orientation using a regular gamepad. The second task was dedicated to the selective listening approach provided by the auditory lens metaphor. In this evaluation, a more complex scene was used, which the participants had to explore utilizing the auditory lens. This lens could be switched on/off, as well as several parameters, eg. selective listening, could be adjusted. The third task employed a complex environment with 12 different sound sources, in which the user had to navigate to a specific sound source (alarm sound) without colliding with other sources. The forth task evaluated the possibilities of a speech-based interface, in which the navigation and scene interaction were mapped to speech commands. The scene sonification itself was also mapped to speech and provided by speech synthesis. In the last task, the participants were asked to interact with a 3D auditory ring-based menu system. Here the gamepad was used for interaction, as well as the Stylus to input basic 3D gestures.

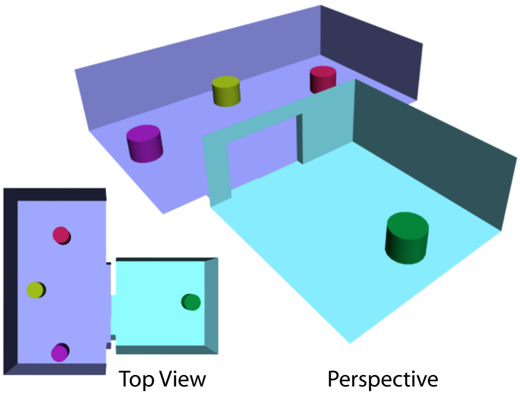


Figure 93: The Sound Stage.

The evaluation of these techniques also required 3D head-tracking and spatial interaction capabilities, and therefore, a similar setup as in the previous section, and as depicted in Figure 92, was used:

- Two desktop computer systems:
  - One computer for the evaluation of the 3D scene sonification and interaction techniques
  - One control computer for the tracking system
- One HiFi headphone system with a microphone for speech input
- One tracking system (Polhemus FASTRAK plus the Stylus sensor)
- One gamepad for regular interactions

14 users (13 male, 1 female) participated in this user evaluation, of which two had a slight visual, as well as one a slight hearing impairment. The majority of the participants ($\geqslant$ 70%) had a high familiarity with 3D interaction techniques, with a few users also being experienced with auditory displays and audio-only computer games ($\leqslant$ 20%). The majority of all users accomplished the tasks without larger difficulties, while two participants (not included in the evaluation) had considerable problems in interpreting and localizing virtual 3D sound sources. These difficulties most likely result from the use of generalized HRTFs, which are not applicable to every individual. All techniques that were evaluated, the auditory lens, as well as several methods for scene sonification and spatial interaction, have proven their applicability towards an interaction and exploration of 3D virtual auditory environments. Especially the integrated head-tracking and the spatial interaction techniques were of high assistance. A complete overview of the evaluation results can be found in the analysis of the questionnaires.

B.3 EVALUATION AND COMPARISON OF AUDIOGAMES

The audio framework devised in Chapter 5 was designed specifically with 3D audio-only computer games in mind. Section 9.4 analyzed and discussed several existing audiogames and compared them with four implementations that are based on this framework. These audiogames utilize 3D sound spatialization, as well as employ head-tracking and spatial interaction techniques. The goal of this evaluation was to explore the potential of an audio-centered gameplay and to assess the applicability of the previously evaluated 3D scene sonification and interaction techniques towards an employment in audio-only computer games. The hypotheses for this evaluation are:

- An audio-centered gameplay is more enjoyable than an adaptation of a visual genre
- Spatial interactions and 3D head-tracking improve the perception and the playability of a 3D audiogame
- Efficient and high-quality 3D sound spatializations are required
- An audio-only gameplay is highly immersive
- Audiogames can be played and enjoyed by unexperienced and sighted users as well

Questionnaire "Audiogames".

In this evaluation, several audiogames were played and compared in terms of playability, scene sonification, immersion, fun, and more. The majority of participants thereby played several of these games. The questionnaire used was divided into two sections and designed to examine each game in more detail, see also Section B.7:

Analysis of Questionnaires.

- General classification of the game
- Interaction and scene sonification

The first part was used to answer more general questions regarding the difficulty and the gameplay itself, while the second part explicitly focussed on the sonification of information and the interaction with the game environment. The participants could thereby play a title for as long as was required to be able to assess the game and to answer the questionnaire.

The play of the regular games required a standard PC only, while the four games that are based on the audio framework required additional hardware to perform spatial interactions. The setup is here similar to the previous evaluations, and as depicted in Figure 93:

- Three desktop computer systems:
  - One computer for the evaluation of regular audiogames
  - One computer for the evaluation of the audiogames that utilize 3D head-tracking and spatial interaction
  - One control computer for the tracking system
- Two regular HiFi headphone systems
- One tracking system (Polhemus FASTRAK with the Stylus sensor)
- Two gamepads for regular interaction

The conclusions to the analysis and comparison of several audiogames can be found in Section 9.4.3, which already stressed the issues for a *rethinking of audiogames* towards a more audio-centered interaction and gameplay. 13 users (12 male, 1 female) participated in this user evaluation, of which two had a slight visual and one a slight hearing impairment. Two participants had no prior experiences with computer games, while three users where also familiar with an auditory gameplay. The participants played and evaluated six games, of which the ones that are ranked highest were also played most often:

- Mosquito (played 11 times)
- Audio Frogger (played 9 times)
- The hidden Secret (played 5 times)
- Der Tag wird zur Nacht (played 5 times)
- Terraformer (played 3 times)
- Shades of Doom (played 3 times)

Interesting to note is that the games with the simplest gameplay (*Mosquito*) were also the ones that were enjoyed most. All of the proposed hypotheses could be confirmed and it was shown that 3D sound spatialization and an audio-centered gameplay are more desirable than an adaptation of audio/visual games. The speech input and output of *AudioQuake* annoyed so many participant that it was excluded from the analysis. Several titles, such as *The hidden Secret* and *Der Tag wird zur Nacht*, concentrated on a strong storyline that was appreciated by all players and found to be very immersive. More details of this evaluation, specifically to each game, can be found in the analysis of the questionnaires.

B.4 EVALUATION OF BONE-CONDUCTING HEADPHONES

The in Chapter 6 devised augmented audio reality system requires a proximaural sound presentation that allows a simultaneous perception of both, the real and the virtual auditory environment. Bone-conducting headphone systems represent here a very promising approach, as this technology does not cover the ears, and hence, does not influence the perception and localization of natural sound sources. Artificial sound sources are perceived via skin and bone, thus *providing a second pair of ears*. Section 9.5.1 presented and discussed the main results of a user evaluation to determine if *bonephones* are really applicable for the perception of 3D auditory environmental information and 3D virtual sound sources. To evaluate this, a direct comparison with regular HiFi headphones was performed. The hypotheses of this evaluation are:

- Sound perception using bone-conducting headphone systems probably causes partial impairments at certain levels of loudness and for certain frequency ranges, with
  - Expected difficulties for low loudness levels
  - Expected difficulties at very low, and very high frequency ranges
- Bone-conducting headphones can be employed for the perception of environmental acoustics and to localize virtual 3D sound sources
- Bone-conducting headphones perform qualitatively similar in the perception of speech, but overall less for music and high-quality acoustics

The task therefore was to let participants listen to both headphone systems and assess to what they have heard. The participants would start here with either the bonephones or the regular HiFi headphones and perform a series of tests and sound quality assessments:

- Perception of varying loudness levels
- Perception of varying frequency ranges
- Perception and quality assessment of different speech, music and environmental acoustic samples
- Source localization of stationary and dynamic 3D virtual sound sources

Questionnaire "Bonephone Evaluation".

Analysis of Questionnaires.

The first test presented a sine tone at different loudness levels to assess what levels of loudness are audible on both headphones. The second test was similar and employed a tone with a constant loudness, but with a varying frequency range. The third test required the participants to rate the perceived quality of several sound, speech and acoustic samples, while the forth task was centered around the localization of stationary and dynamic 3D virtual sound sources. The dynamic sound sources thereby moved along a quarter sphere, ie. from the front to left. No 3D head-tracking was employed, although this would have clearly improved the perception of the dynamic 3D sound sources.

The setup of this evaluation only required a regular desktop computer system with sound output and the possibility to connect two headphone systems:

- One desktop computer for the evaluation of both headphones
- A regular HiFi headphone system
- A bone-conducting headphone system

The test itself was implement using different Powerpoint slides, which explained each test individually, as well as presented the respective sound samples for listening, refer to Appendix C.

In this evaluation, 16 users (13 male, 3 female) participated, with two persons having a slight visual, as well as one a slight hearing impairment. The short conclusion of this evaluation is that bone-conducting headphone systems can be very well applied for the display of 3D virtual auditory environments, as well as employed within an augmented audio reality system. Although the acoustic quality perceived is slightly lower compared to regular HiFi headphones, it does not impede with the perception and localization accuracy of virtual 3D sound sources. In fact, the accuracy of the two systems was almost identical for both, stationary and dynamic 3D sound sources. However, the loudness and frequency evaluations revealed as expected a drop at lower loudness levels as well as for lower frequencies. These facts must be considered for the design of a 3D augmented auditory environment that is presented using bone-conducting headphones. More details of this evaluation can be found in the analysis of the questionnaires.

B.5 EVALUATION OF INTERACTIVE AUDIOBOOKS

Throughout the research it was assumed that an audio-only presentation permits a stronger and deeper immersion than a combined audio/visual representation of a virtual scene. Therefore, Section 9.6 developed the concept of interactive audiobooks, which combines the narrative principles and story representations from audiobooks and radio plays with interactive elements from adventure computer games. Two example stories were prototypically implemented and also evaluated using a user evaluation (Röber et al., 2006b; Huber et al., 2007). These early evaluations, however, have shown several issues within the user interface and the means to interact with the storyline. Consequentially, the concept was revised and a second prototype evaluated with the following hypotheses:

- Interactive audiobooks permit a high level of immersion
- The interface designed is intuitive and easy to use
- The non-linearity of the underlying storyline is perceived and understood
- The play of integrated minigames enhances the perception of the story, as well as increases the level of immersion
- The varying degree of interaction is perceived as seamless

Questionnaire "Interactive Audiobooks".

Analysis of Questionnaires.

The participants could choose to play the first quarter only, or the entire storyline of the interactive audiobook of *"The hidden Secret"*, see also Section 9.6. After the evaluation and the interaction with the audiobook, all participants were asked to answer several questions. The questionnaire was grouped into the following three sections:

- General perception and classification of interactive audiobooks
- Storytelling and narration
- Story interaction and play of the included minigames

The first questions were centered around the general perception of the concept and how well users could interact with the system. The second and third part focussed on specific areas of interactive audiobooks, in which part two concentrated solely on storytelling and narration and part three on the interaction with the included minigames.

The setup for this evaluation was very simple and required a desktop computer system with sound output, a good pair of headphones, as well as a gamepad for the interaction:

- One desktop computer system
- One regular HiFi headphone system
- One regular gamepad for story interaction and play of the minigames

The results of this evaluation confirm the previous analyses and further emphasize the high applicability of auditory environments for storytelling and narration (Röber et al., 2006b; Huber et al., 2007). Although only 7 users (6 male, 1 female) participated in this evaluation (due to the long stories), the results are, nevertheless, meaningful, as the findings orient themselves along those of the previous studies. All participants but one experienced the entire storyline and immersed themselves into the presented story arc. Interesting to note is that only a few participants appreciated the non-linearity of the storytelling, which is most likely due to the fact that the audiobook was only heard once. The interaction with the new designed interface also appears to be much more intuitive, as almost no one missed any of the interactive parts, which was the case with an earlier implementation (Röber et al., 2006b). More details and results can be found in the analysis of the questionnaires.

B.6 AUGMENTED AUDIO REALITY - SYSTEM AND APPLICATION

This last evaluation was performed individually and at a different location. Chapter 6 discussed in detail the concept of augmented audio reality and devised a low-cost system for its realization. The main goal of this user evaluation was a performance analysis of the entire system to determines how well the developed components operate, as well as how good the authored content is perceived. Section 9.5 already discussed and presented here the major results of this evaluation. The main research questions and hypotheses for this evaluation were:

- Evaluation and assessment of the systems overall performance
  - Accuracy of the WiFi-based user positioning
  - Orientation accuracy, ie. the performance of the 3D head-tracking
  - Efficiency and accuracy of the 3D pointing technique utilizing the gyro mouse
- Perception and experience of the auditory overlay, ie. how well are both, the virtual and the artificial, environments perceived as one?
- Expressivity, effectiveness and performance of the employed sonification and interaction techniques
- Presentation and perception of the storyline (immersion)

Questionnaire "Augmented Audio Reality".

Analysis of Questionnaires.

To evaluate the system and its application, two example scenarios have been authored using the environment discussed in Chapter 7. The chosen scenarios required a setup and evaluation of the system within the Cathedral of Magdeburg, refer also to Section 9.5.2. The evaluation itself and the questionnaire used were thereby divided and grouped into three main sections:

- General perception and classification of the augmented audio reality system
- Scenario 1 – Path-tracking and -following
- Scenario 2 – Play and interaction with the augmented audio reality game *"The hidden Secret"*

The first section assessed the general perception and performance of the augmented audio reality system, while the second and third part explicitly concentrated on the two example implementations. Scenario 1 required the tracking and following of a virtual path that was sonified acoustically, while Scenario 2 allowed a partial interaction with the story of *"The hidden Secret"*, this time in its augmented audio reality implementation and played *on location*.

The evaluation of the AAR system along with its application required the largest setup of all evaluations. The radiomap that was employed for the WiFi-based user positioning was measured in advance. The hardware that was employed in this evaluation was:

- One wearable computer system (laptop), equipped with

  - One set of bone-conducting headphones for sound presentation
  - One gamepad for regular interaction
  - One gyro mouse for 3D pointing and 3D interaction
  - One digital compass employed for 3D user head-tracking
  - One WiFi computer card equipped with an external antenna for user positioning

- Nine portable WiFi access points

An extensive analysis and discussion of the results was already provided in Section 9.5.3. A short conclusion of the here developed AAR system is that the system works very well, with an exception being the WiFi-based user positioning. A total number of 13 users (10 male, 3 female) participated in this evaluation, of which three were completely blind. The range in age was between 20 and 59 and the majority of users had no or limited experiences with 3D interactions and auditory display systems. Overall, the system and its functionality was perceived very well, but the inaccuracies of the user positioning also impaired the perception of the other functions. A special analysis tool, however, revealed that all other components performed as expected, refer to Section 9.5.3. More details on the evaluation of this study can also be found in the analysis of the questionnaires.

B.7 QUESTIONNAIRES

The following pages show exemplarily two questionnaires:

- Questionnaire "General and demographic Information" on page 189
- Questionnaire "Evaluation of Audiogames" on page 193

Questionnaire "General and demographic Information"

Fragebogen – Allgemeiner Teil

Nummer:__________

(1) Normalsichtig ☐ Sehbehindert ☐ Blind ☐

(2) Hörbeinträchtigung keine ☐ links ☐ rechts ☐

(3) Geschlecht männlich ☐ weiblich ☐

(4) Alter 10-19 ☐ 20-29 ☐ 30-39 ☐ 40-49 ☐

(5) Ausbildung / Bildungsstand ____

(6) Beruf / Tätigkeit ____

(7) Wie gestalten Sie Ihre Freizeit?

(8) Ich sehe Filme *gar nicht* ☐ ☐ ☐ ☐ ☐ *sehr oft*

(9) Ich sehe Fernsehen *gar nicht* ☐ ☐ ☐ ☐ ☐ *sehr oft*

(10) Ich höre Musik (CDs, MP3, etc.) *gar nicht* ☐ ☐ ☐ ☐ ☐ *sehr oft*

(11) Ich höre Radio *gar nicht* ☐ ☐ ☐ ☐ ☐ *sehr oft*

(12) Ich spiele Computerspiele *gar nicht* ☐ ☐ ☐ ☐ ☐ *sehr oft*

(13) Ich lese Bücher *gar nicht* ☐ ☐ ☐ ☐ ☐ *sehr oft*

(14) Ich höre Hörbücher *gar nicht* ☐ ☐ ☐ ☐ ☐ *sehr oft*

(15) Computerkenntnisse *keine* ☐ ☐ ☐ ☐ ☐ *sehr hoch*

(16) Erfahrung mit 3D Interaktion *keine* ☐ ☐ ☐ ☐ ☐ *sehr hoch*

(17) Erfahrung mit auditiven Displays *keine* ☐ ☐ ☐ ☐ ☐ *sehr hoch*

(18) Computerspielkenntnisse *keine* ☐ ☐ ☐ ☐ ☐ *sehr hoch*

(19) Erfahrung mit Audiogames *keine* ☐ ☐ ☐ ☐ ☐ *sehr hoch*

Beispiel: *nein* ☐ ☒ *ja* , entspricht **ja**

Beispiel: *sehr schlecht* ☐ ☐ ☐ ☒ ☐ *sehr gut* , entspricht **gut**

Nutzerstudien – Übersicht

2D/3D Sonifikation und Interaktion

- Allgemeine 3D Sonifikation- und Interaktionstechniken
- 2D/3D Daten und Volumen Sonifikation
- Sprachsteuerung eines Abenteuer-Computerspiels

Wahrnehmung (Bonephones)

- Sound Wahrnehmung mit normalen Kopfhörern und Knochenschall-Kopfhörern

Audiogames

- Echte 3D Audiogames
- Klassische Audiogames

Interaktive Hörbücher

- Evaluation eines *Interaktiven Hörbuches*

Was gefiel Ihnen am Besten? ____

Was gefiel Ihnen überhaupt nicht? ____

Weitere Anmerkungen, Hinweise, Probleme:

____

____

____

____

____

____

Questionnaire "Evaluation of Audiogames"

Fragebogen – Audiogames

Nummer: __________

Spiel:

- ☐ Mosquito
- ☐ Audio Frogger
- ☐ Dom Saga
- ☐ Matrix Shot
- ☐ Seus Crane – Detective for Hire
- ☐ Der Tag wird zur Nacht
- ☐ Terraformer
- ☐ Shades of Doom
- ☐ Audio Quake

Weitere Anmerkungen, Hinweise, Probleme:

____

____

____

____

____

____

Teil1: Allgemeiner Teil

(1) Persönliche Einordnung des Spiels in Genre: (Mehrfachnennung möglich)

☐ Actionspiel	☐ Geschicklichkeit	☐ Sportspiel
☐ Adventure	☐ Puzzle/Quiz/Rätsel	☐ Rennspiel
☐ Rollenspiel	☐ Ballerspiel	☐ Audiogame

(2) War das Ziel des Spiels klar?	*unklar*	☐ ☐ ☐ ☐ ☐	*klar*
(3) Wie hoch war der Schwierigkeitsgrad des Spiels?	*zu schwer*	☐ ☐ ☐ ☐ ☐	*zu leicht*
(4) Hat es Spaß gemacht?	*überhaupt nicht*	☐ ☐ ☐ ☐ ☐	*sehr viel*
(5) Persönliche Bewertung der Spielidee?	*sehr schlecht*	☐ ☐ ☐ ☐ ☐	*sehr gut*
(6) Würde man es wiederspielen oder weiterempfehlen?	*best. nicht*	☐ ☐ ☐ ☐ ☐	*auf alle Fälle*
(7) Unterhaltsamkeit	*sehr schlecht*	☐ ☐ ☐ ☐ ☐	*sehr gut*
(8) Wie ist die akustische Gestaltung der Szenen?	*sehr schlecht*	☐ ☐ ☐ ☐ ☐	*sehr gut*
(9) Wie ist die akustische Qualität allgemein?	*sehr schlecht*	☐ ☐ ☐ ☐ ☐	*sehr gut*

Teil2: Interaktion und Sonifikation

(10) Fiel es Ihnen schwer die einzelnen Geräusche zu identifizieren?

sehr schwer ☐ ☐ ☐ ☐ ☐ *sehr leicht*

(11) Konnten Sie die Geräusche lokalisieren (3D)? *sehr schwer* ☐ ☐ ☐ ☐ ☐ *sehr gut*

(12) Wie war die Präsentation von Spiel-Informationen? *s. schlecht* ☐ ☐ ☐ ☐ ☐ *sehr gut*

(13) Wussten Sie jederzeit wo Sie sich im Spiel befinden und was Sie dort machen können?

nie ☐ ☐ ☐ ☐ ☐ *immer*

(14) Gab es knifflige Situationen wo Sie nicht weiterwussten? *nie* ☐ ☐ ☐ ☐ ☐ *immer*

(15) Wie gut war die Akustik des Spiels (Simulation der Umgebung)?

sehr schlecht ☐ ☐ ☐ ☐ ☐ *sehr gut*

(16) Konnte man es gut mit den Kopfhörern spielen? *sehr schlecht* ☐ ☐ ☐ ☐ ☐ *sehr gut*

(17) Wie war die Steuerung des Spiels? *zu schwer* ☐ ☐ ☐ ☐ ☐ *zu leicht*

(18) Wie gut funktionierte die 3D Soundwahrnehmung? *s. schlecht* ☐ ☐ ☐ ☐ ☐ *sehr gut*

Fragebogen – Audiogames (Betreuer)

Nummer: __________

Es gab technische Probleme *nein* ☐ *ja* ☐

Wenn ja, welche: ____

Es gab Schwierigkeiten mit der Interaktion / Steuerung *nein* ☐ *ja* ☐

Wenn ja, welche: ____

Es gab Schwierigkeiten bei der auditiven Wahrnehmung *nein* ☐ *ja* ☐

Wenn ja, welche: ____

Fand das Spiel doof, langweilig oder schlecht *nein* ☐ *ja* ☐

Begründung?: ____

Es hat offensichtlich Spass gemacht *nein* ☐ *ja* ☐

C
EXAMPLES AND DVD CONTENT

Accompanying this thesis and its research is a DVD that contains additional sound and video examples, all related publications and presentations, the LaTeXsources of this document along with all images, as well as the SPSS data files from the respective user evaluations.

The pdf document of this thesis located in the root folder on the DVD allows a very easy and intuitive access of all examples and additional documents that are contained on the DVD. Through icons at the page border, these examples can directly be accessed and viewed from within the pdf document itself. Four different icons are available and represent *Sound* or *Video* examples, as well as link to external *Applications* and additional *Documents*.

Sound Examples.

Video Examples.

Applications.

Documents.

The majority of examples originates from within this research, while additional sound and video examples are included to exemplify related work and to clarify certain ideas and techniques. This chapter provides an overview of the accompanying DVD and references all examples available.

This section as well as the content of the DVD is structured as follows:

EXAMPLES contains 8 folders, in which each folder includes all examples referenced in one chapter.

PUBLICATIONS AND PRESENTATIONS contain all publications and presentations that were prepared throughout the research of this thesis. The publications are grouped into folders and are arranged after conference and submission. Several of these folders also include the presentations held at the respective conferences, as well as additional examples that were used.

THESIS FILES includes all images and the complete LaTeXsource code that was used to prepare and compile this document.

USER EVALUATION SPSS DATA FILES contains the questionnaires and the SPSS data files from the various user evaluations, as well as a link to download a free available version of the *SPSS Legacy Viewer* to view the results of the user evaluations.

C.1 THESIS EXAMPLES

The thesis is accompanied by several sound and video examples that exemplify existing and related work, as well as demonstrate the implemented techniques and prototypes. This section is used to organize all examples available and to provide references for *external* examples. The sound examples are encoded as either PCM wav files or MP3, while all videos are encoded using a DivX[1] or Xvid[2] codec.

All files and executable applications that are contained on the DVD are scanned for viruses and guaranteed to be virus free. When listening to the examples, make sure that no environmental effects, such as added acoustic simulations or the like, are active on your sound hardware. Most examples are best listened through headphones, while some (marked) require a presentation using a stereo speaker setting.

1 http://www.divx.com/
2 http://www.xvid.org/

Chapter1

Chapter 1 introduced the topic of this thesis and motivated the research using several examples of existing work.

- A video of playing the Theremin. (http://www.youtube.com/watch?v=d5EzKtn2ARE)
- A video demonstrating Aureal's Wavetracing technology.
- A video showing the ReacTABLE* system in action. (http://www.youtube.com/watch?v=0h-RhyopUmc)
- A video demonstrating the graphics-based 3D waveguide implementation from this research.

Chapter3

While Chapter 2 illuminated the research from a very abstract and theoretical perspective, no additional examples are required for this part. The following Chapter 3 introduced the fundamentals of this research and discussed many aspects of the thesis, ranging from sound perception and sound signal processing to music-centered and audio-only computer games. Several additional sound and video files are available, as well as several demonstrations of audio-only computer games. The examples that are required to be presented and perceived using headphones are additionally marked.

- A sound example for a scale illusion (listen through headphones). (http://philomel.com/musical_illusions/example_scale_illusion.php)
- A sound example for the perception of phantom words (listen through stereo speakers). (http://philomel.com/phantom_words/example_phantom_words.php)
- A sound example for binaural beats (listen through headphones). (http://www.bwgen.com/)
- A 3D sound example (listen through headphones). (http://youtube.com/watch?v=IUDTlvagjJA)
- An environmental acoustics example (listen through headphones). (http://www.soundman.de/deutsch/german.htm)
- A speech synthesis example. (http://www.research.att.com/~ttsweb/tts/demo.php)
- A video visualizing the propagation of sound waves.
- A video trailer of the computer game *Silent Hill Origins*. (http://www.gametrailers.com/player/21634.html)
- A video of the music-centered computer game *REZ*. (http://youtube.com/watch?v=A4EFNWe4mCc)
- A video of the computer game *Metronome*. (http://www.gamespot.com/pc/adventure/metronome/)
- A sound demo of the audio-only computer game *Shades of Doom*. (http://www.gmagames.com/sod.html)
- A sound demo of the audio-only computer game *Terraformers*. (http://www.terraformers.nu)

- A video of the online radio play *Seuss Crane: Detective for Hire*. (http://radio-play.com/seuss/crane)
- A sound sample of the 3D audiobook of Hans Christian Andersen's *"The Nightingale"* (Andersen, 2005).
- A sound sample of the 5.1 audiobook of Jules Verne's *"Journey to the Centre of the Earth"* (Verne, 2005).

Chapter4

In continuation of the previous examples, Chapter 4 concentrated on 2D and 3D auditory display systems and discussed several important aspects. Many of the examples used originate from a publication at the first ICAD conference (Kramer, 1994), while other related and important examples are included as well.

- Seven sound examples demonstrating the principles of auditory Gestalt (Williams).
- Sound examples for a sonification of stock market data (Kramer, 1992).
- Sound examples for static and dynamic beacons (Kramer, 1992).
- Sound examples demonstrating the effectiveness of a 3D auditory display using the application of an air traffic collision avoidance system (Wenzel, 1992; Begault, 1994).
- A sound sample of the *Atmospheric Weather/Works* Project by Polly. (http://andreapolli.com/studio/atmospherics)
- Sound examples demonstrating several audifications of earthquakes and nuclear explosions. (Heyward, 1992).
- Sound examples of earcons for a paint application. (Brewster et al., 1992).
- A sound sample demonstrating a 3D auditory menu system. (http://icad.org/node/402)

Chapter5

The main research started in Chapter 5 with the introduction of 3D virtual auditory environments. The examples that are shown in this chapter still include several important implementations of related work, but the majority of the examples represents auralizations of 3D scenes to acoustically exemplify certain techniques and approaches. These auralizations are spatialized and created using the AM3D sound API, which requires a presentation and perception via regular HiFi headphones (AM3D A/S, 2008).

The listener in these examples is thereby placed at the same location that is indicated in the individual figures, and also rotates around the z-axis to allow a better perception of the auditory scene.

- An auralization of Figure 24b (living room).
- A video demonstrating a musically correct blending of music.
- A sound sample from the album *Pictures at an Exhibition*. (http://www.musopen.com/view.php?type=piece\&id=107)
- An auralization of Figure 26a (normal living room).

- An auralization of Figure 26b (non-realistic sound environment).
- A sound of a Geiger counter. (http://www.mineralab.com)
- A video demonstration of the vOICe system.
- A video demonstrating the sonification of 2D shapes.
- A video demonstrating the sonification of a 3D volumetric data set.
- An auralization of Figure 30b (Soundpipes).
- An auralization of sound and music guides to evoke attraction and repulsion.
- An auralization of Figure 31a (Sonar/Radar).
- An auralization of Figure 31b (Auditory Texture).
- An auralization of Figure 34a (Auditory Cursor).
- An auralization of Figure 34b (Auditory Lens).

Chapter6

Chapter 6 continued the discussions of the previous chapter and directed the research in the area of augmented audio reality. The examples in this section include a demonstration of augmented reality/virtuality, as well as exemplify the principles of augmented audio reality and explain a WiFi-based user positioning technique.

- A video showing examples of augmented reality and augmented virtuality.
- An auralization of Figure 40 (Augmented Audio).
- A video explaining Figure 41 and the functionality of a WiFi-based user positioning.

Chapter7

In Chapter 7 the focus was centered on the authoring and design of 3D virtual auditory environments. Therefore, an authoring framework along with several authoring techniques and guidelines were developed. The videos in this section show the 3D authoring environment applied to the authoring of four basic tasks using the example of an augmented audio reality game.

- A video showing the 3D authoring environment and the creation of a virtual sound source along with the specification of parameters for direction and distance attenuation.
- A video showing the authoring of two auditory textures, as well as the design of object-, time- and input-dependencies.
- A video showing the authoring of a position-dependency (for augmented audio reality).
- A video showing the creation of a 3D ring-based menu system, along with the authoring of auditory textures for the menu items and the specification of an input dependency.

Chapter8

The research in this thesis has shown that all examples require a high-quality auralization combined with a very efficient implementation. Therefore, Chapter 8 conducted additional research in the direction of a GPU-accelerated graphics-based sound rendering and simulation. The examples in this section are comprised of sounds and videos to exemplify the devised techniques and to demonstrate their efficiency.

- A video demonstrating and explaining the GPU-based sound signal processing and filtering.
- A video demonstrating the 3D waveguides implementation of the Cartesian lattice.
- A video demonstrating the 3D waveguides implementation of the BCC lattice.
- A video demonstrating reflection effects in the acoustic ray tracing system.
- A video demonstrating refraction effects in the acoustic ray tracing system.
- A video demonstrating diffraction effects in the acoustic ray tracing system.
- A sound used for the 3D waveguide technique (original sound).
- A sound used for the 3D waveguide technique (low-pass filtered).
- A sound simulation using the 3D waveguides implementation (Cartesian).
- A sound simulation using the 3D waveguides implementation (BCC).

Chapter9

Chapter 9 reviewed the research of the thesis and presented and discussed several areas of application. The examples shown here are very diverse and provide a broad overview of this research and its potential. The majority of the techniques developed in this thesis were examined by user evaluations. This section presents actual examples that were used in these evaluations, as well as shows additional sound and video files to exemplify the developed applications and prototypes.

- A video demonstrating a parallel sonification of stock market data.
- A video demonstrating the sonification of 2D shapes.
- A video showing the sonification of two 3D objects.
- A video demonstrating the sonification of 3D volumetric data sets.
- A video showing the *Sound Stage* for the evaluation of several 3D scene sonification and interaction techniques.
- A video demonstrating the various functions of the auditory lens metaphor.
- A video showing the sonification and interaction with an auditory 3D ring-based menu system.
- A video showing the navigation through a complex 3D auditory environment.
- A video demonstrating both, the soundpipes and the auditory cursor technique.

- A video demonstrating a speech-based interface with the computer game *"Day of the Tentacle"*.
- A sound demo of the audio-only computer game *Shades of Doom*. (http://www.gmagames.com/sod.html)
- A video demonstrating the audiogame *Mosquito*.
- A video demonstrating the audiogame *AudioFrogger*.
- A video demonstrating the audiogame *MatrixShot*.
- A sound demo of the audio-only computer game *Terraformers*. (http://www.terraformers.nu)
- A video of the online radio play *Seuss Crane: Detective for Hire*. (http://radio-play.com/seuss/crane)
- A video demonstrating the auditory adventure game *"The hidden Secret"*.
- Sample data for the evaluation of the bone-conducting headphones.
- A video demonstrating the campus-navigation example.
- A video showing a replay of the augmented audio reality system in the AAR game *"The hidden Secret"*.
- A video showing the interactive audiobook of *"The Pit And The Pendulum"* (Narrative Interaction).
- A video showing the interactive audiobook of *"The Pit And The Pendulum"* (Minigame Interaction).
- A sound demo of the interactive audiobook of *"The hidden Secret"*.
- A video showing the waveguides implementation applied to wavefield synthesis.
- A sound demonstration using virtual HRIR simulations (horizontal plane) (listen through headphones).
- A sound demonstration using virtual HRIR simulations (median plane) (listen through headphones).

The concluding Chapter 10 summarized the thesis and discussed interesting ideas for future research. Although several references were provided, no additional examples are found in this chapter.

C.2 PUBLICATIONS AND PRESENTATIONS

Through the years of research, several publications and articles have been prepared and written along with several presentations that were held at various conferences. Table 16 provides an overview of all publications, as well as directly links to the pdf documents and conference presentations.

Conference	Year	Title
DIGRA	2003	*Game Graphics Beyond Realism: Then, Now, and Tomorrow* (pdf/ppt)
ICAD	2004	*Interacting with Sound: An interaction Paradigm for virtual auditory Worlds* (pdf/ppt)
CGAIDE	2004	*Auditory Game Authoring: From virtual Worlds to auditory Environments* (pdf/ppt)
ICAD	2005	*Leaving the Screen: New Perspectives in Audio-only Gaming* (pdf/ppt)
GCDC	2005	*PS2 Game Development under Linux* (pdf/ppt)
DIGRA	2005	*Playing Audio-only Games: A compendium of interacting with virtual, auditory Worlds* (pdf/ppt)
Graduate Day	2006	*Interacting with Sound: Techniques for interacting with virtual auditory Environments* (pdf/ppt)
Eurographics	2006	*Enhanced Cartoon and Comic Rendering* (pdf/ppt)
TR-2006-4	2006	*HRTF Simulations through acoustic Raytracing* (pdf)
TR-2006-5	2006	*Soundpipes: A new way of Path Sonification* (pdf)
TR-2006-8	2006	*Flexible Film: Interactive Cubist-style Rendering* (pdf)
AudioMostly	2006	*Authoring of 3D virtual auditory Environments* (pdf/ppt)
AudioMostly	2006	*Composition and Arrangement Techniques for Music in Interactive Immersive Environments* (pdf/ppt)
ICMC	2006	*Waveguide-based Room Acoustics through Graphics Hardware* (pdf/ppt)
TIDSE	2006	*Interactive Audiobooks: Combining Narratives with Game Elements* (pdf/ppt)
Cost ConGAS	2007	*Interaction with Sound in auditory Computer Games* (pdf)
DAFx	2007	*Ray Acoustics using Computer Graphics Technology* (pdf/ppt)
AudioMostly	2007	*Evolution of Interactive Audiobooks* (pdf/ppt)
AudioMostly	2008	*A Musical Instrument based on 3D Data and Volume Sonification Techniques* (pdf)

Table 16: Thesis Publications and Presentations.

C.3 THESIS FILES

This thesis was prepared using LaTeX and compiled using MikTeX 2.4.1779. The *Thesis Files* folder on the DVD contains the complete source code and all necessary files to compile this thesis. The style used in this document is a slight variation of the *Classic Thesis* style developed by Miede (Miede, 2007).

The Root Folder contains the main tex file (diss.tex), as well as the style files and the individual bibtex sources. The additional subfolders are organized as follows:

APPENDIX The tex files of the appendices of the thesis.

BACKMATTER The *backmatter* of the thesis, ie. listings, bibliography, index, acronyms and declaration.

CHAPTER The main tex files for the chapters of the thesis.

FRONTMATTER The *frontmatter* of the thesis, ie. title pages, table of contents, notation, acknowledgements, abstract, dedication and publications.

IMAGES All image data used for the figures in this thesis.

C.4 QUESTIONNAIRES AND SPSS DATA FILES

Several user evaluations have been performed throughout this research to evaluate and assess the designed techniques and the implemented systems. All evaluations were based on user observations and questionnaires, but also sometimes employed user log-files to determine the user's performance and to estimate the system's functionality. The DVD contains all of the questionnaires used, as well as the SPSS data files from the evaluation. *SPSS* is a powerful data mining and statistical analysis software that is often used to analyze user evaluations (SPSS Inc., 2008). In order to inspect the SPSS data files, either a working copy of SPSS, or the *SPSS Legacy Viewer*[3] is required (SPSS Inc., 2008).

3 http://support.spss.com/Student/Utilities/SPSS/LegacyViewer/readme.html

BIBLIOGRAPHY

ARTICLES AND PROCEEDINGS

Aszódi and Czuczor(2002) – Barnabás Aszódi and Szabolcs Czuczor. Calculating 3D Sound-Field using 3D Image Synthesis and Image Processing. In *Central European Seminar on Computer Graphics*, 2002. (Cited on page 108.)

Azuma(1997) – Ronald T. Azuma. A survey of Augmented Reality. In *Presence: Teleoperators and Virtual Environments 6*, pages 355–385, 1997. (Cited on pages 79, 80, 81, 83, and 85.)

Ballas(1992) – James A. Ballas. Delivery of Information through Sound. In Gregory Kramer, editor, *Proceedings of ICAD*, pages 79–94, 1992. (Cited on page 34.)

Berndt et al.(2006) – Axel Berndt, Knut Hartmann, Niklas Röber, and Maic Masuch. Composition and Arrangement Techniques for Music in Interactive Immersive Environments. In *Audio Mostly Conference*, pages 53–59, 2006. (Cited on page 53.)

Bertram et al.(2005) – Martin Bertram, Eduard Deines, Jan Mohring, Jevgenij Jegorovs, and Hans Hagen. Phonon Tracing for Auralization and Visualization of Sound. In *IEEE Transactions on Computer Graphics and Visualization*, Minneapolis, USA, 2005. (Cited on page 109.)

Blattner et al.(1992) – Merra M. Blattner, Albert L. Papp III, and Ephraim P. Glinert. Sonic Enhancement of Two-Dimensional Graphic Displays. In Gregory Kramer, editor, *Proceedings of ICAD*, pages 447–470, 1992. (Cited on pages 40 and 41.)

Blattner et al.(1989) – M.M. Blattner, D.A. Sumikawa, and R.M. Greenberg. Earcons and Icons: Their Structure and Common Design Principles. *ACM Human-Computer Interaction*, pages 11–44, February 1989. (Cited on pages 37 and 57.)

Bölke and Gorny(1995) – Ludger Bölke and Peter Gorny. Direkte Manipulation akustischer Objekte durch blinde Rechnerbenutzer. In *Software Ergonomie, German Chapter ACM and GI Darmstadt*, pages 93–106, 1995. (Cited on pages 37 and 57.)

Boone(2001) – Marinus M. Boone. Acoustic Rendering with Wavefield Synthesis. In *ACM SIGGRAPH Campfire: Acoustic Rendering for Virtual Environments*, 2001. (Cited on pages 28, 50, 159, and 160.)

Bowman and Hodges(1997) – Doug A. Bowman and Larry F. Hodges. An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in immersive virtual Environments. In *SI3D '97: Proceedings of the 1997 symposium on Interactive 3D graphics*, pages 35–ff. ACM, 1997. (Cited on page 63.)

Brewster et al.(1992) – Stephen A. Brewster, Peter C. Wright, and Alistair D. N. Edwards. A Detailed Investigation into the Effectiveness of Earcons. In Gregory Kramer, editor, *Proceedings of ICAD*, pages 471–498, 1992. (Cited on pages 41 and 201.)

Brooks Jr.(1999) – F.P. Brooks Jr. What's Real About Virtual Reality? *IEEE Computer Graphics & Applications*, 19(9):16–27, 1999. (Cited on page 47.)

Brungart et al.(2005) – Douglas Brungart, Brian Simpson, and Alexander Kordik. The Detectability of Head-Tracker Latency in virtual Auditory Displays. In *Proceedings of 11th ICAD*, Limerick, Ireland, 2005. (Cited on page 75.)

Campos and Howard(2005) – Gulherme R. Campos and David M. Howard. On the Computational Efficiency of Different Waveguide Mesh Topologies for Room Acoustic Simulation. *IEEE Transactions on Speech and Audio Processing*, 13(5):1063–1072, 2005. (Cited on pages 114 and 116.)

Caudell and Mizell(1992) – T.P. Caudell and D.W. Mizell. Augmented Reality: An Application of Heads-up Display Technology to manual Manufacturing Processes. In *International Conference on Systems Sciences*, pages 659–669, 1992. (Cited on page 79.)

Ciger et al.(2003) – Jan Ciger, Mario Gutierrez, Frederic Vexo, and Daniel Thalmann. The magic Wand. In *SCCG '03: Proceedings of the 19th Spring Conference on Computer Graphics*. ACM, 2003. (Cited on page 63.)

Cohen et al.(1993) – M. Cohen, S. Aoki, and N. Koizumi. Augmented Audio Reality: Telepresence/VR hybrid acoustic Environments. In *Proceedings of 2nd IEEE International Workshop on Robot and Human Communication*, pages 361–364, 1993. (Cited on pages 79, 81, and 83.)

Cohen(1994) – Michael Cohen. Augmented Audio Reality: Design for a Spatial Sound GPS PGS. In *Proceedings of 2nd IEEE International Workshop on Robot and Human Communication*, 1994. (Cited on pages 81, 82, and 84.)

Cohern(1992) – Jonathan Cohern. Monitoring Background Activities. In Gregory Kramer, editor, *Proceedings of ICAD*, pages 499–532, 1992. (Cited on pages 37, 39, and 233.)

Coomans and Timmermans(1997) – M.K.D. Coomans and H.J.P. Timmermans. Towards a Taxonomy of Virtual Reality User Interfaces. In *Proceedings of International Conference on Information Visualization*, 1997. (Cited on pages 46 and 47.)

Cooper and Taylor(1998) – M. Cooper and M. Taylor. Ambisonic Sound in virtual Environments and Applications for the blind People. In *Proceedings of the Second European Conference on Disability, Virtual Reality, and Associated Technologies*, pages 113–118, 1998. (Cited on page 108.)

Crawford(2003) – C. Crawford. Assumption Underlying the Erasmatron Interactive Storytelling Engine. In M. Mateas and P. Sengers, editors, *Narrative Intelligence*. John Benjamins, Amsterdam, 2003. (Cited on pages 18, 47, and 55.)

Crispien and Fellbaum(1996) – Kai Crispien and Klaus Fellbaum. A 3D-Auditory Environment for Hierarchical Navigation in Non-visual Interaction. In *ICAD*, 1996. (Cited on pages 38, 41, and 231.)

Dachselt and Hinz(2005) – R. Dachselt and M. Hinz. Three-Dimensional Widgets Revisited - Towards Future Standardization. In *Proceedings of the Workshop 'New Directions in 3D User interfaces'*, 2005. (Cited on pages 18 and 19.)

Deines et al.(2006b) – Eduard Deines, Martin Bertram, Jan Mohring, Jevgenij Jegorovs, Frank Michel, Hans Hagen, and Gregory M. Nielson. Comparative Visualization for Wave-based and Geometric Acoustics. *IEEE Transactions on Visualization and Computer Graphics*, 12(5), September/October 2006b. (Cited on page 109.)

Eckel(2001a) – G. Eckel. Audio Augmented Environments. In *Symposium on Arts and Technology*, 2001a. Conneticut College, New London, USA. (Cited on page 39.)

Eckel(2001b) – G. Eckel. Immersive audio-augmented Environments: The LISTEN Project. In *Proceedings of IEEE Information Visualization*, pages 571–573, 2001b. (Cited on page 39.)

Ekman(2007) – Inger Ekman. Sound-based Gaming for Sighted Audiences - Experiences from a Mobile Multiplayer Location Aware Game. In *Proceedings of AudioMostly Conference*, pages 148–153, 2007. (Cited on page 84.)

Feiner et al.(1993) – S. Feiner, B. MacIntyre, and D Seligmann. Knowledge-based Augmented Reality. *Communications of the ACM*, 37(7):52–62, 1993. (Cited on pages 79 and 81.)

Fitch1992 and Kramer(1992) – W. Tecumseh Fitch1992 and Gregory Kramer. Sonifying the Body Electric: Superiority of an Auditory Display in a Complex Multivariate System. In Gregory Kramer, editor, *Proceedings of ICAD*, pages 307–326, 1992. (Cited on page 39.)

Fontana and Rocchesso(2001) – F. Fontana and D. Rocchesso. Signal-Theoretic Characterization of Waveguide Mesh Geometries for Models of Two-Dimensional Wave Propagation in Elastic Media. *IEEE Transactions on Speech and Audio Processing*, 9(2): 152–161, Februar 2001. (Cited on page 114.)

Frauenberger and Noisternig(2003a) – C. Frauenberger and M. Noisternig. 3D Audio Interfaces for the Blind. In Eoin Brazil and Barbara Shinn-Cunningham, editors, *ICAD*, volume 9, pages 280–283, 2003a. (Cited on page 41.)

Freudenberg et al.(2001a) – Bert Freudenberg, Maic Masuch, Niklas Röber, and Thomas Strothotte. The Computer-Visualistik-Raum: Veritable and Inexpensive Presentation of a Virtual Reconstruction. In *VAST 2001: Virtual Reality, Archaelogy, and Cultural Heritage*, pages 97–102. ACM SIGGRAPH,, 2001a. (Cited on pages 47, 51, 80, and 231.)

Freudenberg et al.(2001b) – Bert Freudenberg, Maic Masuch, and Thomas Strothotte. Walk-Through Illustrations: Frame-Coherent Pen-and-Ink-Style in a Game Engine. In *Computer Graphics Forum: Proceedings Eurographics*, 2001b. (Cited on page 50.)

Funkhouser et al.(1999a) – T. Funkhouser, P. Min, and I. Carlbom. Real-Time Acoustic Modeling for Distributed Virtual Environments. *ACM Computer Graphics, SIGGRAPH'99 Proceedings*, pages 365–374, August 1999a. (Cited on pages 107 and 124.)

Funkhouser et al.(2002c) – Thomas Funkhouser, Nicolas Tsingos, Ingrid Carlbom, Gary Elko, Mohanand Sondhi, and James West. Modelling Sound Reflection and Diffraction in Architectural Environments with Beam Tracing. In *Forum Acusticum*, 2002c. (Cited on page 118.)

Gallo and Tsingos(2004) – Emmanuel Gallo and Nicolas Tsingos. Efficient 3D Audio Processing with the GPU. In *GP2, ACM Workshop on General Purpose Computing on Graphics Processors*, 2004. (Cited on page 110.)

Gaver(1992) – William W. Gaver. Using and Creating Auditory Icons. In Gregory Kramer, editor, *Proceedings of ICAD*, pages 417–446, 1992. (Cited on page 41.)

Gaver(1989) – W.W. Gaver. The Sonic Finder: An Interface that uses auditory Icons. *ACM Human-Computer Interaction*, pages 67–94, 1989. (Cited on pages 2, 3, and 37.)

Gaye et al.(2003) – L. Gaye, R. Maze, and L. E Holmquist. Sonic City: The Urban Environment as a Musical Interface. In *International Conference on New Interfaces for Musical Expression*, Montreal, Canada, May 2003. (Cited on page 39.)

Gaye(2002) – Lalya Gaye. A Flexible 3D Sound System for Interactive Applications. In *Proceesings of CHI*, 2002. (Cited on pages 67 and 69.)

Globus and Raible(1992) – A. Globus and E. Raible. *13 Ways to say Nothing with scientific Visualization*. Technical Report RNR-92-006, NASA Ames Research Center, 1992. (Cited on page 16.)

Gräfe et al.(2007) – Andreas Gräfe, Martin Dausel, and Andreas Franck. Wave Field Synthesis for Games using the OpenAL Interface. In *Audio Mostly Conference*, pages 116–120, 2007. (Cited on pages 28 and 160.)

Grossman et al.(2004) – T. Grossman, D. Wigdor, and R. Balakrishnan. Multi-Finger Gestural Interaction with 3D Volumetric Displays. In *Proceedings of UIST, ACM Symposium on User Interface Software and Technology*, pages 61–70, 2004. (Cited on page 18.)

Grossman et al.(2007) – T. Grossman, D. Wigdor, and R. Balakrishnan. Exploring and Reducing the Effects of Orientation on Text Readability in Volumetric Displays. In *Proceedings of CHI 2007*, pages 483–492, 2007. (Cited on page 18.)

Hand(1997) – Chris Hand. A Survey of 3D Interaction Techniques. *Computer Graphics Forum*, 16(5):269–281, 1997. (Cited on pages 18 and 19.)

Harris(2005) – Mark Harris. Mapping Computational Concepts to GPUs. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Courses*, page 50, 2005. (Cited on page 176.)

Hartmann et al.(2005b) – K. Hartmann, S. Hartmann, and M. Feustel. Motif Definition and Classification to Structure Non-Linear Plots and to Control the Narrative Flow in Interactive Dramas. In *3rd International Conference on Virtual Storytelling*, pages 158–167, 2005b. (Cited on pages 154 and 155.)

Heyward(1992) – Chris Heyward. Listening to the Earth Sing. In Gregory Kramer, editor, *Proceedings of ICAD*, pages 369–404, 1992. (Cited on pages 37, 40, 201, and 231.)

Hoffmann et al.(2003) – M. Hoffmann, R. Dachselt, and K. Meißner. An Independent Declarative 3D Audio Format on the Basis of XML. In *Proceedings of ICAD*, Boston, MA, USA, 2003. (Cited on pages 73, 74, 75, and 233.)

Hofman et al.(1998) – P. M. Hofman, J.G.A. Van Riswick, and A.J Van Opstal. Relearning Sound Localisation with new Ears. *Nature Neuroscience*, 1(5):417–421, September 1998. (Cited on pages 22 and 107.)

Holland and Morse(2001) – S. Holland and D.R. Morse. AudioGPS – Spatial Audio in a minimal Attention Interface. In *3rd International Workshop on Human Computer Interaction with mobile Devices*, 2001. (Cited on page 39.)

Hoorn et al.(2003) – J. F. Hoorn, E. A. Konijn, and G. C. Van der Veer. Virtual Reality: Do not augment Realism, augment Relevance. *Upgrade - Human-Computer Interaction: Overcoming Barriers*, 4(1):18–26, 2003. (Cited on page 43.)

Härmä et al.(2003) – a. Härmä, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, H. Nironen, and S. Vesa. Techniques and Applications of wearable augmented Reality Audio. In *the 114th Convention of the Audio Engineering Society*, 2003. (Cited on pages 82, 84, and 85.)

Huber et al.(2007) – Cornelius Huber, Niklas Röber, Knut Hartmann, and Maic Masuch. Evolution of Interative Audiobooks. In *Audio Mostly Conference*, pages 166–167, 2007. (Cited on pages 48, 75, 139, 155, 157, 167, and 186.)

Ise and Otani(2002) – Shiro Ise and Makoto Otani. Real Time Calculation of the Head Related Transfer Function based on the Boundary Element Method. In *Internation Confererence on Auditory Display*, Kyoto, Japan, 2002. (Cited on page 107.)

Ishii et al.(1998) – H. Ishii, C. Wisneski, S. Brave, A. Dahley, M. Gorbet, B. Ullmer, and P. Yarin. ambientRoom – Integrating Ambient Media with Arcitectural Space. In *Proceedings of CHI'98, ACM Press*, pages 18–23, 1998. (Cited on pages 39 and 41.)

Ivanov and Schemmer(2007) – Svilen Ivanov and Stefan Schemmer. Lokalisierung im WLAN - Neue Möglichkeiten und ihre Grenzen. In *Proceedings of 9th Conference on Wirless Technologies*, 2007. (Cited on page 85.)

Janata and Childs(2004) – Petr Janata and Edward Childs. Marketbuzz: Sonification of real-time Financial Data. In *Proceedings of ICAD*, 2004. (Cited on pages 39 and 57.)

Jedrzejewski(2004) – Marcin Jedrzejewski. Computation of Room Acoustics Using Programmable Video Hardware. In *International Conference on Computer Vision and Graphics*, 2004. (Cited on pages 108, 109, and 118.)

Kahana et al.(1999) – Y. Kahana, P.A. Nelson, M. Petyt, and S. Choi. Numerical Modeling of the Transfer Functions of a Dummy-Head and of the external Ear. In *Audio Engineering Society, 16th International Conference*, pages 330–345, Rovaneimi, 1999. (Cited on page 107.)

Kaltenbrunner et al.(2006) – S. Kaltenbrunner, M. an Jordà, G. Geiger, and M. Alonso. The reacTable*: A Collaborative Musical Instrument. In *Proceedings of the Workshop on Tangible Interaction in Collaborative Environments (TICE), at the 15th International IEEE Workshops on Enabling Technologies (WETICE 2006)*, Manchester, U.K., 2006. (Cited on pages 2 and 3.)

Kanno et al.(2006) – K. Kanno, N. Fernando, A. Bolhassan, S. Narita, and M. Cohen. Personal Practically Panoramic: Multimodal Interfaces. In *International Conference on Virtual Reality*, page 322, 2006. (Cited on page 82.)

Kapralos et al.(2004) – B. Kapralos, M. Jenkin, and E. Milios. Sonel Mapping: Acoustic Modeling utilizing an acoustic Version of Photon Mapping. In *IEEE International Workshop on Haptic Audio Visual Environments and their Applications*, 2004. (Cited on page 109.)

Kehoe and Pitt(2006) – A. Kehoe and I.J. Pitt. Using Speech Technology to Provide Help in Games. In *Fun 'n Games Conference*, Preston, England, 2006. (Cited on page 23.)

Kiegler and Moffat(2006) – Katarina Kiegler and David C. Moffat. Investigating the Effects of Music on Emotions in Games. In *Proceedings of AudioMostly Conference*, pages 37–41, 2006. (Cited on pages 27 and 60.)

Kobayashi and Schmandt(1997) – M. Kobayashi and C. Schmandt. Dynamic SoundScape: Mapping Time to Space for Audio Browsing. In *Proceedings of CHI'98, ACM Press*, pages 194–201, 1997. (Cited on pages 39 and 41.)

Kramer and Ellison(1991) – G. Kramer and S. Ellison. Audification: The use of Sound to Display Multivariate Data. In *Proceedings of ICMC*, pages 214–221, 1991. (Cited on page 40.)

Kramer(1992) – Gregory Kramer. Some Organizing Principles for Representating Data with Sound. In Gregory Kramer, editor, *Proceedings of ICAD*, pages 185–222, 1992. (Cited on pages 31, 34, 35, 37, 39, 41, 201, and 231.)

Lane et al.(1998) – J.D. Lane, S.J. Kasian, J.E. Owens, and G.R. Marsh. Binaural auditory Beats affect Vigilance Performance and Mood. *Journal of Physiolocial Behavior*, 63(2): 249–252, 1998. (Cited on page 21.)

Li et al.(2004) – Zhiyun Li, Ramani Duraiswami, and Nail A. Gumerov. Capture and Recreation of higher Order 3D Sound Fields via Reciprocity. In *ICAD*, 2004. (Cited on pages 120 and 121.)

Lindeman et al.(2008) – R.W. Lindeman, H. Noma, and P.G. de Barros. An Empirical Study of Hear-Through Augmented Reality: Using Bone Conduction to Deliver Spatialized Audio. In *IEEE Virtual Reality*, pages 35–42, 2008. (Cited on page 146.)

Lodha et al.(1997) – Suresh K. Lodha, John Beahan, Travis Heppe, Abigail Joseph, and Brett Zane-Ulman. MUSE: A Musical Data Sonification Toolkit. In *Proceedings of ICAD 97*, 1997. (Cited on page 51.)

Ludowici(2000) – Babette Ludowici. Ottonische Aula Regia oder unbekannter Kirchenbau? Ein Arbeitsbericht zum Stand der Auswertung der Grabungen von 1959-1968 auf dem Magdeburger Domplatz. *Archäologisches Korrespondenzblatt*, 30(3), 2000. (Cited on page 51.)

Madhyastha and Reed(1992) – T.M. Madhyastha and D.A. Reed. A Framework for Sonification Design. In Gregory Kramer, editor, *Proceedings of ICAD*, pages 267–290, 1992. (Cited on page 40.)

Mariette(2006) – Nick Mariette. Perceptual Evaluation of Spatial Audio for "Audio Nomad" Augmented Reality Artworks. In *Proceedings Engage*, 2006. (Cited on page 84.)

Mariette(2007a) – Nick Mariette. From Backpack to Handheld: The Recent Trajectory of Personal Location Aware Spatial Audio. In *Proceedings of 2007 Perth Digital Arts Conference*, 2007a. (Cited on page 84.)

Mayer-Kress et al.(1992) – G. Mayer-Kress, R. Bargar, and I. Choi. Musical Structures in in Data from Chaotic Attractors. In Gregory Kramer, editor, *Proceedings of ICAD*, pages 341–368, 1992. (Cited on page 40.)

McCabe and Rangwalla(1992) – Kevin McCabe and Akil Rangwalla. Auditory display of computational fluid dynamics data. In Gregory Kramer, editor, *Proceedings of ICAD*, pages 327–340, 1992. (Cited on page 40.)

Meckseper(1986) – Cord Meckseper. Das Palatium Ottos des Großen in Magdeburg. *Burgen und Schlösser*, 27, 1986. (Cited on page 51.)

Meijer(1992) – Peter B.L. Meijer. An Experimental System for Auditory Image Representations. *IEEE Transactions on Biomedical Engineering*, 39(2):112–121, 1992. (Cited on page 57.)

Micea et al.(2001) – Mihai V. Micea, Mircea Stratulat, Dan Ardelean, and Daniel Aioanei. Implementing Professional Audio Effects with DSPs. *Transactions on Automatic Control and Computer Science*, 40(60):55–60, 2001. (Cited on page 111.)

Milgram et al.(1994) – P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. Augmented Reality: A class of Displays on the Reality-Virtuality Continuum. In *SPIE Proceedings of Telemanipulator and Telepresence Technologies*, volume 2351, 1994. (Cited on pages 44, 79, and 231.)

Milgram et al.(1995) – Paul Milgram, David Drascic, J. Julius J. Grodski, Anu Restogi, Shumin Zhai, and Chin Zhou. Merging real and virtual Worlds. In *IMAGINA'95*, pages 218–220, 1995. (Cited on pages 43, 80, and 81.)

Minghim and Forrest(1995) – R. Minghim and A. R. Forrest. An Illustrated Analysis of Sonification for Scientific Visualisation. In *Proceedings of the 6th conference on Visualization '95*, 1995. (Cited on page 59.)

Mischke and Scardovelli(2005) – Thilo Mischke and Simone Scardovelli. Lauschangriff – Blinde Videospieler. *GEE – Games Entertainment Education*, pages 60–64, September 2005. (Cited on page 29.)

Moeck et al.(2007) – Thomas Moeck, Nicolas Bonneel, Nicolas Tsingos, George Drettakis, Isabelle Viaud-Delmon, and David Aloza. Progressive Perceptual Audio Rendering of Complex Scenes. In *ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, pages 189–196, 2007. (Cited on page 169.)

Mynatt(1992) – Elizabeth D. Mynatt. Auditory presentation of graphical user interfaces. In Gregory Kramer, editor, *Proceedings of ICAD*, pages 533–556, 1992. (Cited on pages 37, 41, and 64.)

Nickel(1973) – Ernst Nickel. Magdeburg in karolingisch-ottonischer Zeit. *Zeitschrift für Archäologie*, 7, 1973. (Cited on page 51.)

Owens et al.(2005) – John D. Owens, David Luebke, Naga Govindaraju, Mark Harris, Jens Krüger, Aaron E. Lefohn, and Timothy J. Purcell. A Survey of General-Purpose Computation on Graphics Hardware. In *Eurographics 2005, State of the Art Reports*, pages 21–51, August 2005. (Cited on page 109.)

Paelke and Reimann(2006) – V. Paelke and C. Reimann. Authoring von Augmented-Reality Anwendungen. *Deutsche Gesellschaft für Karthographie; Kartographische Schriften. Aktuelle Entwicklungen in Geoinformation und Visualisierung*, 10, 2006. (Cited on page 94.)

Palomäki(2006) – Henni Palomäki. Meanings conveyed by the simple auditory rhythm. In *Proceedings of ICAD*, 2006. (Cited on page 57.)

Pollack and Ficks(1954) – I. Pollack and L. Ficks. Information of elementary multidimensional Auditory Displays. *Journal of Acoustic Society of America*, pages 150–158, 1954. (Cited on page 31.)

Polli(2004) – Andrea Polli. Atmospherics/Weather Works: A multi-channel Storm Sonification Project. In *Proceedings of 10th ICAD*, 2004. Sidney, Australia. (Cited on page 39.)

Pope et al.(1999) – Jackson Pope, David Creasey, and Alan Chalmers. Realtime Room Acoustics Using Ambisonics. In *Proceedings of the 16th International AES Conference: Spatial Sound Reproduction*, pages 427–435, 1999. (Cited on page 108.)

Purcell et al.(2002) – Timothy J. Purcell, Ian Buck, William R. Mark, and Pat Hanrahan. Ray Tracing on Programmable Graphics Hardware. In *Proceedings of ACM SIGGRAPH*, pages 703–712, 2002. (Cited on pages 118 and 122.)

Purwins et al.(2000) – H. Purwins, B. Blankertz, and K. Obermayer. Computing Auditory Perception. *Organised Sound*, 5(3):159–171, 2000. (Cited on page 34.)

Quinn and Meeker(2001) – Marty Quinn and Loren David Meeker. Research set to Music: The Climate Symphony and Other Sonifications of Ice Core, Radar, DNA, Seismic and Solar Wind Data. In *Proceedings of ICAD*, pages 127–150, 2001. (Cited on page 57.)

Rabenhorst et al.(1990) – D.A. Rabenhorst, E.J. Farell, D.H. Jameson, T.D. Linton, and J.A. Mandelmann. Complementory Visualization and Sonification of Multidimensional Data. In *Proceedings of SPIE*, pages 147–153, 1990. (Cited on page 40.)

Röber et al.(submitted) – Niklas Röber, Martin Spindler, Ulrich Kaminski, and Maic Masuch. Graphics-based acoustic simulations. *IEEE Transactions on Visualization and Computer Graphics*, submitted. (Cited on pages 111 and 167.)

Richardson and Kaiwi(2002) – John F. Richardson and Jerry Kaiwi. Individualized Head Related Transfer Functions: Generation and Utility of Synthetic 3-D Audio in Simulation. In *The Proceedings of the 2002 Summer Computer Simulation Conference*, 2002. (Cited on pages 22 and 107.)

Rick and Mathar(2007) – T. Rick and R. Mathar. Fast Edge-Diffraction-Based Radio Wave Propagation Model for Graphics Hardware. In *Proceedings of ITG INICA*, 2007. (Cited on page 86.)

Ritter et al.(1999) – Alf Ritter, Joachim Böttger, Oliver Deussen, Matthias König, and Thomas Strothotte. Hardware-Based Rendering of Full-Parallax Synthetic Holograms. *Applied Optics*, 38(11):1364–1369, April 1999. (Cited on page 112.)

Röber and Masuch(2004a) – Niklas Röber and Maic Masuch. Auditory Game Authoring: From virtual Worlds to auditory Environments. In *Proceedings of CGAIDE 2004 Conference*, 2004a. London, England. (Cited on page 167.)

Röber and Masuch(2004b) – Niklas Röber and Maic Masuch. Interacting with Sound: An interaction Paradigm for virtual auditory Worlds. In *Proceedings of 10th ICAD*, 2004b. Sidney, Australia. (Cited on pages 51, 61, 62, 69, 137, 142, 144, and 167.)

Röber and Masuch(2005a) – Niklas Röber and Maic Masuch. Leaving the Screen: New Perspectives in Audio-only Gaming. In *Proceedings of 11th ICAD*, Limerick, Ireland, 2005a. (Cited on pages 53, 68, 139, and 167.)

Röber and Masuch(2005b) – Niklas Röber and Maic Masuch. Playing Audio-only Games: A Compendium of interacting with virtual, auditory Worlds. In *Proceedings of 2nd DIGRA Gamesconference*, Vancouver, Canada, 2005b. (Cited on pages 22, 142, 144, and 167.)

Röber et al.(2006a) – Niklas Röber, Eva C. Deutschmann, and Maic Masuch. Authoring of 3D virtual auditory Environments. In *Audio Mostly Conference*, pages 15–21, 2006a. (Cited on pages 22, 86, 87, 148, and 167.)

Röber et al.(2006b) – Niklas Röber, Cornelius Haber, Knut Hartmann, Matthias Feustel, and Maic Masuch. Interactive Audiobooks: Combining Narratives with Game Elements. In *Proceedings of TIDSE 2006 Conference*, pages 358–369, 2006b. (Cited on pages 48, 52, 155, 156, 167, 186, and 187.)

Röber et al.(2006c) – Niklas Röber, Martin Spindler, and Maic Masuch. Waveguide-based Room Acoustics through Graphics Hardware. In *ICMC*, 2006c. (Cited on pages 107, 108, 117, 158, and 167.)

Röber et al.(2007) – Niklas Röber, Ulrich Kaminski, and Maic Masuch. Ray Acoustics using Computer Graphics Technology. In *10th International Conference on Digital Audio Effects (DAFx)*, 2007. (Cited on pages 107, 109, 158, and 167.)

Roberts(1986) – Lind A. Roberts. Consonance Judgments of musical Chords by Musicians and untrained Listeners. *Acustica*, 62(5):163–171, 1986. (Cited on page 57.)

Rose et al.(1995) – E. Rose, D. Breen, K. Ahlers, D. Greer, C. Crampton, M. Tuceryan, and R. Whitaker. Annotating Real-World Objects Using Augmented Vision. In *Computer Graphics International '95*, 1995. (Cited on page 85.)

Rossiter and Ng(1996) – David Rossiter and Wai-Yin Ng. A System for the complementary Visualization of 3D Volume Images using 2D and 3D binaurally processed Sonification Representations. In *Proceedings of the 7th IEEE Visualization conference*, pages 351–354, 1996. (Cited on pages 16, 17, 55, and 59.)

Rozier et al.(2000) – Joseph Rozier, Karrie Karahalios, and Judith Donath. Hear & There: An Augmented Reality System of Linked Audio. In *Proceedings of ICAD*, 2000. (Cited on page 84.)

Rutherford and Geiger(1908) – Ernest Rutherford and Hans Geiger. An Electrical Method of Counting the Number of α-Particles from Radio-active Substances. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 81(546):141–161, August 1908. (Cited on pages 36, 55, and 56.)

Savioja et al.(1995) – L. Savioja, J. Backman, A. Järvinen, and T. Takala. Waveguide Mesh Method for Low-Frequency Simulation of Room Acoustics. In *15th Int. Congress on Acoustics (ICA'95)*, pages 637–640, Trondheim, Norway, 1995. (Cited on pages 113 and 114.)

Savioja et al.(2002) – Lauri Savioja, Tapio Lokki, and Jyri Huopaniemi. Auralization applying the Parametric Room Acoustic Modelling Technique - The DIVA Auralization System. In *Proceedings of ICAD*, Japan, 2002. (Cited on page 118.)

Scaleety and Craig(1991) – C. Scaleety and A. Craig. Using Sound to Extract Meaning from Complex Data. In *Proceedings of SPIE*, pages 207–219, 1991. (Cited on page 40.)

Seznec(2007) – Yann Seznec. The Wii Loop Machine: Musical Software Development for the Nintendo Wii Remote. In *Proceedings of AudioMostly Conference*, pages 48–51, 2007. (Cited on page 68.)

Sheridan(1992) – Thomas B. Sheridan. Musings on Telepresence and Virtual Presence. *Presence: Teleoperators and Virtual Environents*, 1(2):120–125, 1992. (Cited on page 46.)

Sloman(1971) – A. Sloman. Interactions between Philosophy and A.I. — The Role of Intuition and Nonlogical Reasoning in Intelligence. In *2nd IJCAI*, 1971. (Cited on page 36.)

Smith(1992) – Julius O. Smith. Physical modelling using digital Waveguides. *Computer Music Journal*, 16(4):75–87, 1992. (Cited on page 113.)

Smith et al.(1998) – Shamus Smith, Tim Marsh, David Duke, and Peter Wright. Drowning in Immersion. In *Proceedings UK-VRSIG'98*, Exeter, UK, 1998. (Cited on pages 47 and 48.)

Smith et al.(1992) – Stuart Smith, M. Pickett, Ronald, and Marian G. Williams. Environments for Exploring Auditory Representations of Multidimensional Data. In Gregory Kramer, editor, *Proceedings of ICAD*, pages 167–184, 1992. (Cited on pages 16 and 33.)

Speeth(1961) – S. D. Speeth. Seismometer Sounds. In *Journal of the Acoustical Society of America*, volume 33, pages 909–916, 1961. (Cited on page 31.)

Spindler et al.(2006) – Martin Spindler, Niklas Röber, Robert Döhring, and Maic Masuch. Enhanced Cartoon and Comic Rendering. In *EUROGRAPHICS 06*, 2006. Vienna, Austria. (Cited on page 50.)

Stockmann et al.(2008) – Lars Stockmann, Axel Berndt, and Niklas Röber. A Musical Instrument based on 3D Data and Volume Sonification Techniques. In *Audio Mostly Conference, to appear*, 2008. (Cited on pages 76, 128, and 167.)

Strothotte et al.(1995) – Thomas Strothotte, Helen Petrie, Valerie Johnson, and Lars Reichert. Mobic: An Aid to increase the independent Mobility of blind and elderly Travellers. In *2nd TIDE Congress*, 1995. (Cited on page 41.)

Taylor II et al.(2001) – Russell M. Taylor II, Thomas C. Hudson, Adam Seeger, Hans Weber, Jeffrey Juliano, and Aron T. Helser. VRPN: A Device-Independent, Network-Transparent VR Peripheral System. In *ACM Symposium on Virtual Reality Software & Technology*, 2001. (Cited on pages 76 and 129.)

Theußl et al.(2001b) – Thomas Theußl, Torsten Möller, and Eduard Gröller. Optimal Regular Volume Sampling. In *Proceedings IEEE Visualization*, pages 91–98, October 2001b. (Cited on page 115.)

Tory et al.(2001) – Melanie Tory, Niklas Röber, Torsten Möller, Anna Celler, and M. Stella Atkins. 4D Space-Time Techniques: A Medical Imaging Case Study. In *Proceedings of IEEE Visualization*, pages 473–476, San Diego, USA, 2001. (Cited on page 16.)

Tsingos and Drettakis(2004) – Emmanuel Tsingos, Nicolas Gallo and George Drettakis. Perceptual Audio Rendering of Complex Virtual Environments. In *Proceedings of ACM SIGGRAPH*, 2004. (Cited on pages 107 and 118.)

Tsingos(2007) – Nicolas Tsingos. Perceptually-based Auralization. In *Proceedings of 19th Intl. Congress on Acoustics*, 2007. (Cited on page 107.)

Tsingos et al.(2002) – Nicolas Tsingos, Ingrid Carlbom, Gary Elko, Tom Funkhouser, and Bob Kubli. Validation of Acoustical Simulations in the Bell Labs Box. *IEEE Computer Graphics & Applications, Special Issue on Virtual World, Real Sounds*, 22(4):28–37, 2002. (Cited on pages 169 and 232.)

VanDuyne and Smith(1993) – S. VanDuyne and J.O. Smith. Physical Modelling of the 2-D digital Waveguide Mesh. In *Int. Computer Music Conference*, pages 40–47, Tokyo, Japan, 1993. (Cited on page 113.)

Vickers and Hogg(2006) – Paul Vickers and Bennett Hogg. Sonification abstraite / Sonification concréte: An 'Æsthetic Perspective Space for classifying Auditory Displays in the Ars Musica Domain. In *Proceedings of ICAD*, 2006. (Cited on page 57.)

von Ehrenfels(1890) – Christian von Ehrenfels. Über Gestaltqualitäten. *Vierteljahrsschrift für wissenschaftliche Philosophie*, 14:249–292, 1890. (Cited on page 34.)

Walker and Lindsay(2005) – B. N. Walker and J. Lindsay. Using Virtual Reality to Prototype Auditory Navigation Displays. *Assistive Technology Journal*, 17(1):72–81, 2005. (Cited on page 84.)

Walker and Stanley(2005) – B. N. Walker and R. Stanley. Thresholds of Audibility for bone-conduction Headsets. In *11th International Conference on Auditory Display (ICAD)*, 2005. (Cited on page 146.)

Walker and Lindsay(2006) – Bruce N. Walker and Jeffrey Lindsay. Navigation Performance With a Virtual Auditory Display: Effects of Beacon Sound, Capture Radius, and Practice. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(14): 265–278, 2006. (Cited on page 146.)

Wand and Straßer(2004) – Michael Wand and Wolfgang Straßer. Multi-Resolution Sound Rendering. In *EUROGRAPHICS Symposium on Point-Based Graphics*, 2004. (Cited on page 118.)

Ware and Osborne(1990) – Colin Ware and Steven Osborne. Exploration and virtual Camera Control in virtual three-dimensional Environments. In *Proceedings of the 1990 symposium on Interactive 3D graphics*, pages 175–183, Snowbird, Utah, USA, 1990. ACM. (Cited on page 18.)

Webster and Weir(2005) – Gregory D. Webster and Catherine G. Weir. Emotional Responses to Music: Interactive Effects of Mode, Texture, and Tempo. *Motivation and Emotion)*, 29(1):19–39, 2005. (Cited on page 57.)

Weiser et al.(1999) – Mark Weiser, Rich Gold, and John Seely-Brown. The Origins of Ubiquitous Computing Research at PARC in the late 1980s. *IBM systems journal*, 38(4), 1999. (Cited on page 44.)

Wenzel(1992) – Elizabeth M. Wenzel. Spatial Sound and Sonification. In Gregory Kramer, editor, *Proceedings of ICAD*, pages 127–150, 1992. (Cited on pages 20, 21, 38, 201, and 231.)

Williams() – Sheila M. Williams. [perceptual principles in sound grouping. (Cited on pages 21, 34, 35, and 201.)

Witmer and Singer(1998) – Bob G. Witmer and Michael J. Singer. Measuring Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environents*, 7(3):225–240, 1998. (Cited on page 46.)

Yatim and Masuch(2007) – Maizatul Hayati Mohamad Yatim and Maic Masuch. GATE-LOCK - A Game Authoring Tool for Children. In *6th International Conference on Interaction Design and Children*, pages 173–174, 2007. (Cited on page 4.)

Youssef and Agrawala(2005) – Moustafa Youssef and Ashok K. Agrawala. The Horus WLAN Location Determination System. In *3rd International Conference on Mobile Systems, Applications, and Services (MobiSys 2005)*, 2005. (Cited on pages 85, 86, and 87.)

Youssef et al.(2003) – Moustafa Youssef, Ashok K. Agrawala, and A. Udaya Shankar. WLAN Location Determination via Clustering and Probability Distributions. In *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2003. (Cited on pages 85 and 86.)

THESES, REPORTS AND MANUALS

AGEIA Corp.(2006) – AGEIA Corp. AGEIA brings first Dedicated Physics Processor to Market; Launches New Age in Interactive Gaming. Press Release, March 2006. (Cited on page 170.)

AKG Acoustics GmbH(2008) – AKG Acoustics GmbH. *Hear0999 Audiosphere Broschure*, 2008. URL http://www.akg.com/mediendatenbank2/psfile/datei/28/hearo_999_4055c15b60e1f.pdf. (Cited on pages 73, 147, and 148.)

Amandusson(2003) – Sarah Amandusson. *Auditory Display and the VTK Sonification Toolkit*. Master's Thesis, Department of Numerical Analysis and Computer Science, Royal Institute of Technology, Stockholm, Sweden, 2003. (Cited on page 17.)

Andres(2005) – Sven Andres. *Entwicklung eines Audio Raytracing Systems zur Simulation von Head-Related Transfer Functions*. Master's Thesis, Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg, 2005. (Cited on pages 158, 160, and 161.)

Ascension Technology(2007) – Ascension Technology. *3D Navigator Broschure*, 2007. URL http://www.est-kl.com/hardware/tracking/ascension/3dnav_barco2.jpg. (Cited on page 18.)

Aureal(2000) – Aureal. Aureal announces Vortex3. Press Release, 2000. (Cited on pages 3, 26, and 106.)

Böttcher(2005) – Ralf Armin Böttcher. *Flow in Computerspielen*. Master's Thesis, Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg, 2005. (Cited on page 47.)

Creative Labs(2005) – Creative Labs. Creative introduces Sound Blaster X-Fi - Delivering dramatic Enhancements to any Music or Audio plus accelerated PC Gaming. Press Release, August 2005. (Cited on pages 26 and 106.)

Deutschmann(2006) – Eva C. Deutschmann. *Entwicklung eines Autorensystems für Augmented-Audio-Anwendungen*. Master's Thesis, Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg, 2006. (Cited on pages 74 and 148.)

Fey(2003) – Antje Fey. *Die Entwicklung des Hörbuchs in Deutschland – Geschichte, Formen und Rezeption*. Master's Thesis, Johann-Wolfgang-Goethe-Universität, Frankfurt am Main, Germany, 2003. (Cited on pages 30, 51, and 52.)

Govindaraju and Manocha(2007) – Naga K. Govindaraju and Dinesh Manocha. *Cache-Efficient Numerical Algorithms using Graphics Hardware*. Technical Report, Department of Computer Science, UNC Chapel Hill, 2007. (Cited on page 112.)

Gyration(2004) – Gyration. *Gyration Ultra Cordless Optical Mouse – User's Manual*, 2004. URL http://www.gyration.com/descriptions/document/GP110-MANUAL-EN.pdf. (Cited on pages 86 and 87.)

Hämäläinen(2003) – Perttu Hämäläinen. *3D Sound Rendering and Cinematic Sound in Computer Games*. Technical Report, Helsinki University of Technology, Telecommunications Software and Multimedia Laboratory, Finnland, 2003. (Cited on page 28.)

Hermann(2002) – Thomas Hermann. *Sonification for Exploratory Data Analysis*. PhD Thesis, Bielefeld University, 2002. (Cited on page 33.)

Hoppe(1998) – Axel Hoppe. *Validierung und Nachbearbeitung von gerenderten Bildern*. PhD Thesis, Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg, 1998. (Cited on page 44.)

Huber(2004) – Cornelius Huber. *Erweiterbarkeit von AUXplorer und Implementierung eines interaktiven Hörbuchs*. Technical Report, Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg, 2004. Lab-Intern Report. (Cited on page 143.)

Huber(2006) – Cornelius Huber. *HörSpiele - Entwicklung eines Konzepts zur Gestaltung interaktiver Hörbücher*. Master's Thesis, Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg, 2006. (Cited on pages 155 and 156.)

Hugenberg(2007) – Hanno Hugenberg. *Audiosimulation auf der Ageia PhysX PPU*. Technical Report, Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg, 2007. Lab-Intern Report. (Cited on page 170.)

InterSense(2004) – InterSense. *InterSense Wireless InertiaCube3 – Broschure*, 2004. URL http://www.isense.com/uploadedFiles/Products/WirelessInertiaCube3.pdf. (Cited on pages 86 and 87.)

Kaminski(2007) – Ulrich Kaminski. *3D Strahlenbasiertes Sound Rendering mit Computergrafik-Hardware*. Master's Thesis, Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg, 2007. (Cited on pages 122, 158, and 160.)

Malyszczyk and Mewes(2005) – Arvid Malyszczyk and Maik Mewes. *SIA - Das Speech-Interface for Adventuregames*. Technical Report, Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg, 2005. Lab-Intern Report. (Cited on pages 23, 71, and 138.)

Miede and Futterlieb(2005) – André Miede and Jörg Futterlieb. *Entwicklung von Authoringtools für virtuelle, auditive Welten*. Technical Report, Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg, 2005. Lab-Intern Report. (Cited on pages 133 and 138.)

Moreno-Fortuny(2004) – Gabriel Moreno-Fortuny. *Distributed Stream Processing Ray Tracer*. Technical Report, University of Waterloo, http://www.cgl.uwaterloo.ca/~gmoreno/streamray.html 2004. (Cited on pages 118 and 122.)

Neumann(2004) – Thomas Neumann. *Eine Geometrie-Engine zur Berechnung von 3D-Sound mit Raumakustik-Effekten in Echtzeit*. Master's Thesis, Tchnische Universität Braunschweig, Institut für Compuergraphik, 2004. (Cited on page 118.)

Nvidia(2007) – Nvidia. *NVIDIA Tesla | GPU Computing Solutions for HPC*. Santa Clara, USA, 2007. URL http://www.nvidia.com/tesla. (Cited on page 117.)

OseanServer Technology Inc.(2007) – OseanServer Technology Inc. *Digital Compass Users Guide, OS5000 Series*, 2007. URL http://www.ocean-server.com/download/OS5000_Compass_Manual.pdf. (Cited on pages 86, 87, and 89.)

Otto and Domke(2007) – Mathias Otto and Daniel Domke. *Entwurf und Entwicklung eines Wireless-LAN-Positionierungssystems für eine Augmented Audio Anwendung*. Technical Report, Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg, 2007. Lab-Intern Report. (Cited on page 86.)

Otto and Kurth(2008) – Mathias Otto and Hannes Kurth. WLAN Positioning Accuracy. Personal Communication, 2008. (Cited on pages 87 and 153.)

Polhemus(2008) – Polhemus. *FASTRAK Broschure and Specs*, 2008. URL http://polhemus.com/?page=Motion_Fastrak. (Cited on pages 67, 68, and 72.)

Purcell(2004) – Timothy J. Purcell. *Ray Tracing on a Stream Processor*. PhD Thesis, Stanford University, March 2004. (Cited on pages 118 and 122.)

Ritter(2005) – Felix Ritter. *Interaktives Illustrieren von Informationsräumen: Räumliche und funktionale Zusammenhänge spielerisch begreifen*. PhD Thesis, Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg, 2005. (Cited on page 44.)

Röber(2000) – Niklas Röber. *Multidimensional Analysis and Visualization Software for Dynamic SPECT*. Technical Report, Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg, 2000. Internship Report. (Cited on page 16.)

Röber(2001) – Niklas Röber. *Realtime-Rendering Nicht-Photorealistischer Computergraphiken mit OpenGL*. Technical Report, Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg, 2001. Lab-Intern Report. (Cited on pages 51, 80, and 231.)

Röber(2002) – Niklas Röber. *Visualization of Fuel Cell Simulations*. Master's Thesis, Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg, 2002. (Cited on page 115.)

Röber and Masuch(2006) – Niklas Röber and Maic Masuch. *Soundpipes: A new way of Path Sonification*. Technical Report 5, Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg, 2006. (Cited on pages 62, 132, 137, and 167.)

Röber et al.(2006) – Niklas Röber, Sven Andres, and Maic Masuch. *HRTF Simulations through acoustic Raytracing*. Technical Report 4, Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg, 2006. (Cited on pages 158, 161, and 167.)

Sasse(2007) – Dennis Sasse. *User Interfaces für Audiospiele und Interaktive Hörbücher*. Technical Report, Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg, 2007. Lab-Intern Report. (Cited on pages 155 and 157.)

Stockmann(2007) – Lars Stockmann. *Designing an Audio API for Mobile Platforms*. Technical Report, Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg, 2007. (Cited on pages 72, 75, 90, and 158.)

Stockmann(2008) – Lars Stockmann. *Interaktive Sonifikation von 2D/3D Daten*. Master's Thesis, Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg, 2008. (Cited on pages 18, 50, 57, 76, and 128.)

Theremin(1924) – Leo Sergejewitsch Theremin. *Method of and Apparatus for the Generation of Sounds*. United States Patent Office, December 5th 1924. URL http://mixonline.com/online_extras/theremin-patent.pdf. 73,529. (Cited on pages 2 and 3.)

THX Consortium(2000) – THX Consortium. *THX Certified Training Programm – Presentation Materials*. San Rafael, USA, June 2000. (Cited on page 27.)

Vonia Corporation(2008) – Vonia Corporation. *EZ-80P/S20 Broschure*, 2008. URL http://www.dowumi.com/eng_product/html/view.php?no=43&categoryid=17. (Cited on pages 82, 89, 145, 147, 148, and 232.)

Walz(2004a) – Daniel Walz. *Techniken zur Erkundung von auditiven virtuellen Umwelten*. Master's Thesis, Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg, Germany, 2004a. (Cited on pages 53 and 74.)

Whalen(2005) – Sean Whalen. *Audio and the Graphics Processing Unit*. Technical Report, 2005. http://www.node99.org/projects/gpuaudio/gpuaudio.pdf. (Cited on pages 108, 110, and 111.)

BOOKS, CHAPTERS AND MONOGRAPHES

3rd Baron Rayleigh(1877/1878) – John William Strutt 3rd Baron Rayleigh. *The Theory of Sound*, volume I/II. Cambridge University, 1877/1878. (Cited on page 1.)

Begault(1994) – Durand R. Begault. *3D Sound - For Virtual Reality and Multimedia*. Academic Press Professional, 1994. (Cited on pages 15, 38, 50, and 201.)

Beranek(1986) – Leo L. Beranek. *Acoustics*. Published for the Acoustical Society of America through the American Institute of Physics, 1986. (Cited on page 119.)

Bilbao(2004) – S. Bilbao. *Wave and Scattering Methods for Numerical Simulation* . John Wiley & Sons, 2004. ISBN 0470870176. (Cited on page 114.)

Bimber and Raskar(2005) – Oliver Bimber and Ramesh Raskar. *Spatial Augmented Reality: Merging Real and Virtual Worlds*. A K Peters, Ltd., 2005. (Cited on pages 81 and 84.)

Boer(2002b) – James R. Boer. *Game Audio Programming*. Charles River Media, 2002b. (Cited on pages 25, 26, 27, 32, 72, and 106.)

Bowman et al.(2004) – Doug A. Bowman, Ernst Kruijff, Joseph J. LaViola, and Ivan Poupyrev. *3D User Interfaces: Theory and Practice*. Addison-Wesley Professional, 2004. (Cited on pages 43, 44, 47, 49, 55, 66, 68, 81, and 84.)

Bregman(1990) – A. S. Bregman. *Auditory Scene Analysis: The perceptual Organization of Sound*. MIT Press, 1990. (Cited on pages 20 and 34.)

Conway and Sloane(1976) – J. H. Conway and N. J. A. Sloane. *Sphere Packings, Lattices and Groups*. Springer, 2nd edition, 1976. ISBN 0-387-96617-X. (Cited on page 115.)

Crawford(2002) – Chris Crawford. *The Art of interactive Design - An Euphonious and Illuminating Guide to build sucessfull Software*. No Starch Press, San Francisco, USA, 2002. (Cited on pages 18, 47, and 55.)

Csákszentmihályi(1975) – Mihály Csákszentmihályi. *Beyond Boredom and Anxiety*. Jossey-Bass, San Francisco, 1975. (Cited on page 47.)

Dempinski and Viale(2005) – Kelly Dempinski and Emmanuel Viale. *Advanced ligthing and materials with shaders*. Wordware Publishing, Inc., 2005. (Cited on pages 108 and 170.)

Dove et al.(1842) – Heinrich Wilhelm Dove, Ludwig Ferdinand, and Ludwig Moser. *Repertorium der Physik: Eine Zusammenstellung der neueren Fortschritte dieser Wissenschaft.*, volume 6. Veit & comp., 1842. (Cited on page 21.)

Dutre et al.(2003) – Philip Dutre, Philippe Bekaert, and Kavita Bala. *Advanced Global Illumination*. AK Peters Ltd., 2003. (Cited on pages 119 and 121.)

Everest(1994) – F. Alton Everest. *The Master Handbook of Acoustics (3rd edition)*. TAB Books (Imprint of McGraw-Hill), 1994. (Cited on page 22.)

Everest(2001) – F. Alton Everest. *The Master Handbook of Acoustics*, volume 4th edition. McGraw-Hill/TAB Electronics, 2001. (Cited on pages 22 and 118.)

Faulkner(1998) – Christine Faulkner. *The Essence of Human-Computer Interaction*. Prentice Hall Europe, 1998. (Cited on pages 17 and 50.)

Goldstein(2007) – E. Br. Goldstein. *Cognitive Psychology*. Wadsworth Publishing, 2007. (Cited on pages 15, 19, 21, 22, 23, 34, 67, and 112.)

Gooch and Gooch(2001) – Bruce Gooch and Amy Gooch. *Non-Photorealistic Rendering*. AK Peters, 2001. (Cited on pages 50 and 58.)

Heilbrun and Stacks(1991) – Adam Heilbrun and Barbara Stacks. *Cyberspace. Ausflüge in virtuelle Wirklichkeiten*, chapter Was heißt virtuelle Realität? Ein Interview mit Jaron Lanier. Reinbek, Hamburg, 1991. (Cited on page 43.)

Hiebert(2006) – Garin Hiebert. *OpenAL Programmer's Guide - Version 1.0 and 1.1*. Creative Technology Limited, 2006. (Cited on pages 25, 26, 27, 28, 101, 105, and 109.)

Huang and Benesty(2004) – Yiteng Arden Huang and Jacob Benesty, editors. *Audio Signal Processing for Next-Generation Multimedia Communication Systems*. Springer, 2004. (Cited on page 161.)

Jackson(1991) – A. G. Jackson. *Handbook of Crystallography*. Springer, 1991. (Cited on page 115.)

Knight(1960) – David C. Knight. *The first Book of Sound: A basic Guide to the Science of Acoustics*. Franklin Watts, Inc. New York, 1960. (Cited on pages 2 and 3.)

Kramer(1994) – Gregory Kramer. *Auditory Display: Sonification, Audification, and auditory Interfaces*, chapter An Introduction to Auditory Display, pages 1–78. Addison Wesley, 1994. (Cited on pages 1, 3, 21, 22, 31, 33, 36, 38, 40, 50, 57, 59, 201, 231, and 233.)

Kramer et al.(1997) – Gregory Kramer, Bruce Walker, Terri Bonebright, Perry Cook, John Flowers, Nadine Miner, and John Neuhoff. *Sonification Report: Status of the Field and Research Agenda*. ICAD Community, 1997. Prepared for the National Science Foundation by members of the International Community for Auditory Display. (Cited on pages 32 and 55.)

Kuttruff(2000) – Heinrich Kuttruff. *Room Acoustics*. Spon Press, London, 4th edition, 2000. (Cited on pages 24 and 118.)

Leinung and Stumvoll(1904) – W. Leinung and R. Stumvoll. *Aus Magdeburgs Sage und Geschichte*. Neumann, 1904. (Cited on page 143.)

Mach(1886) – Ernst Waldfried Josef Wenzel Mach. *Beiträge zur Analyse der Empfindungen*. G. Fischer, Jena, 1886. (Cited on page 34.)

Matlin(1987) – Margaret W. Matlin. *Sensation and Perception*. Pearson Allyn & Bacon, 1987. (Cited on pages 15 and 19.)

Moore(1989) – B.C.J. Moore. *An Introduction to the Psychology of Hearing*. Academic Press, London, 3rd edition, 1989. (Cited on page 34.)

Mosco(2005) – Vincent Mosco. *The Digital Sublime: Myth, Power and Cyberspace*. The MIT Press, 2005. (Cited on page 170.)

Murray(1998) – J. H. Murray. *Hamlet on the Holodeck. The Future of Narrative in Cyberspace*. MIT Press, 1998. (Cited on pages 47 and 154.)

Newton(1687) – Isaac Newton. *Philosophiae Naturalis Principia Mathematica*. Royal Society, London, 1687. (Cited on page 1.)

Peacock et al.(2006) – Daniel Peacock, Peter Harrison, Andrea D'Orta, Valery Carpentier, and Edward Cooper. *OpenAL Effects Extension Guide - Version 1.0 and 1.1*. Creative Technology Limited, 2006. (Cited on pages 25, 27, 28, 105, and 109.)

Poe(1843) – Edgar Allan Poe. *The Gift: A Christmas and New Year's Present*, chapter The Pit and the Pendulum. Carey & Hart, 1843. (Cited on page 156.)

Raskin(2000) – J. Raskin. *The Human Interface: New Directions for Interactive Systems*. ACM Press, Addison Wesley, Massachusetts, USA, 2000. (Cited on page 17.)

Rocchesso(2003) – Davide Rocchesso. *Introduction to Sound Processing*. GNU Free Documentation License, 2003. http://www.scienze.univr.it/~rocchess. (Cited on pages 23 and 25.)

Rolland et al.(2001) – Jannick P. Rolland, Yohan Baillot, and Alexei A. Goon. *Fundamentals of Wearable Computers and Augmented Reality*, chapter A Survey of Tracking Technology for virtual Environments, pages 67–112. Lawrence Erlbaum Associates., 2001. (Cited on pages 85 and 87.)

Salen and Zimmerman(2003) – Katie Salen and Eric Zimmerman. *Rules of Play: Game Design Fundamentals*. MIT Press, 2003. (Cited on pages 67 and 72.)

Schirra(2000) – J.R.J. Schirra. *Vom Realismus der Bilder: Interdisziplinäre Forschungen zur Semantik bildhafter Darstellungsformen*, chapter Täuschung, Ähnlichkeit und Immersion: Die Vögel des Zeuxis, pages 119–135. Magdeburg: Scriptum, 2000. (Cited on page 46.)

Schmitt(1968) – Erich Schmitt. *Das dicke Schmitt-Buch*. Eulenspiegel-Verlag, Berlin, 1968. (Cited on pages 10 and 231.)

Schroeder et al.(2004) – William Schroeder, Ken Martin, and Bill Lorensen. *The Visualization Toolkit*. Kitware Inc., General Electric GE, 3rd edition, 2004. (Cited on pages 15, 16, and 17.)

Schumann and Müller(2000) – Heidrun Schumann and Wolfgang Müller. *Visualisierung: Grundlagen und allgemeine Methoden*. Springer, 2000. (Cited on pages 3, 15, 16, 32, 35, and 50.)

Shilling and Shinn-Cunningham(2002) – Russell D. Shilling and Barbara Shinn-Cunningham. *Handbook of Virtual Environments: Design, Implementation, and Applications.*, chapter Virtual Auditory Displays, pages 65–92. Mahwah, NJ: Erlbaum., 2002. (Cited on pages 33, 50, and 53.)

Shneiderman(2004) – Ben Shneiderman. *Designing the User Interface*. Addison-Wesley, Amsterdam, 2004. (Cited on pages 17, 18, and 50.)

Smith(1997) – Steven W. Smith. *The Scientist and Engineer's Guide to Digital Signal Processing*. California Technical Pub., 1997. (Cited on pages 23 and 25.)

Strothotte and Schlechtweg(2002) – Thomas Strothotte and Stefan Schlechtweg. *Non-Photorealistic Computer Graphics: Modeling, Rendering and Animation*. Morgan Kaufmann, 2002. (Cited on page 50.)

Stuart(2001) – Rory Stuart. *Design of Virtual Environments*. Barricade Books, 2001. (Cited on page 46.)

Vorländer(2007) – Michael Vorländer. *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Springer, RWTHedition, 2007. (Cited on pages 21, 22, 50, and 112.)

Warren(1999) – Richard M. Warren, editor. *Auditory Perception - A new Analysis and Synthesis*. Cambridge University Press, UK, 1999. (Cited on pages 19 and 22.)

Wells(1984a) – Alexander Frank Wells. *Structural Inorganic Chemistry*. Oxford University Press, 5th edition, 1984a. (Cited on page 115.)

Wendemuth et al.(2004) – Andreas Wendemuth, Edin Andelic, and Sebastian Barth. *Grundlagen der stochastischen Sprachverarbeitung*. Oldenbourg, 2004. (Cited on pages 23 and 71.)

Wickens et al.(2003) – Christopher D. Wickens, John D. Lee, Yili Liu, and Sallie E. Gordon-Becker. *An Introduction to Human Factors Engineering*. Prentice Hall, 2003. (Cited on page 18.)

Yamamoto(1710-1717/2002) – Tsunetomo Yamamoto. *Hagakure. The Way of the Samurai*. Hokuseido Press (Reprint), Tokyo, 1710-1717/2002. originally recorded by Tashiro Tsuramoto, translated by Takao Mukoh. (Cited on page xi.)

Zander(1995) – Horst Zander. *Audio am PC*. Pearson Education Deutsch – Markt & Technik Verlag, 1995. (Cited on pages 25 and 27.)

Zölzer(2002) – Udo Zölzer, editor. *DAFX - Digital Audio Effects*. John Wiley & Sons, West Sussex, England, 2002. (Cited on pages 23, 24, 25, 109, and 231.)

MUSIC, COMPOSITIONS, AUDIOBOOKS AND MOVING PICTURES

Andersen(2005) – Hans Christian Andersen. The Nightingale – 3D Audiobook. AM3D A/S, CD, 2005. (Cited on pages 30 and 201.)

Curtis(1966-1971) – Dan Curtis. Dark Shadows. TV Series, 1966-1971. (Cited on page 49.)

Mozart(1787) – W. A. Mozart. Musikalisches Würfelspiel: Anleitung so viel Walzer oder Schleifer mit zwei Würfeln zu componieren ohne musikalisch zu seyn noch von der Composition etwas zu verstehen. Köchel Catalog of Mozart's Work KV1 Appendix 294d or KV6 516f, 1787. (Cited on page 28.)

Mussorgski(1874/1886) – Modest Mussorgski. Pictures at an Exhibition, 1874/1886. (Cited on page 52.)

Prince(2004) – Prince. Musicology. Smi Col (Sony BMG), CD, 2004. (Cited on page vii.)

Prokofjew(1936) – Sergej Prokofjew. Peter and the Wolf, 1936. Op. 67, A Children's Tale for Narrator and Orchestra. (Cited on page 52.)

Ravel(1922) – Maurice Ravel. Pictures at an Exhibition, 1922. (Cited on page 52.)

Tomita(1975) – Isao Tomita. Pictures at an Exhibition, 1975. for Synthezizer. (Cited on page 52.)

Verne(2005) – Jules Verne. Die Reise zum Mittelpunkt der Erde. MDR, Der Hörverlag, 5.1 Surround Audio-DVD, 2005. (Cited on pages 30 and 201.)

Wachowski and Wachowski(1999) – Andy Wachowski and Larry Wachowski. The Matrix. Movie, 1999. (Cited on page 142.)

COMPUTER GAMES

Atkinson and Gucukoglu(2008) – Matthew Tylee Atkinson and Sabahattin (AGRIP) Gucukoglu. AudioQuake2008, 2008. URL http://www.agrip.org.uk/. PC. (Cited on pages 23, 30, 71, 141, and 142.)

Buena Vista Games(2006) – Buena Vista Games. Every Extend Extra. PSP, UMD, 2006. (Cited on page 29.)

Cohen et al.(2004) – Yishay Cohen, Jolanda Dekker, Arnout Hulskamp, David Kousemaker, Tim Olden, Cees Taal, and Wouter Verspaget. Demor - Location based 3D Audiogame, 2004. URL http://student-kmt.hku.nl/~g7/site/. (Cited on pages 30, 84, and 85.)

Dannecker et al.(2003) – Fachhochschule Stuttgart (Tanja Dannecker, Matthias Pasedag, Christa Stoll, and Heinrich Sturm). Der Tag wird zur Nacht, 2003. PC. (Cited on pages 140 and 142.)

Destiny Media(1999) – Destiny Media. Seuss Crane - Detective For Hire. Webgame, 1999. (Cited on pages 30, 142, and 143.)

Eidos Interactive(1998) – Eidos Interactive. Thief: The Dark Project. Windows, CD-ROM, 1998. (Cited on page 49.)

GMA Games(2001) – GMA Games. Shades of Doom. Published by GMA Games, 2001. PC. (Cited on pages 30, 140, and 141.)

Harmonix Music Systems(2005) – Harmonix Music Systems. Guitar Hero. Playstation2, CD-ROM, 2005. (Cited on pages 28 and 29.)

Indies Zero(2005) – Indies Zero. Electroplankton. NintendoDS, 2005. (Cited on page 29.)

Infocom(1982) – Infocom. Zork: The Great Underground Empire. PS, Disk, 1982. (Cited on page 52.)

Konami(2001) – Konami. Dance Dance Revolution. Playsation2, CD-ROM, 2001. (Cited on page 29.)

Konami(2007) – Konami. Silent Hill Origins. PlaystationPortable(PSP), UMD, 2007. (Cited on pages 27 and 49.)

Lucas Arts(1993) – Lucas Arts. Day of the Tentacle. Lucas Arts. Entertainment, 1993. PC. (Cited on page 138.)

Microïds(2003) – Microïds. Syberia. PC/Windows, CD-ROM, 2003. (Cited on pages 20, 22, and 231.)

Namco(2007) – Namco. Ridge Racer 2. PlaystationPortable(PSP), UMD, 2007. (Cited on page 27.)

NaNaOn-Sha(1996) – NaNaOn-Sha. PaRappa the Rapper. Playstation, CD-ROM, 1996. (Cited on page 29.)

Nintendo Europe(2007) – Nintendo Europe. Wii Fit. Wii, CD-ROM, 2007. (Cited on page 68.)

Nintendo Europe(2006) – Nintendo Europe. Wii Sports. Wii, CD-ROM, 2006. (Cited on page 68.)

Pin Interactive(2003) – Pin Interactive. Terraformers, 2003. PC. (Cited on pages 30, 140, 142, and 143.)

Q Entertainment(2004) – Q Entertainment. Lumines. PlaystationPortable (PSP), UMD, 2004. (Cited on page 29.)

Quantic Dream(2005) – Quantic Dream. Fahrenheit – Indigo Prophecy. Windows, CD-ROM, 2005. (Cited on page 52.)

Sega(2000) – Sega. Samba De Amigo, 2000. (Cited on page 28.)

Sierra On-Line(1983) – Sierra On-Line. Frogger, 1983. PC. (Cited on page 141.)

Sony Entertainment(2003) – Sony Entertainment. EyeToy: Play, 2003. (Cited on page 68.)

Sony Entertainment(2004) – Sony Entertainment. SingStar. Playstation2, CD-ROM, 2004. (Cited on page 29.)

Sumo Digital(2007) – Sumo Digital. Driver76. PlaystationPortable(PSP), UMD, 2007. (Cited on page 27.)

Tarsier Studios(2008) – Tarsier Studios. Metronome. PC, XBox360, CD-ROM, 2008. (Cited on page 29.)

thatgamecompany(2008) – thatgamecompany. flOw. PlaystationPortable(PSP), Download, 2008. (Cited on pages 29 and 47.)

United Game Artists(2001) – United Game Artists. REZ. Dreamcast, Playstation2, CD-ROM, 2001. (Cited on pages 29 and 47.)

Warp(1999) – Warp. Real Sound: Kaze no Regret. Published by Sega/Sony, 1999. Sega Saturn, Dreamcast. (Cited on pages 30 and 41.)

SOFTWARE AND INTERNET REFERENCES

AM3D A/S(2008) – AM3D A/S. AM:3D Positional Audio. Product Website, 2008. URL http://www.am3d.com. (Cited on pages 25, 26, 105, 147, and 201.)

Andresen(2002) – Gavin Andresen. Playing by Ear: Creating Blind-Accessible Games, 2002. URL http://www.gamasutra.com/resource_guide/20020520/andersen_pfv.htm. (Cited on page 29.)

audiobooks.com(2008) – audiobooks.com. Audiobooks Store, 2008. URL http://www.audiobooks.com/. (Cited on page 30.)

Autodesk(2008) – Autodesk. 3D Studio MAX 9, 2008. URL http://images.autodesk.com/adsk/files/3dsmax_2009_brochure_us.pdf. (Cited on pages 95 and 100.)

Bartiméus Accessibility Foundation(2008) – Bartiméus Accessibility Foundation. Audio Game Maker, 2008. URL http://www.audiogamemaker.net. (Cited on page 29.)

Brode et al.(2008) – Tatjana Brode, Jens Krisinger, Mathias Ott, Marc Buhl, Jens Sparschuh, Ulrich Peltzer, Kathrin Röggla, Tanja Langer, Gerhard Falkner, Tanja Dückers, Michal Hvorecky, Tina Schimansky, and Das Helmi. Ein GPS-basiertes Literaturprojekt im Netz und auf den Straßen Berlins. Web Article, 2008. URL http://www.landvermesser.tv. (Cited on page 158.)

CNET.COM(2005) – CNET.COM. Play games in surround sound. CNET.COM – Web Article, January 2005. URL http://www.cnet.com/4520-7384_1-6244092-1.html. (Cited on page 22.)

Deutsch(1995/2003) – Diana Deutsch. Musical Illusions and Paradoxes, 1995/2003. URL http://deutsch.ucsd.edu/psychology/deutsch_research1.php. (Cited on page 21.)

ELAC Technische Software(2008) – ELAC Technische Software. CARA 2.1/2.2 PLUS: Computer Aided Room Acoustics. Software, 2008. URL http://www.cara.de. Kiel, Germany. (Cited on page 159.)

Evergreen Technologies(2005) – Evergreen Technologies. RumbleFX ForceFeedback Headphones. Product Website, 2005. URL http://www.evergreennow.com. (Cited on pages 70 and 73.)

Firelight Technologies Pty, Ltd(2001-2008) – Firelight Technologies Pty, Ltd. Fmod music & sound effects system. Product Website, 2001-2008. URL http://www.fmod.org. (Cited on pages 25, 26, and 105.)

Fraunhofer IFF(2008) – Fraunhofer IFF. Virtual Development and Training Centre VDTC, 2008. URL http://www.vdtc.de. (Cited on page 47.)

Gardner and Martin(2000) – Bill Gardner and Keith Martin. HRTF Measurements of a KEMAR Dummy-Head Microphone. Web Article/Archive, 2000. URL http://sound.media.mit.edu/KEMAR.html. MIT Media Lab. (Cited on page 124.)

Gasior(1999) – David Gasior. Creative Labs versus Aureal. DSP Wiki – Web Article, December 1999. URL http://dspwiki.com/index.php?title=Creative_Labs_v._Aureal&printable=yes. (Cited on page 26.)

Giovanni(2008) – A. Giovanni. Adventure Maker, 2008. URL http://www.adventuremaker.com/. (Cited on page 93.)

Girvan(2005) – Ray Girvan. Sound Sense: Sonification, 2005. URL http://www.scientific-computing.com/features/feature.php?feature_id=58. (Cited on page 58.)

Hermann(2006) – Thomas Hermann. An Overview of Auditory Displays and Sonification. Web Article, 2006. URL http://www.sonification.de/main-ad.shtml. (Cited on pages 31, 32, and 231.)

ICAD Community(1992 - 2008) – ICAD Community, 1992 - 2008. URL http://www.icad.org. (Cited on page 31.)

Jones(2008) – Chris Jones. Adventure Game Studio, 2008. URL http://www.adventuregamestudio.co.uk/. (Cited on page 93.)

Kennedy(2007) – Brendan Kennedy. Bubbles of Sound in the Public Space: How iPods are Changing the Way we Interact. The Queen's Journal – Web Article, January 2007. URL http://www.queensjournal.ca/story/2007-01-23/postscript/bubbles-sound-public-space/. (Cited on pages 170, 171, and 232.)

Kitware Inc.(2008) – Kitware Inc. VTK Home Page, 2008. URL http://www.vtk.org. (Cited on pages 16 and 17.)

Lee(2007) – Johnny Chung Lee. Wiimote Projects, 2007. URL http://www.cs.cmu.edu/~johnny/projects/wii/. (Cited on page 68.)

Meijer(2008) – Peter B.L. Meijer. Vision Technology for the Totally Blind, 2008. URL http://www.seeingwithsound.com. (Cited on pages 55, 57, and 58.)

Menshikov(2003) – Aleksei Menshikov. Modern Audio Technologies in Games. Web Article, 2003. URL http://www.digit-life.com/articles2/sound-technology/. Game Developers Conference. (Cited on pages 28, 38, and 231.)

Miede(2007) – Andre Miede. Classic Thesis Latex Style, 2007. URL http://www.ctan.org/tex-archive/macros/latex/contrib/classicthesis/. (Cited on page 205.)

New Scientist(2008) – New Scientist. Music Special: Five great auditory Illusions. Web Article, February 2008. URL http://www.newscientist.com/article/dn13355-music-special-five-great-auditory-illusions-.html. (Cited on page 21.)

OpenAL(2008) – OpenAL. OpenAL: Cross-Platform 3D Audio. Product Website, 2008. URL http://www.openal.org. (Cited on pages 27 and 75.)

Polly(2003) – Andrea Polly. Atmospheric/Weather Works: The Sonification of Meteorological Data, 2003. URL http://andreapolli.com/studio/atmospherics/. (Cited on pages 39 and 201.)

Quinn(2007) – Marty Quinn. Design Rhythmics Sonification Research Lab, 2007. URL http://www.drsrl.com. (Cited on page 57.)

Ramelet(2000a) – Philippe Ramelet. Guide du son 3D, Partie 1 l'A3D. Web Article at Hardware.fr, 2000a. URL www.hardware.fr/art/imprimer/145. (Cited on pages 2, 3, 26, and 27.)

Ramelet(2000b) – Philippe Ramelet. Guide du son 3D, Partie 1 l'EAX. Web Article at Hardware.fr, 2000b. URL www.hardware.fr/art/imprimer/147. (Cited on pages 26 and 27.)

Reiners and Voss(2008) – Dirk Reiners and Gerrit Voss. OpenSG: Open Source Scenegraph System, 2008. URL http://www.opensg.org/. (Cited on page 73.)

Schneider and Muschett(1998) – Toni Schneider and Mark Muschett. Interview with Toni Schneider From Aureal Semiconductor about Wavetracing, 1998. URL http://www.3dsoundsurge.com/interviews/aureal2.html. (Cited on page 3.)

SDL(2008) – SDL. SDL – Simple Direct Media Layer. Product Website, 2008. URL http://www.libsdl.org/. (Cited on page 26.)

Sennheiser(2008) – Sennheiser. guidePORT System, 2008. URL http://www.guideport.de. (Cited on pages 39, 81, 84, and 85.)

Skyhook Wireless(2008) – Skyhook Wireless. WiFi Positioning System (WPS), 2008. URL http://www.skyhookwireless.com. (Cited on page 87.)

Sony Creative Software(2008) – Sony Creative Software. Sound Forge 9, 2008. URL http://www.sonycreativesoftware.com/soundforge. (Cited on pages 95 and 100.)

SPSS Inc.(2008) – SPSS Inc. SPSS, Data Mining, Statistical Analysis Software, Predictive Analysis, Predictive Analytics, Decision Support Systems, 2008. URL http://www.spss.com/. (Cited on pages 128 and 206.)

Surlykke and Kalko(2008) – Annemarie Surlykke and Elisabeth K. V. Kalko. Echolocating Bats Cry Out Loud to Detect Their Prey. PLoS ONE – Online Science Journal, 2008. URL http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0002036. (Cited on page 63.)

Taylor II et al.(2008) – Russell M. Taylor II, Thomas C. Hudson, Adam Seeger, Hans Weber, Jeffrey Juliano, and Aron T. Helser. Virtual Reality Peripheral Network, 2008. URL http://www.cs.unc.edu/Research/vrpn/. (Cited on pages 76 and 129.)

Tidwell(1999) – Jenifer Tidwell. Common Ground: A Pattern Language for Human-Computer Interface Design, 1999. URL http://www.mit.edu/~jtidwell/common_ground_onefile.html. (Cited on pages 17 and 18.)

van der Honing(2003) – M. van der Honing. 3D Adventure Studio, 2003. URL http://3das.noeska.com/. (Cited on page 93.)

van Tol and Huiberts(2006) – Richard van Tol and Sander Huiberts. Audiogames Website. http://www.audiogames.net, 2006. (Cited on pages 29, 38, and 140.)

Zizza(2000) – Keith Zizza. Your Audio Design Document: Important Items to Consider in Audio Design, Production, and Support, 2000. URL http://www.gamasutra.com/features/20000726/zizza_pfv.htm. (Cited on page 47.)

LIST OF FIGURES

LIST OF LISTINGS

LIST OF TABLES

ACRONYMS

2D	Two Dimensional
3D	Three Dimensional
AAR	Augmented Audio Reality
AD	Auditory Display
AGP	Accelerated Graphics Port
API	Application Programming Interface
AR	Augmented Reality
A/S	Analogic/Symbolic (Continuum)
AV	Augmented Virtuality
AUI	Auditory User Interface
BCC	Body Centered Cubic (Lattice)
BEM	Boundary Element Methods
BRDF	Bidirectional Reflection Distribution Function
CAVE	Cave Automatic Virtual Environment
CC	Cubic Cartesian (Lattice)
CD	Compact Disc
Cg	C for Graphics
CPU	Central Processing Unit
DOF	Degree Of Freedom
DSP	Digital Signal Processor
DVD	Digital Versatile Disc
EAX/EFX	Environmental Audio/Effects Extension
eg.	*exempli gratia* For Example
FCC	Face Centered Cubic (Lattice)
FBO	Framebuffer Object
FEM	Finite Element Methods
ie.	*id est* That Is
(I)FFT	(Inverse) Fast Fourier Transformation
FIR	Finite Impulse Response
fps	Frames Per Second
GLSL	Graphics Library Shading Language

GPGPU	General Purpose computations using a Graphics Processing Unit
GPS	Global Positioning System
GPU	Graphics Processing Unit
GUI	Graphical User Interface
HCI	Human Computer Interaction
HiFi	High Fidelity (Audio)
HLSL	High Level Shading Language
HRIR	Head-related Impulse Response
HRTF	Head-related Transfer Function
IR	Impulse Response
MR	Mixed Reality
NPR	Non-photorealistic Rendering
NRS	Non-realistic Sound Presentation
OpenAL/EFX	Open Audio Library plus Environmental Effects Extension
OpenGL	Open Graphics Library
OpenSG	Open Scene Graph
PC	Personal Computer
PCIe	Express Peripheral Component Interface
PCM	Pulse Code Modulation
PSP	PlayStation Portable
RIR	Room Impulse Response
SDL	Simple Direct Media Layer
SLI	Scalable Link Interface
SPSS	*Statistical Package for the Social Sciences* – A statistical data analysis software.
UI	User Interface
UMD	Universal Media Disc
UML	Unified Modeling Language
VAE	Virtual Auditory Environment
VE	Virtual Environment
VR	Virtual Reality
VRML	Virtual Reality Markup Language
VRPN	Virtual Reality Peripheral Network
VTK	Visualization Tool Kit
WiFi	Wireless Fidelity – Wireless Local Area Network
XML	Extended Markup Language

INDEX

DECLARATION

I herby declare that this thesis was complied solely by myself and only with the the references marked and cited.

Magdeburg, September 2008

Niklas Röber