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

# 9 CASE STUDIES

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

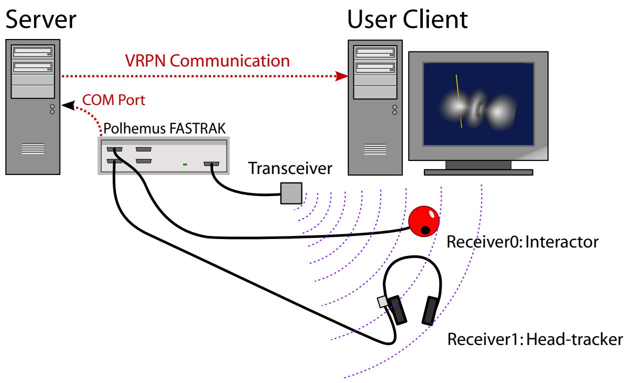
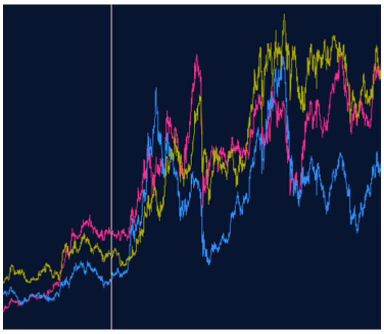



Figure 70: User Evaluation Setup.

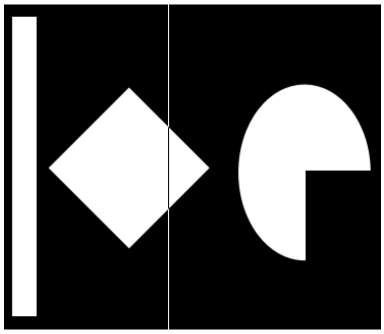
The tasks for this evaluation are discussed and explained in the following two sections. The setup for this evaluation can be seen in Figure 70. Overall, the following equipment was employed in the evaluation:

- 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*

*Stock Data Sonification.*

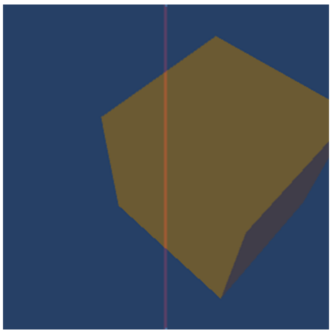
For several data sets and applications, sound and acoustics provide a more intuitive way to represent and understand the inherent information. An example that was employed in previous discussions is the sonification of stock market data, see also Figure 71a and the demonstration on the left. While the 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.

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.

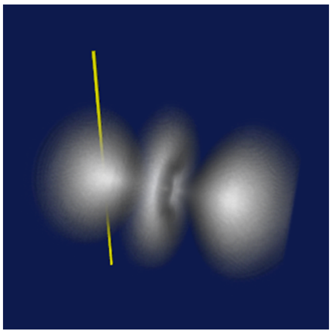
The second and the third assignment in this evaluation were the sonification of stock market data, as well as an acoustic representation of several 2D shapes. Two examples that were used can be seen in Figure 71. Figure 71a shows an example of the stock market data, while Figure 71b shows several 2D 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.*

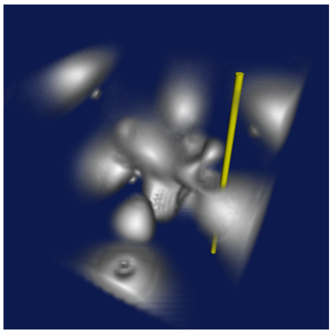
The third assignment was concerned with the identification of several 2D shapes using the acoustic scanline technique that was introduced in Section 5.3.1. Figure 71b shows an example of the 2D shapes to be identified, which almost all participants answered correctly, see also the demonstration on the 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 |

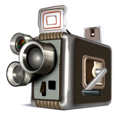
Table 10: Data and Volume Sonification Results.

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.

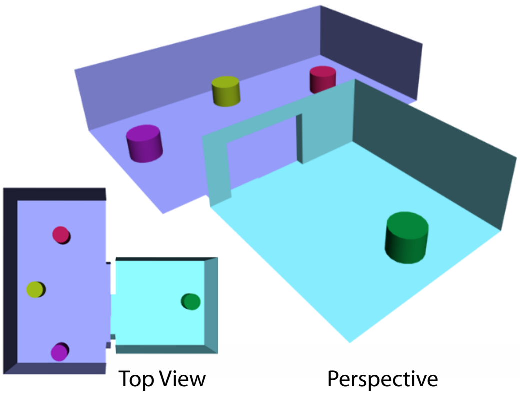
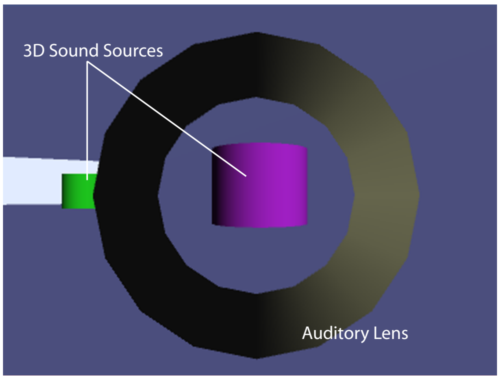


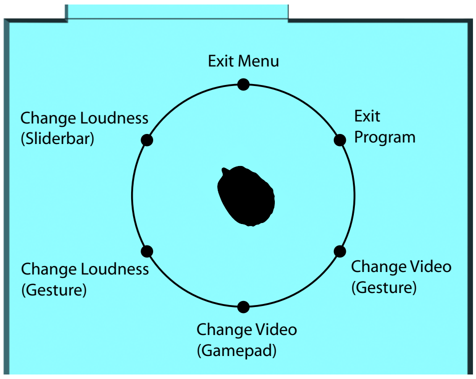

Figure 73: The Sound Stage.

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:

- 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.*

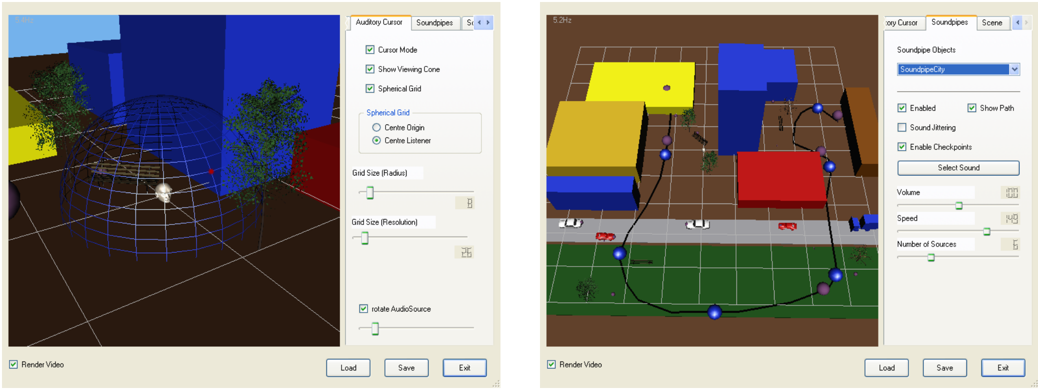
The second task included the exploration of differently complex scenes with and without the assistance of the auditory lens system, refer to Figure 74a. Out of 14 participants, 10 declared that both, orientation and navigation were easier to perform with the auditory lens system, while 4 stated that 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.

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.

*Ring-based Menu System.*

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.

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

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*.

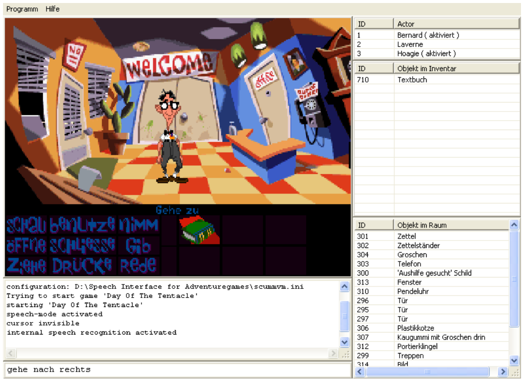



Figure 76: Speech-based Gameplay.

### 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)

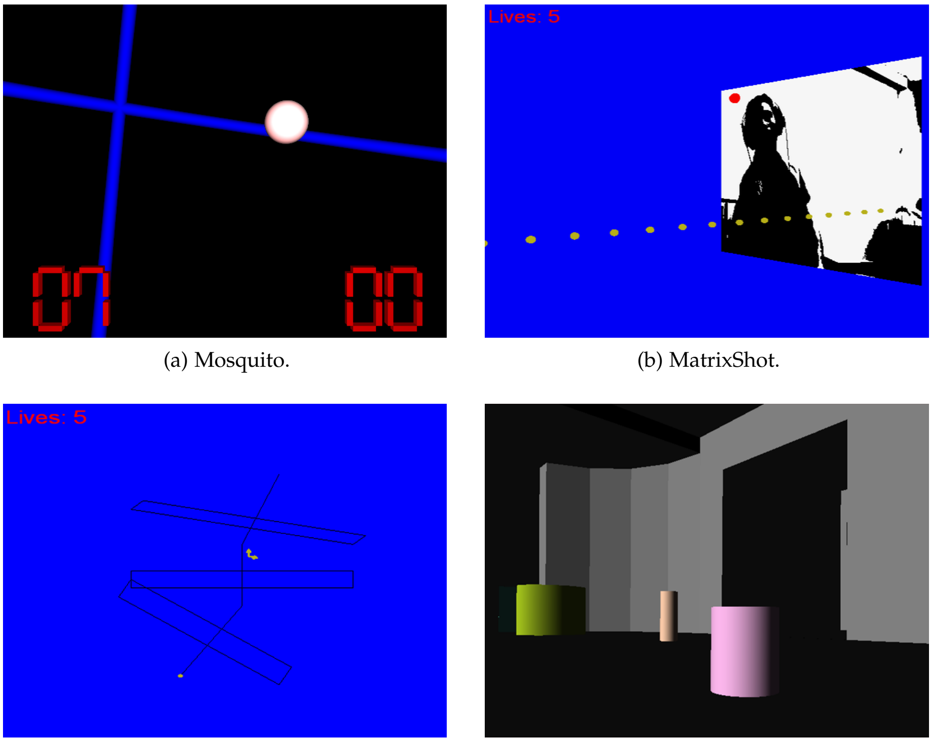
The results of this evaluation are discussed throughout the following two sections and are summarized in Table 12 and in Section B.3. A play of the regular audiogames requires a standard PC only, while the four games that are based on the audio framework demand additional hardware to perform the spatial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

*Video of "AudioFrogger".*

*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.

*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.

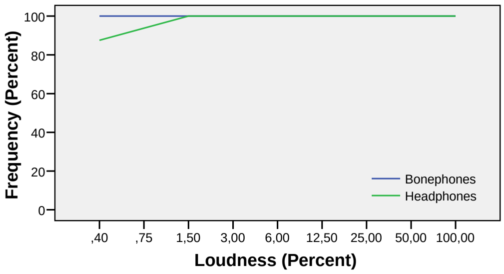
A large portion of the discussions in Chapter 6 was centered around suitable techniques and technology for a combined presentation of real and virtual auditory environments. The focus was centered around a presentation that allows the perception of both environments as one, and described the use of so 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*

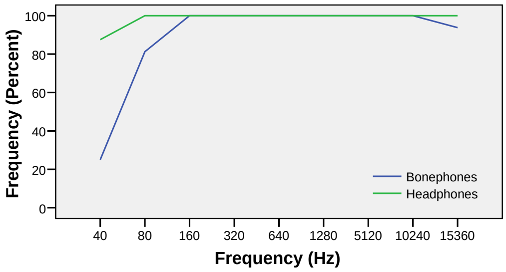
An inherent component of the augmented audio reality system devised in Chapter 6 is the presentation of the virtual auditory environment using so called *Bone-conducting Headphones*. Such bonephones have multiple advantages, but the perception of sound via skin and bone also has several drawbacks. Bonephones 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).

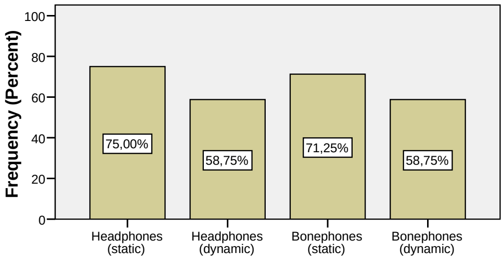
1 http://www.dowumi.com



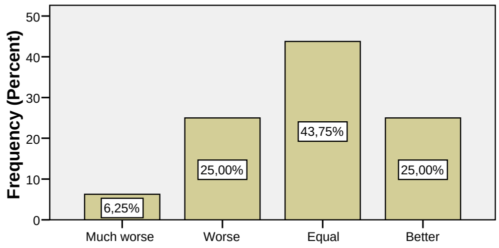

(a) Perception of Loudness.




(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.

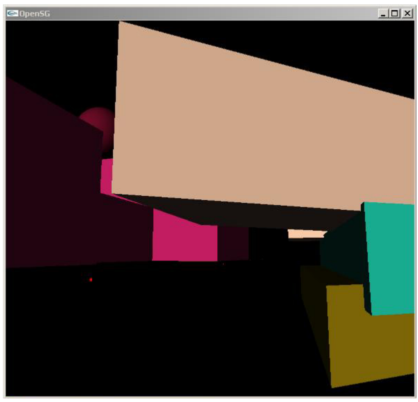
### 9.5.2 *Examples and Applications*

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.

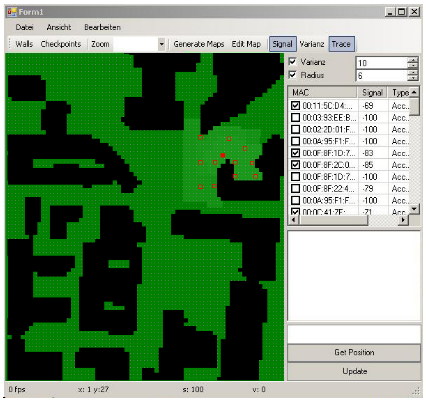
Figure 80 provides an overview of both example scenarios. The first example represents an assistance and training simulation for the visually impaired and is located on the campus of the University of Magdeburg, while the second example is an augmented audio reality implementation of the story of *"The 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*

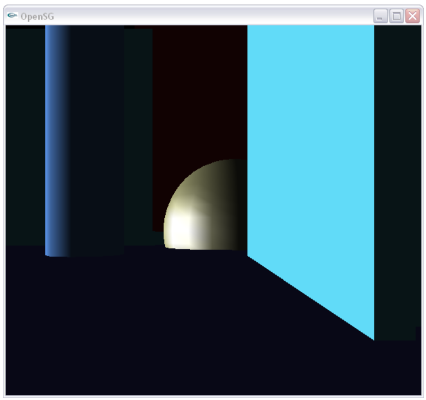
The first example represents a training simulation for the visually impaired that is set on the campus of the University of Magdeburg. The major objective was to devise an application that allows an acoustic enhancement of the University campus to improve the navigation and orientation of visually impaired 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



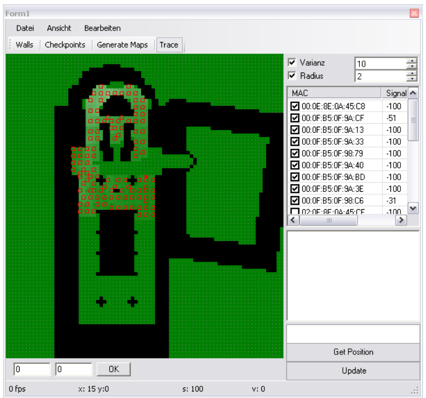

(a) Campus Navigation – VR View.




(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 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*.

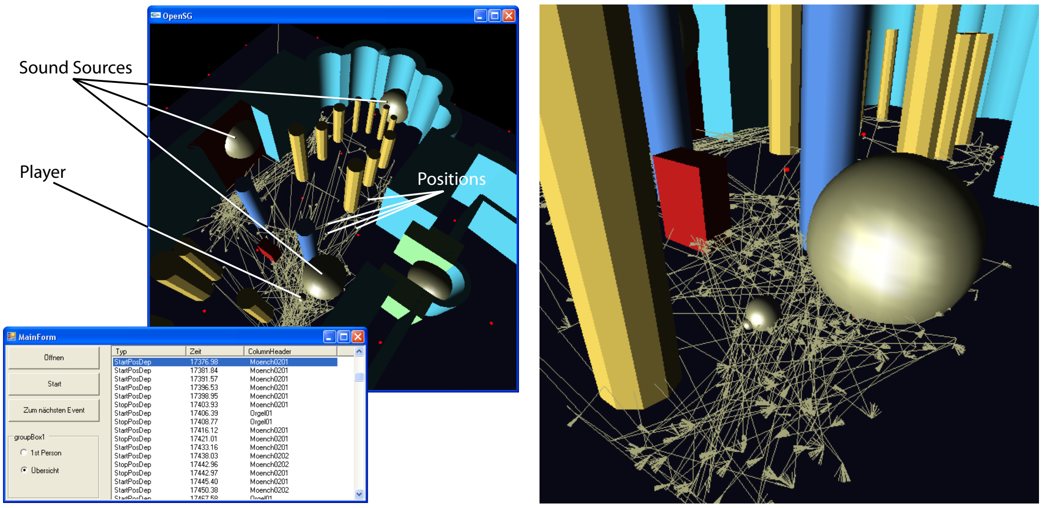
The evaluation of the AAR system required the largest and most complex setup. The radiomap that was employed for the WiFi-based user positioning was measured in advance to allow a quick start of the actual evaluation. The hardware that was employed for this evaluation was:

- 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

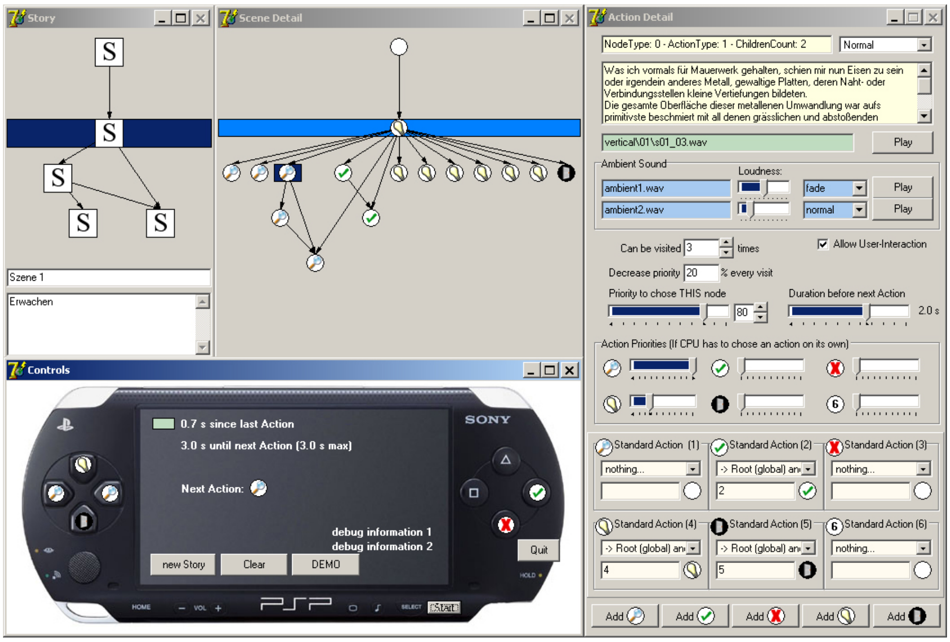



Figure 83: Interactive audiobook authoring environment.

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).

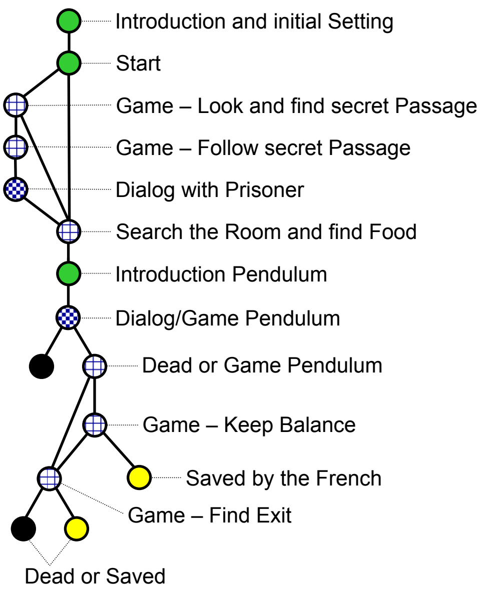



Figure 84: Simplified Story-graph.

*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.

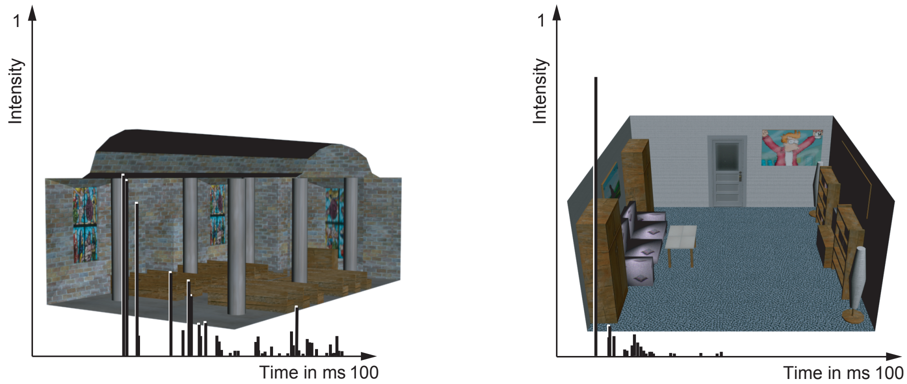
The story in this new example is very similar to the story employed in two previous examples, refer to Section 9.4 and Section 9.5. However, the here employed story arc contains additional narratives that is not available in the other implementations. A comparison with the other forms of presentation 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.

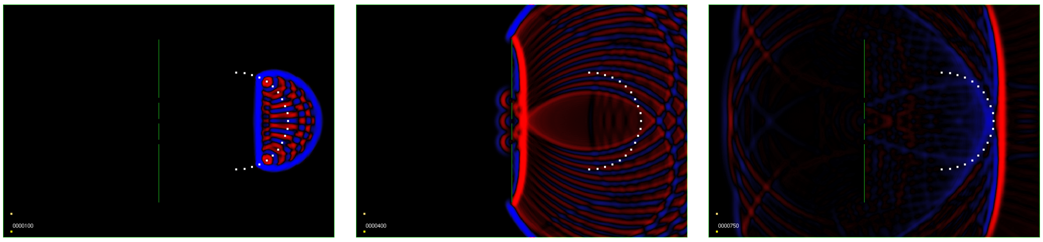
which is already available as the *PAudioDSP* API developed by Stockmann (Stockmann, 2007). An interesting project was here described in 2008 by Brode et al., in which listeners experience the narration of fictional stories at real locations in Berlin (Brode et al., 2008). Their system is based on a 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.

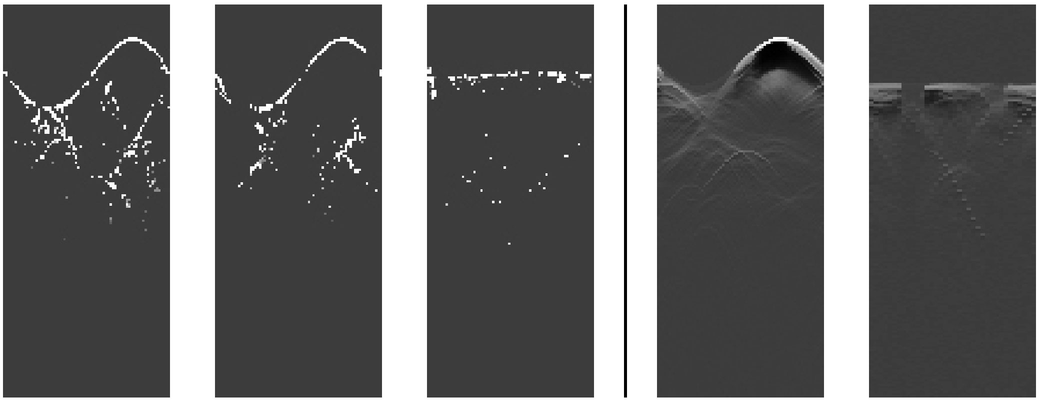
### 9.7.1 *Room Acoustic Simulations*

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).

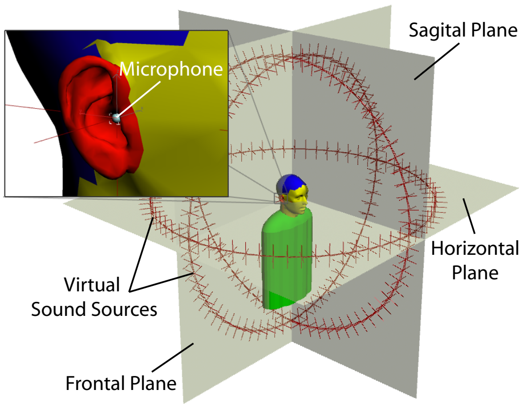
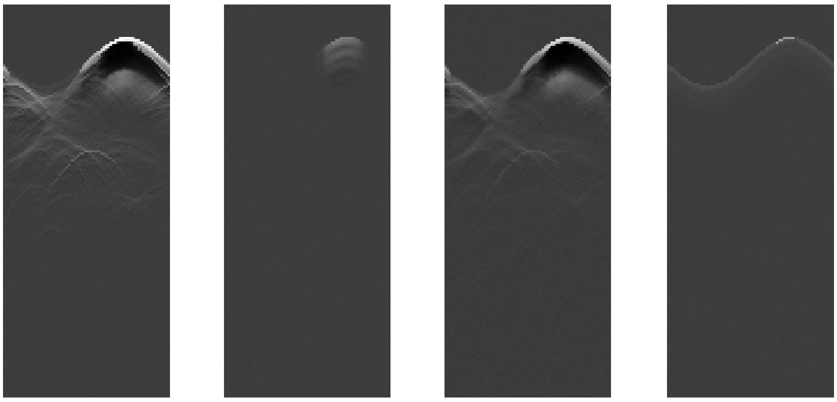



Figure 88: HRIR Simulation System.

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.