10

After this, we will all have to live a little differently. — Rainer Maria Rilke[1]

CONCLUDING REMARKS

In this thesis, a number of ideas for an INTERACTION WITH SOUND were discussed, analyzed and illuminated from various perspectives. Several examples have been provided to emphasize the discussions, as well as to evaluate the results achieved. Concluding this analysis, this last chapter serves as a summary for the work, discusses open problems and current limitations, as well as provides several possibilities for future improvements. After the research in this thesis, the exploration of 3D virtual auditory environments will continue, but with new directions set and additional goals provided.

10.1 SUMMARY

The research in this thesis examined 3D virtual auditory environments and explored several associated areas of application. After a short introduction and motivation of the topic in Chapter 1, Chapter 2 started with an in-depth analysis of the subject matter and discussed the related areas of research. In a first step, several hypotheses were devised and a schedule of the research was developed. Both Chapter 3 and Chapter 4 discussed several required fundamentals, as well as related and existing research in the areas of sound & acoustics and auditory display systems. Chapter 3 provided a broad perspective on the entire area and discussed topics ranging from sound synthesis, propagation and perception, towards an employment of sound in entertainment computing. The succeeding Chapter 4 continued this discussion, but with a more focussed perspective on auditory display systems, and here especially on 3D spatial auditory displays along the established standards and applications.

Several conclusions towards the research goal could already be drawn from these initial discussions, and led towards a definition and design of 3D virtual auditory environments in Chapter 5. This chapter exclusively focussed on the modeling of 3D auditory environments, as well as on the techniques required for a sonification of 3D scene information and spatial interaction. Starting out with a research objective in the design of a 3D auditory display system that supports an efficient and intuitive perception, quickly methods for a non-realistic auditory display of 3D auditory environments moved into focus.

Proposition 1 *A* ***Non-realistic auditory Display*** *is essential for an efficient and intuitive auditory presentation of a 3D virtual auditory environment. For that, the display is altered towards a non-physically based acoustic representation of a 3D scene and the objects therein. This is achieved by integrating additional virtual sound objects, by an exaggeration and/or reduction of certain physical parameters/laws, as well as through the use of situation-based auditory display styles for object sonification.*

Despite this non-realistic approach in the auditory display, the majority of the applications discussed are based on and exploit a physically correct 3D sound spatialization, as well as 3D head-tracking techniques for an intuitive perception and 3D scene interaction. 3D sound spatialization is essential for the display of 3D virtual auditory environments, as it provides directional and distance cues for localized sources. Additionally, it enhances the segregation between several sound sources and streams, and thereby improves the

1 upon seeing Constantin Brancusi's 1919 sculpture *Bird in Space*

overall perception of the 3D environment. Head-tracking mimics as a technique a natural human listening behavior, which further improves the perception and evaluation of 3D sound sources.

Proposition 2 ***3D Sound Spatialization and Head-Tracking*** *are key elements for the display and the interaction with 3D virtual auditory environments. Both concepts enhance the perception of 3D auditory spaces and represent imperative techniques that are required for an adequate interaction/sonification of 3D virtual auditory environments.*

A large portion of Chapter 5 was dedicated to the exploration of suitable sonification and interaction techniques to convey information from a virtual auditory scene to the user, as well as to input information. Chapter 5 started with an examination of abstract data sonification techniques and devised methods for a global and local 3D scene sonification. Sonification and interaction are bonded and require each other in order to derive/input information from/into a 3D scene.

Proposition 3 *A task-dependent* ***Sonification and Interaction Design****, that is customized to the specific requirements of an application is required for all applications. Techniques of 3D scene sonification are thereby used to convey abstract information of a virtual 3D scene and the objects therein, while interaction techniques allow an input of information into the virtual environment. The techniques employed must enable an adequate interaction, orientation, navigation and wayfinding, and thereby convey local and global information of the auditory environment. The sonification techniques should be based on a non-realistic auditory design and aim at the most intuitive display of an environment's semantics.*

Utilizing this information, the last section of Chapter 5 was dedicated to the design and development of an audio framework, suitable for an evaluation of this research. The design of this framework was based on the sonification and spatial interaction techniques discussed, as well as employed 3D sound sources and user tracking technology. This section also discussed several areas of applications, of which some examples have been implemented and examined in Chapter 9.

Chapter 6 extended these concepts and techniques towards a design of an augmented audio reality framework. The discussion started with the requirements for such a system, and later explored spatial interaction techniques, as well as methods for a combined display of a real-world location that is augmented by a 3D virtual auditory environment.

Proposition 4 ***Augmented Audio Reality*** *describes a system and techniques that support an extension of a real-world environment using additional auditory information. The underlying 3D virtual auditory environment must be synchronized with the real location in terms of position, orientation and time. Further requirements for such a system are a non-realistic 3D auditory display, efficient techniques for user-orientation and -positioning, as well as methods for spatial interaction. The system itself is realized as a wearable computer that requires an implementation on mobile and lightweight hardware.*

A spatial interaction design allows an intuitive interaction with virtual environments based on a natural real-world interaction behavior. 3D sound spatialization and spatial interaction techniques are both required for an adequate representation of 3D auditory environments. Chapter 5 introduced the basic concepts, which can, due to a direct applicable spatial mapping, be used for an interaction with augmented auditory environments as well.

Proposition 5 *A* ***Spatial Interaction Design*** *mimics a real-world interaction behavior and provides a more intuitive interface for an interaction with 3D auditory environments. The* *techniques require additional user-tracking equipment, which measure the user's orientation and position and translates this information into virtual 3D scene interaction techniques. A variety of spatial interaction designs can be employed, ranging from 3D gestures and 3D pointing towards real object interactions. The techniques are related to the application's task and the tracking technology available.*

Several possibilities for an efficient and low-cost design of user-orientation and positioning techniques have been discussed and implemented. The augmented audio reality system developed is based on a WiFi-enabled user-positioning, a digital compass for 3D head-tracking and gyro-technology to implement various 3D interaction designs. Using this system, two examples have been prototypically implemented and were examined in more detail in Chapter 9.

Besides the actual techniques for an interaction with 3D virtual/augmented auditory environments, also authoring and design are of high importance and were discussed in Chapter 7. This chapter first explored common principles for the design of both 3D virtual and augmented auditory environments, and derived therefrom several authoring guidelines and design principles. These guidelines have been implemented within an additional 3D authoring environment, which was also deployed as authoring system in a variety of example applications.

Proposition 6 *Techniques of* ***Authoring and Design*** *assist the user in the design and setup of 3D virtual auditory environments. The authoring process includes the creation of 3D geometry, the design of sound, speech and music samples, the definition of dependencies and a selection of suitable sonification and interaction techniques, as well as the actual auditory scene authoring through the definition and setup of auditory textures. During the authoring and auditory scene design, one must adhere to an appropriate balance between a display's functionality and its aesthetic appearance.*

Both Chapter 5 and Chapter 6 expressed the importance of high-quality, yet efficient methods and techniques for 3D sound spatialization and a simulation of environmental acoustics. Using the current state of the art audio APIs, these requirements can only partially be fulfilled. Therefore, Chapter 8 examined sound and light wave propagation principles, and discussed the possibilities of using efficient computer graphics rendering techniques and hardware to aid 3D sound rendering and simulation. The chapter especially concentrated on the development of a GPU-based sound rendering for both, room acoustic simulations and 3D sound spatialization. A second focus of Chapter 8 was centered around the development of personalized HRIR filters to remedy several artifacts introduced by standard HRIR filters. Although the proposed techniques were not fully integrated into the 3D audio framework, it could be shown in the concluding analyses of both Chapter 8 and Chapter 9 that the devised concepts and techniques indeed largely enhance the quality and efficiency for 3D sound rendering and simulation.

Proposition 7 ***Graphics-based acoustic Simulations*** *greatly improve both quality and efficiency of 3D sound rendering and acoustic simulation. Several similarities between sound and light propagation can be exploited to exploit dedicated computer graphics hardware for an expedited and improved sound simulation. Graphics-based sound simulations can also be used for the simulation and measurement of individual HRIR filters to improve the perception of 3D sound sources and 3D virtual auditory environments.*

A number of applications and example scenarios were discussed and analyzed throughout the research of this thesis. Chapter 9 summarized the majority of these examples and presented them in a chapter- and topic-overlapping manner. Chapter 9 evaluated techniques for 2D and 3D data sonification, analyzed methods to interact with and display information within 3D virtual auditory environments, as well as examined several very specific applications, such as an augmented audio reality system, and audio-only computer games with an introduction of the concept of interactive audiobooks.

A central role in this thesis research is the evaluation of 3D auditory environments for entertainment and edutainment purposes. As a result, the audio framework developed was primarily designed with entertainment and edutainment applications in mind. A large section of Chapter 9 was therefore dedicated to the analysis of audio-only computer games. This section compared four own developments with existing audiogames in terms of playability, immersion, complexity and enjoyment. The audiogames developed were implemented on top of the audio framework and utilized the techniques for 3D scene sonification and spatial interaction.

Proposition 8 *An* ***Audio-centered Gameplay*** *concentrates on the specifics of an auditory perception and proposes a new way of play and interaction with audio-only computer games. Using adequate scene sonification and interaction techniques, the content and the story of an audiogame can be perceived with a high level of immersion. These characteristics make this form of presentation very suitable for a display of narrative content, ie. adventure-based computer games. Audiogames benefit effectively from the use of 3D spatial interaction techniques, as well as from a general employment of 3D sound spatialization in combination with user head-tracking. Similar to audio/visual computer games, also audiogames benefit from a simple and clear design, which in this respect focusses on an auditory perception and an audio-centered gameplay.*

Another very interesting application that was introduced in Chapter 9 are so called *Interactive Audiobooks*. They combine the immersive advantages of an auditory storytelling (ie. based on audiobooks and radio plays) with interactive elements from adventure based computer games. Further advantages of interactive audiobooks are a non-linear storytelling, as well as a varying degree of interaction. This allows to reexperience the same or a similar storyline, which can be either perceived passively as a regular audiobook, or actively as an audio adventure game. This becomes possible through the use of a story-graph structure that consists of narration and interaction nodes, and allows the construction of several different storylines with various endings.

Proposition 9 ***Interactive Audiobooks*** *combine the advantages of an immersive, non-linear storytelling with interactive elements from adventure-based audio-only computer games. The result is a highly immersive presentation of narrative content, which becomes even more effective through the added interaction. The level of interaction can be varied smoothly, which allows a free blending between a passive audiobook and an interactive adventure computer game. Interactive audiobooks can be played and controlled using only few commands that influence the main characters conduct, which is also responsible for controlling the audiobook in a non-interactive mode.*

The last section of Chapter 9 evaluated and analyzed the results from the graphics-based acoustic simulations that were developed in Chapter 8. An emphasis in this analysis was an application of these techniques for general room acoustic simulations, as well as especially a discussion of virtual HRIR simulations using ray-based sound simulations.

10.2 DISCUSSION

After this summary of the thesis and its research, the following section discusses the contributions of this work in respect to the individual research areas, as well as critically reflects the results achieved.

10.2.1 *Contributions*

The research that has been presented in this thesis has contributed to a variety of scientific areas and was published in a number of conference articles. The two major contributions are a redefinition and extensive analysis of 3D virtual auditory environments, as well as the research for a more efficient, graphics-inspired sound rendering and simulation.

2D/3D DATA SONIFICATION — The area of 2D/3D data, image and volume sonification has been advanced through a refinement of existing sonification techniques, as well as through an added spatial sonification for the exploration of 3D objects and volumetric data sets. It could be shown that an added spatialization and rhythmic sequencing improved the parallel perception of linear data (stocks), and that 3D interaction improved the understanding of 3D shapes and the topology of 3D volumetric data sets (Stockmann et al., 2008).

3D VIRTUAL AUDITORY ENVIRONMENTS — Due to ambiguities in its definition, the term *3D virtual auditory Environment* has been redefined in the confines of virtual reality and 3D auditory display systems using an abstract definition of VR/MR environments. The new definition focusses on an audio-centered design, as well as employs a non-realistic auditory scene description. A number of 3D scene sonification and 3D spatial interaction techniques were devised, implemented and their applicability also evaluated, including the concepts of an auditory cursor, -guides, -landmarks, -lens, sonar/radar and soundpipes system. Furthermore, the concepts of dependency modeling and auditory textures were devised and implemented to allow a broad spectrum for an interaction design (Röber and Masuch, 2004b, 2005b,a, 2006).

INTERACTIVE AUDIOBOOKS — The concept of *Interactive Audiobooks*, which combines the advantages of a non-linear narration with interactive computer game elements, has been developed and implemented. In two user evaluations, the concepts functionality could be confirmed and a high level of immersion was shown (Röber et al., 2006b; Huber et al., 2007).

AUGMENTED AUDIO REALITY — The concept of augmented audio reality has been advanced in terms of 3D spatial interaction, as well as through an evaluation of new areas of application. A low-cost, yet efficient system has been devised and implemented, and was employed in a user-guidance scenario targeting the visually impaired, as well as in a narrative augmented audio reality game (Röber et al., 2006a).

AUTHORING AND 3D SCENE DESIGN — Further contributions were made through an analysis of 3D virtual/augmented auditory environments and the development of 3D scene authoring and design techniques. An additionally implemented 3D authoring framework demonstrated the developed guidelines and principles in practice (Röber and Masuch, 2004a).

3D SCENE AURALIZATION AND SOUND RENDERING — Due to the high demands in acoustic realism, additional research was spent in the exploration of more efficient 3D sound rendering and simulation techniques. A number of graphics-inspired methods were discussed and implemented, including a technique for general GPU (sound) signal processing, as well as two systems for a GPU-centered ray- and wave-based sound simulation. Furthermore, advancements in the direction of virtual HRIR simulations for the creation of personalized HRTFs were discussed and developed (Röber et al., 2006,c, 2007; Röber et al., submitted).

10.2.2 *Critical Reflections*

One of the major differences between the research on auditory display systems and visual (computer graphics) display techniques is the availability of related work and the amount of research conducted. As we primarily live in a visually-centered environment, more research has been accomplished in the visual domain. Although the area of auditory display has left its infancy many years ago, the awareness of its potential, also in respect to new areas of application, is still underdeveloped. The research that was conducted in this thesis explicitly focussed in this direction and discussed and advanced 3D auditory display systems, both in technology, as well as in areas of application. Although the techniques that were developed to perform a 3D scene sonification and interaction are applicable to a broad variety of applications, a slight focus was placed in this research on entertainment and edutainment tasks.

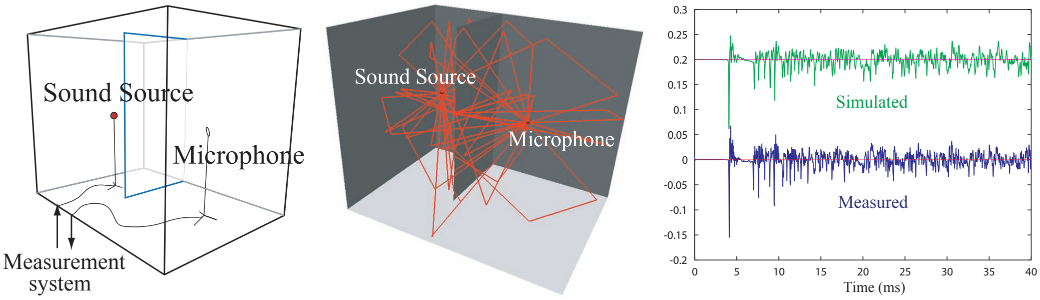
Chapter 9 discussed a broad spectrum of different applications that were prototypically implemented and evaluated. Selected applications and proposed techniques have been analyzed in more detail using user evaluations. Although the main goal of the research was not specifically the development of techniques to aid the visually impaired, the techniques discussed are, nevertheless, very applicable and useful in this domain. The focus on this specific group of users could have been stronger in specific evaluations, see also Appendix B. As this would also have shifted the center of this research, these analyses are left for future development.

During the development and implementation of the techniques and applications, several approximations had to be applied. These approximations include the realism of the sound rendering and synthesis techniques used, as well as in certain cases the design and the presentation of the auditory display systems themselves. The chosen applications and example scenarios focus on a single user presentation and interaction. The binaural sound rendering that is employed is – due to the use of 3D head-tracking – generally only applicable for a single person perception. Although a multi-user presentation and interaction environment can also be realized using the audio framework developed, but this would require major modifications in terms of sound rendering and especially the display of content. The spatial interaction techniques, however, are directly applicable in such a setting as well.

The audio framework that is used as basis in the majority of applications was developed based on OpenAL/EFX for sound spatialization and rendering. This API, however, is limited to standardized HRTF filters only, as well as uses several severe approximations for the simulation of room and environmental acoustics. Within the conducted user evaluations, several participants claimed to have difficulties in the localization of virtual 3D sound sources, which, no doubt, resulted from the use of generalized HRTFs. Chapter 8 examined therefore the possibilities of using computer graphics hardware and rendering techniques to improve the sound rendering process in terms of quality and efficiency. Very promising results were achieved, but could not yet be integrated into the audio framework itself. This future development could, through an increased perception of the 3D auditory scene, shed new light on the performances of the developed 3D scene sonification and interaction techniques as well.

10.3 FUTURE IMPROVEMENTS

A thorough research poses often more (*new*) questions than it is able to answer. The research within this thesis has been extensive, yet several areas and ideas remain for future development and improvement. Specific possibilities for future improvements



(a) System Setup. (b) Simulation. (c) Comparison.

Figure 90: Evaluation of Acoustic Rendering Techniques using the *Bell Labs Box* (Tsingos et al., 2002).

have already been outlined and discussed previously in the respective chapters. This final section summarizes the most interesting and promising areas of future research, and provides directions for further improvements. The discussion is thereby divided into aspects of interface design, technical improvements and further areas of application.

Interface Design

The last section already started a discussion on important interface related issues, such as multi-user presentations and interactions, but also a stronger focus on the requirements of the visually impaired. This not only requires additional research regarding sound rendering and the auditory display of 3D virtual environments for group perception, but also includes different and modified applications to enable group interaction. An interesting direction is here, which is also applicable to single-user systems, a perceptual presentation. This technique, similar to the MP3 music format, only considers those sources and acoustical effects that are audible in the final presentation. Moeck et al. have studied a first approach for a progressive perceptual audio rendering of large and complex scenes (Moeck et al., 2007). This approach is highly applicable to the sonification of 3D auditory environments as well, and could also reduce the requirements and complexity of acoustic rendering.

But also *classic* future improvements of interface design are important, such as a detailed analysis of the individual techniques, especially in combination with the enhanced graphics-based sound simulations, as well as a generally improved user interface and a further exploration of advanced spatial interaction techniques.

Acoustic Rendering

At the time of writing, the enhanced sound rendering and simulation techniques are not yet fully integrated into the audio framework developed. Although their function and general applicability has been confirmed using several tests and evaluations within this thesis, a real-world comparison would finally reassure their performance. One possibility is here to use the so called *Bell Labs Box* as is depicted in Figure 90. This system has been used by Tsingos et al. to study virtual sound simulations and to compare them with real-world measurements (Tsingos et al., 2002). A similar approach could be used to further evaluate the graphics-based wave- and ray-acoustic simulation techniques that were developed in this thesis.

Another interesting direction of research is here the perception of sound and which propagation effects are most significant. This is especially important for the advancement of virtual HRIR simulations, and could enhance the overall 3D sound perception.

A third technique that is potentially interesting for a graphics-accelerated implementation is ambisonics. Ambisonics can be realized in graphics hardware using an implementation of spherical harmonics, which are currently applied in rare cases for global illumination effects (Dempinski and Viale, 2005).

Physics Simulations

In 2006, Ageia introduced a physics processing unit (PPU) as additional hardware for the PC system, dedicated to the simulation of real-world physics within computer games (AGEIA Corp., 2006). Although this hardware seemed potentially interesting to perform sound simulations as well, a first attempt of implementing ray- and wave-based sound simulations failed due to a less efficient hardware design and a game-centered API that only permits an implementation of *game physics* (Hugenberg, 2007). However, the availability of dedicated physics hardware is intriguing and the hardware might evolve in the same direction as computer graphics hardware has a decade ago. Until then, advantages of employing physics hardware in sound-based simulations do not exist.

Augmented Audio Reality

A large part of the research was directed to an implementation of 3D spatial interaction techniques and the design of an augmented audio reality system. An implementation of augmented reality is only possible using additional, generally very expensive, hardware for user-tracking and positioning. The proposed system that was conceptualized within this research is based on a low-cost approach and only employs commodity hardware that is broadly available. An evaluation revealed that the developed system worked very well, but the WiFi-based user positioning turned out to be a weak point of the implementation. Without the need to resort to expensive hardware, which no doubt would remedy all problems, additional sensors and technologies, such as Bluetooth, can be integrated to advance the system.

More and more everyday devices, such as mobile game consoles, PDAs and telephones, feature additional sensors that can be easily employed in the direction of an audio-centered ubiquitous computing and used in an ambient intelligence design. Auditory user interfaces and auditory display systems will no doubt play a larger role in the near future. A lot of the information that is conveyed visually today might be perceived and displayed acoustically tomorrow.

And Beyond...

Beyond all these technical questions, the perspectives and future possibilities of the applications and user scenarios discussed are very intriguing. In a web article, Kennedy wrote about the *"Bubbles of Sound in Public Space"* (Kennedy, 2007), see also Figure 91. These bubbles emerge through the large availability of mobile MP3 players, which immerse their listeners into their own personal auditory environment:

> *"... the music they're listening to (...) becomes a kind of background soundtrack to the experience of public space. Public space, or the urban environment, has become a kind of background scenery to the music we listen to rather than the primary focus of our experience."* (Mosco, 2005)

This alludes to a discussion of the social effects of music. People who are shutting each other off by plugging their ears in public spaces are achieving the opposite of what music was initially created for – a means of communication and to bring people together. Therefore, a more philosophical challenge is the question of how to unite and connect different people over sound and music together, rather than immersing themselves in their own secluded worlds. A huge influence is here the application itself and the interaction it allows, as well as the content that is displayed.

Leaving this last statement unanswered, this concludes this chapter and the research in this thesis that hopefully provides a strong and firm basis for further research and future developments.

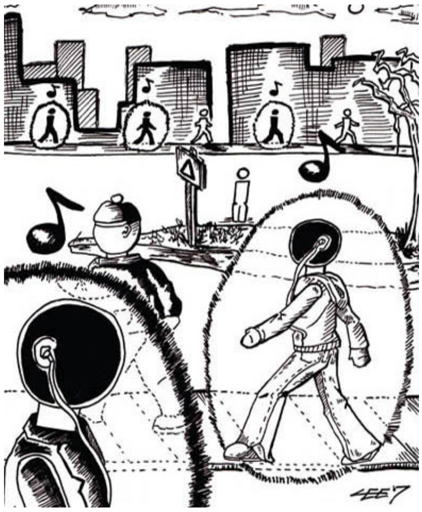


Figure 91: *"Bubbles of Sound in Public Space"* by Dave Lee (Kennedy, 2007).