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

5 AUDITORY ENVIRONMENTS

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

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

5.1 VIRTUAL REALITY AND AUDITORY ENVIRONMENTS

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

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

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

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

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

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

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

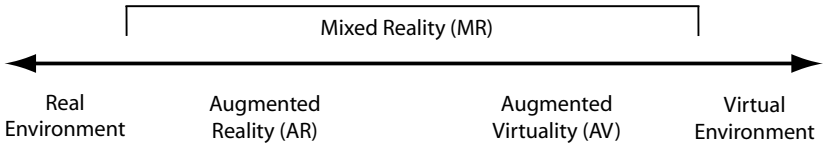


Figure 21: Reality-Virtuality Continuum (Milgram et al., 1994).

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

5.1.1 *Virtual Reality*

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

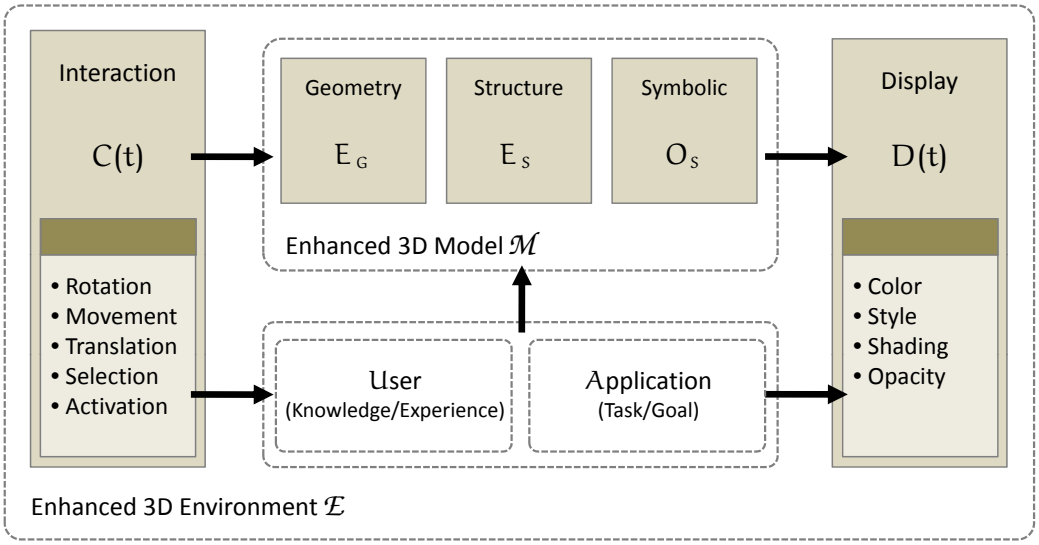


Figure 22: Formal Description for 3D Virtual Environments.

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

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

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

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

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

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

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

2 $C(t)$ = from Control

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

5.1.2 *Virtual auditory Environments*

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

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

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

The formal model of Equation 5.3 for describing virtual environments can directly be applied to define virtual *auditory* environments as well. Some modifications are, however, necessary in order to accommodate the differences in perception and presentation. This

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

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

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

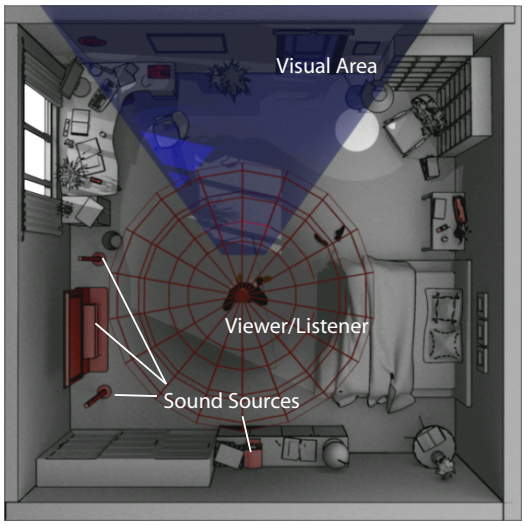
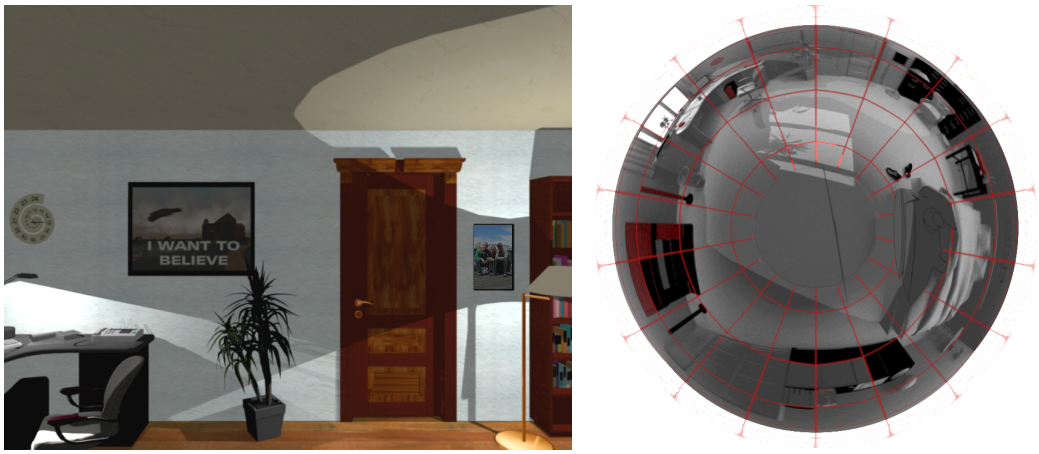


Figure 23: Different Perceptual Areas: Visual (Cone) and Auditory (Sphere).

Auralization of Figure 24b.

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



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

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

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

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

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

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

itself, the task to accomplish and the prospective user. As this discussion would lead deep into the area of software design, references to the literature are provided for a further study (Shneiderman, 2004; Faulkner, 1998).

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

5.2 AUDITORY PRESENTATION AND DISPLAY

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

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

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

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

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

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

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

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

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

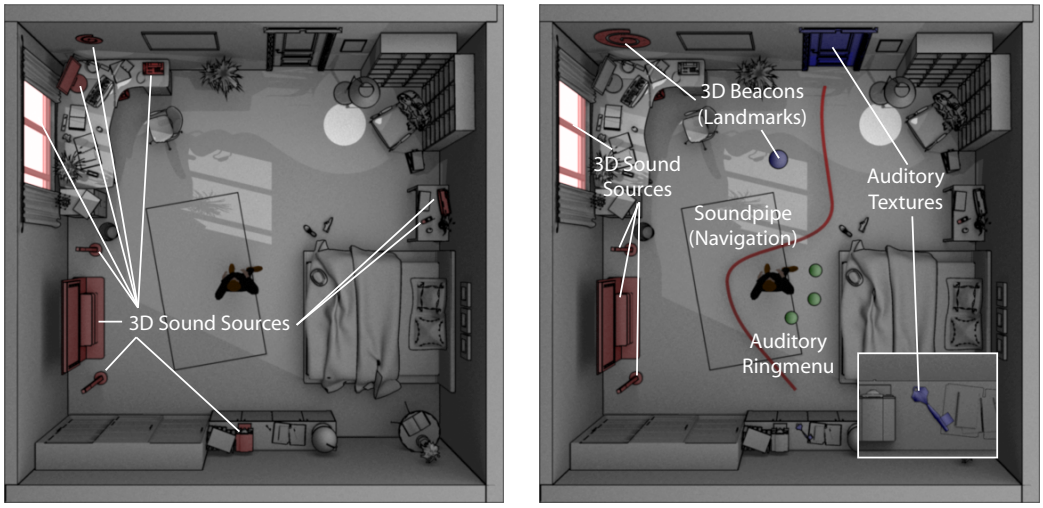
5.2.1 *Scene Auralization*

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

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

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

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



(a) 3D Scene Auralization.

(b) Non-realistic auditory Display.

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

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

Pictures at an Exhibition – Promenade.

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

Correct blending of Music.

Computer games, as a modern form of storytelling, also employ music to express emotions and to enhance the perception of the story and the game play. As this form of interactive narration is not bound by a timed schedule as in operas and film, the

transition between scenes and the blending between two pieces of music is often not predictable and can not be timed in advance. Therefore, most implementations utilize a gradual transition between two pieces, which often destroys the immersion due to the introduction of disharmonies that are caused by blending artifacts. A solution was found by examining the problem from a musician's perspective and through the development of techniques that allow a harmonic blending between various pieces of music (Berndt et al., 2006). The music is now composed and divided into several pieces, which can be blended correctly at certain points. This allows a much smoother presentation, and to maintain the listener's immersion.

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

Auralization of Figure 26a.

Auralization of Figure 26b.

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

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

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

5.2.2 *Non-realistic Auditory Display*

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

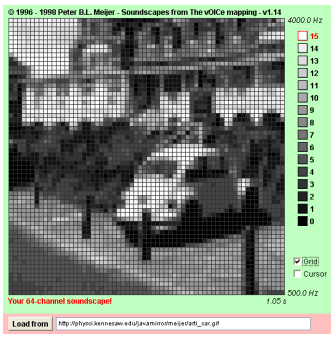
Definition *Non-realistic Sound Rendering* describes a principle for the display of 3D auditory environments that focusses on an intuitive and *non-realistic* auditory presentation of 3D scene information. This presentation is not required to adhere to the laws of physics, instead it concentrates on the most intuitive and direct presentation of information available, by employing additional sound sources and artificial cues, as well as through a deliberate alteration of an objects acoustical appearance. The appearance of objects

thereby might include different auditory representations, as well as a change of their physical attributes, such as Doppler factor, velocity, or distance. The basic fundamentals identifying such a *non-realistic* auditory presentation for a 3D scene are:

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

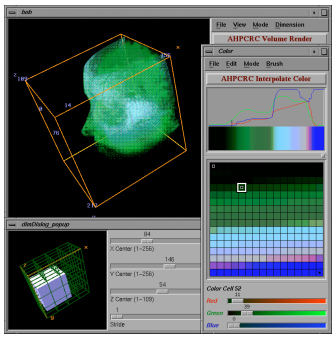
Figure 26 shows a comparison of a regular 3D scene auralization and a non-realistic auditory display of the same environment. The normal auralization applied in Figure 26a shows the familiar living room scene with a virtual avatar in its center. The majority of sound sources in this room is based on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

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



(a) The vOICe System[3] (Meijer, 2008).

(b) Radiation Counter[4] (Rutherford and Geiger, 1908).



(c) 3D Volume Sonification (Rossiter and Ng, 1996).

Figure 27: Data Sonification Examples.

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

5.3 DATA AND 3D SCENE SONIFICATION

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

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

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

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

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

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

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

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

5.3.1 *Data Sonification*

Geiger Counter.

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

An intuitive sonification is able to directly represent the semantic information and establishes a link between the underlying data and the sounds displayed (symbolic representation). Contrary to visualization, a sonification of data is always perceived over time and requires a certain amount to interpret the sounds heard. Depending on the techniques used, this time ranges from a few milliseconds to seconds and even minutes for very complex sonifications that utilize harmonics and music in their display. Sonifications that are based on physical attributes, such as loudness, pitch, or timbre, are very fast to perceive and analyze. As a result, many sonification examples employ

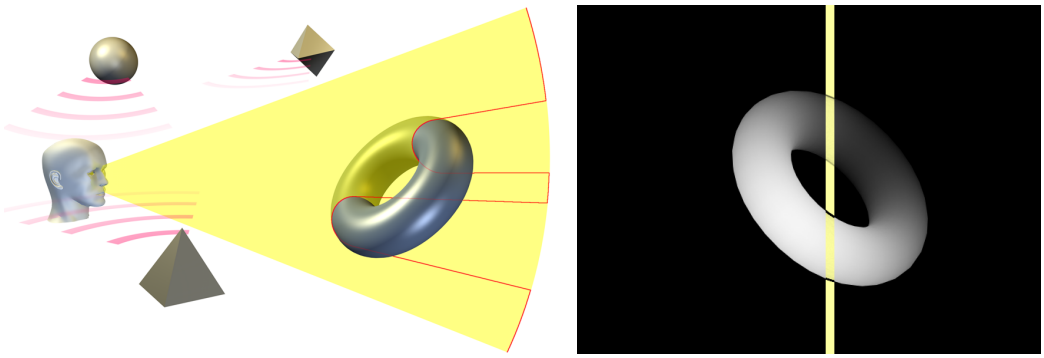
such a direct audification based on these three parameters. In some cases, however, it is also necessary to integrate an evaluation within the sonification itself; to not only display the raw data, but to additionally inform the listener about trends and assessments. An example is the sonification of stock market data, in which a different evaluation of the falling and rising of stocks may be employed depending on the ownership. Music and harmonic melodies can be easily used for such a task and display a falling stock using a minor, and a rising one with a major scale (Janata and Childs, 2004; Stockmann, 2008). Dissonances can be employed to sonify an imminent danger that requires an immediate and fast interaction (Roberts, 1986).

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

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

Shapes and Images

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



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

Figure 28: Scanline Sonification for 3D Objects.

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

Demonstration of the vOICe System.

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

2D Shape Sonification.

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

3D Objects and Data Volumes

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

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

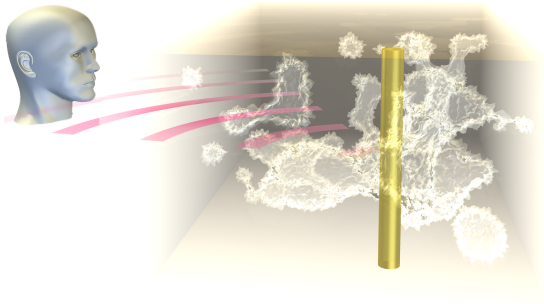


Figure 29: 3D Volume Sonification using an interactive Chimes.

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

3D Volumetric Data Sonification.

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

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

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

5.3.2 *3D Scene Sonification*

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

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

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

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

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

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

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

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

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

Table 4: 3D Scene Sonification Techniques.

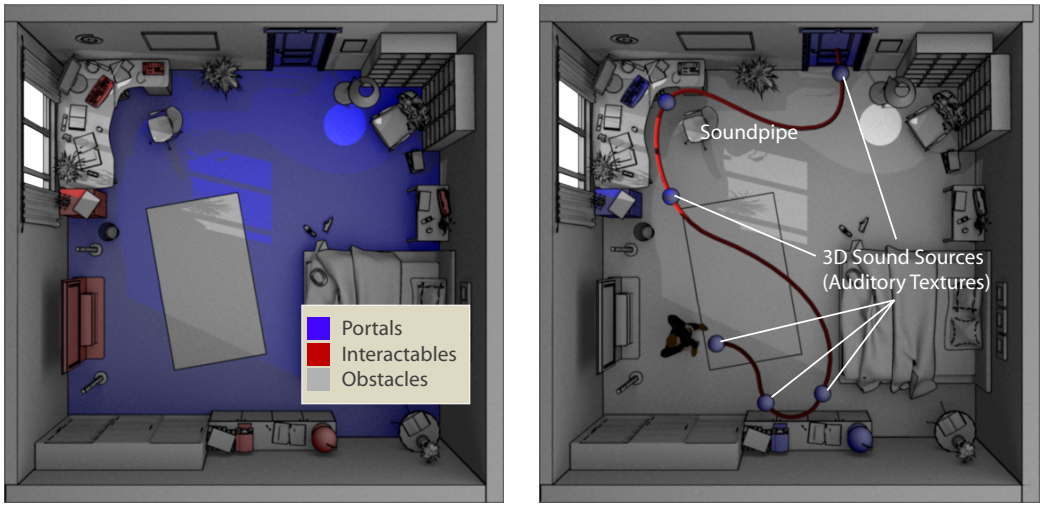
An overview of the employed 3D scene sonification techniques can be seen in Table 4. The list shows the major characteristics for each technique, such as application, presentation and perception. Application thereby describes how the technique is applied, ie. for sonifying global/environmental, or local/object-based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

Global Sonification Techniques

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

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

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



(a) Scene Object Classification.

(b) Pathway Sonification using *Soundpipes*.

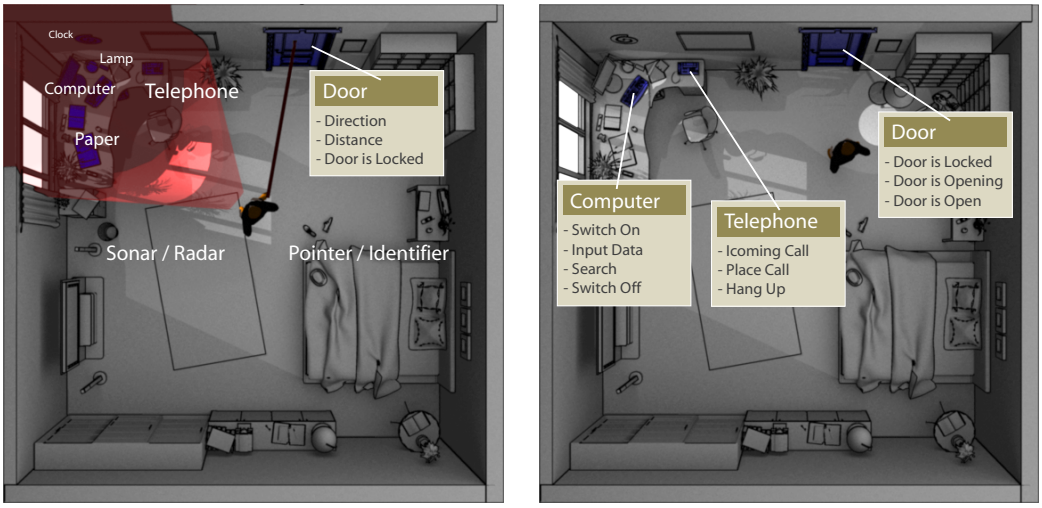
Figure 30: Global 3D Scene Sonification Techniques.

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

Auralization of Figure 30b.

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

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



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

(b) Three Auditory Textures.

Figure 31: Local 3D Scene Sonification Techniques.

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

Sound and Music Guides.

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

Local Sonification Techniques

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

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

Auralization of Figure 31a.

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

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

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

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

Auditory Texture

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

Figure 32: Auditory Texture for a Telephone.

Auralization of Figure 31b.

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

Summary

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

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

The discussed local sonification techniques are:

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

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

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

5.4 INTERACTION CONCEPTS

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

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

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

Table 5: 3D Scene Interaction Devices and Techniques.

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

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

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

5 DOF = Degree of Freedom

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

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

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

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

5.4.1 *Common 3D Scene Interaction*

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

(a) 5DT Data Glove[6].

(b) 3D Space Navigator[7].

(c) Polhemus FASTRAK[8] (Polhemus, 2008).

Figure 33: Spatial Interaction Devices.

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

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

5.4.2 *Spatial Interaction*

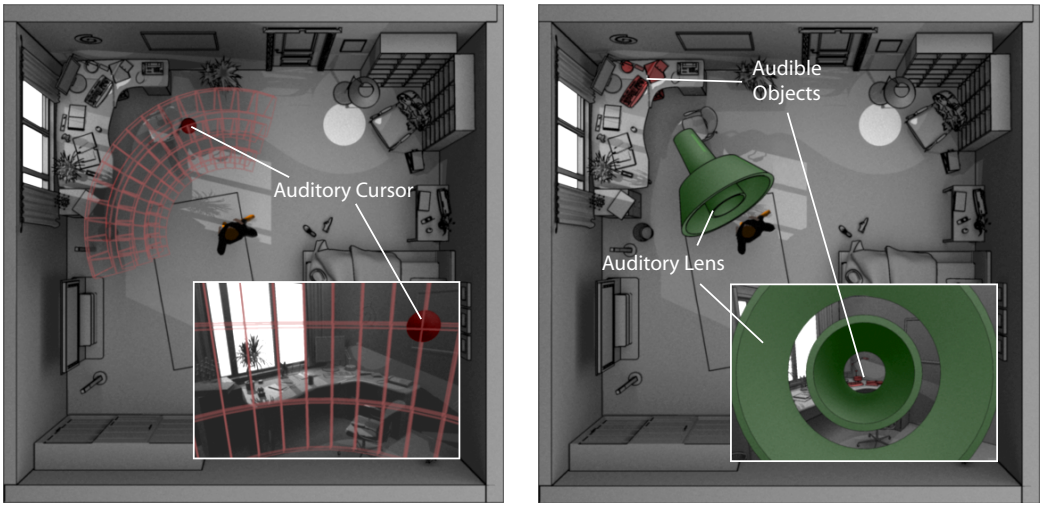
A spatial interaction design that mimics a real-world interaction is in many cases more desirable, as it more intuitively maps the interaction onto the virtual environment:

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

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

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

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



(a) Auditory Cursor and 3D Interactor.

(b) Auditory Lens.

Figure 34: Spatial Interaction Techniques.

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

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

Auralization of Figure 34a.

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

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

Auralization of Figure 34b.

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

Auralization of Figure 31a.

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

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

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

5.4.3 *Speech-based Interaction*

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

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

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

5.5 FRAMEWORK DESIGN

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

5.5.1 *Design Essentials*

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

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

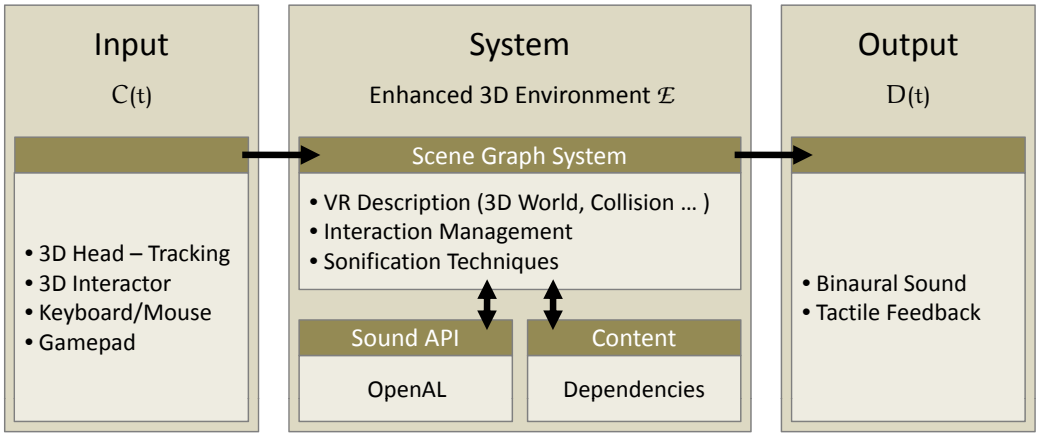


Figure 35: Framework Overview.

  - 3D pointing and interaction devices.

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

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

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

The devised system is based on standard PC hardware, as this allows an easier development and evaluation of the entire system and the single techniques. However, at several points throughout this research, alternatives that allow an implementation on mobile hardware and devices are discussed (Stockmann, 2007). Although it was shown that regular sound hardware has certain difficulties and inefficiencies with 3D sound spatializations and the simulation of room acoustics, it serves as an initial basis to gather first results. For the tracking of the user's head-orientation, as well as for the performance of spatial interactions and 3D gestures, a Polhemus FASTRAK system is employed. This is a 6 DOF magnetic field-based tracking solution that allows the use of up to four independent sensors (Polhemus, 2008), see also Figure 33c. The final binaural sound display and rendering is performed using regular HiFi headphones. The headphone systems employed consist of a regular high-quality HiFi system (*Hearo999 Audiosphere*

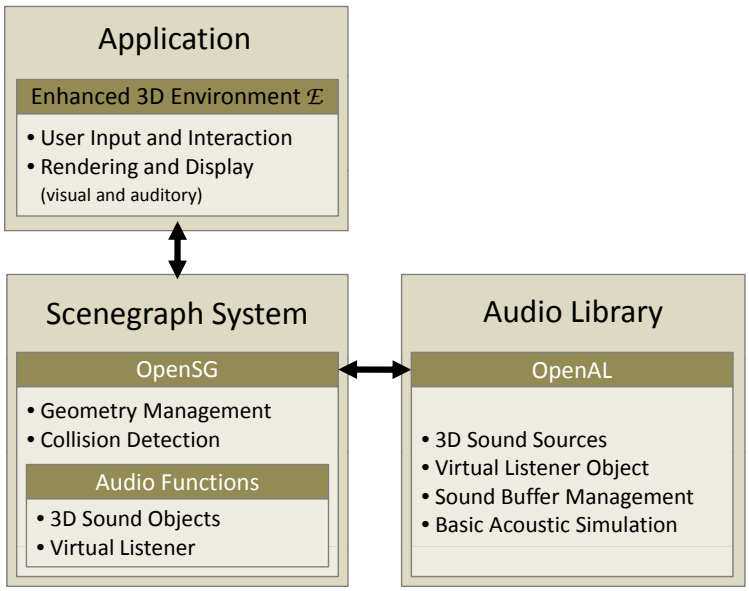


Figure 36: Framework Implementation.

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

5.5.2 *Implementation*

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

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

core system, as can be seen in Figure 36. This allows to substitute the sound rendering API by a more capable system, refer to the discussions of Chapter 8.

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

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

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

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

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

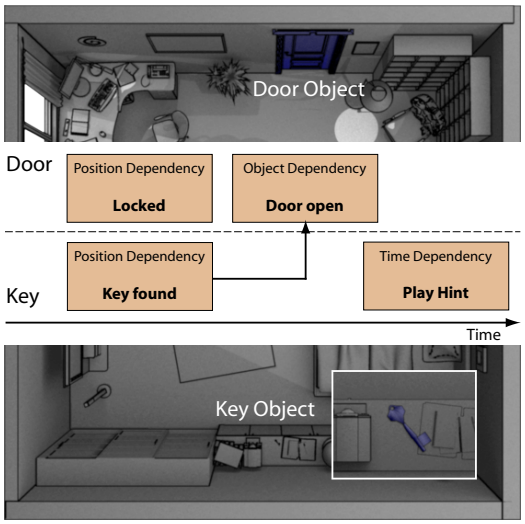


Figure 37: Interaction Dependencies.

A *Position Dependency* describes the influence of the listener's position onto surrounding objects within the virtual auditory environment. Depending on the listener's position, certain objects are audible, while others are silent. A position dependency is described through a threshold ϵ – the distance to a certain object in the scene. If the listener approaches this object and if the distance is below this threshold ϵ, a certain action is evoked. Position dependency is very general and applicable in many forms, most notably for augmented audio reality applications, which will be introduced and discussed in Chapter 6. An *Object Dependency* describes an inter-object dependency, in which one object can change the state and appearance of another object or group of objects. This dependency is very helpful for using additional *dummy* objects to model dependencies based on

the connection of the environment with other secondary systems. A *Time Dependency* is simply a counter that triggers an event after a certain amount of time has lapsed. This dependency can also be used for the modeling of additional narratives in the form of a state machine. Figure 37 displays a simple dependency graph for the door and the key object. In this familiar setting, a small part of a virtual 3D auditory computer game is described. The task of the user in this example is to find a missing *Key* object to unlock and open the *Door*. Both objects are described through a position dependency, which, if the user approaches, display their current condition. Here the door appears as being *locked*, while the key identifies itself as key to open the locked door. The unlocking itself is performed using an object dependency that unlocks the door after the key is found. If the user has difficulties finding the key, a hint is played after a certain amount of time trough an added time dependency. The concept of dependency modeling allows a very broad design and implementation of various auditory environments.

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

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

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

Another key component of the system is the connection to the various interaction devices that are employed, see also Figure 35. A low latency of the tracking equipment is thereby essential, and must be below 80ms for the head-tracking system, as otherwise perceptional artifacts occur (Brungart et al., 2005). The implemented spatial interaction techniques and 3D gestures also require a low latency ($\leqslant$ 200ms). For an application

towards music and the design of virtual instruments, the latency must be as low as 30-40ms, as otherwise a continuous and focussed play is difficult to achieve (Stockmann, 2008; Stockmann et al., 2008). For the systems connection to the tracking equipment, the VRPN library was employed, which supports several VR solutions and is controlled via a network system (Taylor II et al., 2001, 2008), refer to Figure 70. The latency of this API is low enough for a regular interaction with 3D auditory environments, but required a re-implementation due to an application for the design of virtual computer music instruments (Stockmann et al., 2008). The additional gamepad, as well as the keyboard and mouse connections to the system have been implemented in a straightforward manner using the Direct Input API from Microsoft[9].

5.5.3 *Areas of Application*

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

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

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

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

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

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

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

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

5.6 SUMMARY

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

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