3 FUNDAMENTALS

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

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

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

3.1 INFORMATION ANALYSIS AND DISPLAY

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

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

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

3.1.1 *Visualization and Information Presentation*

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

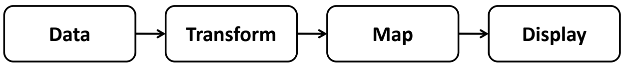


Figure 4: Visualization Pipeline.

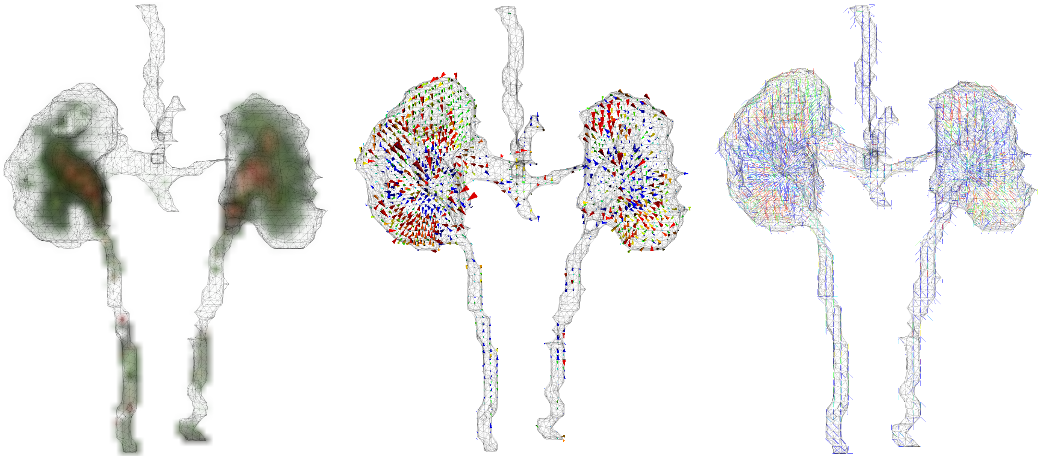
One of the critical and often most difficult steps in the visualization pipeline is the extraction of important information. Smith et al. describe it as to

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

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

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

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



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

Figure 5: 3D Visualization Examples.

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

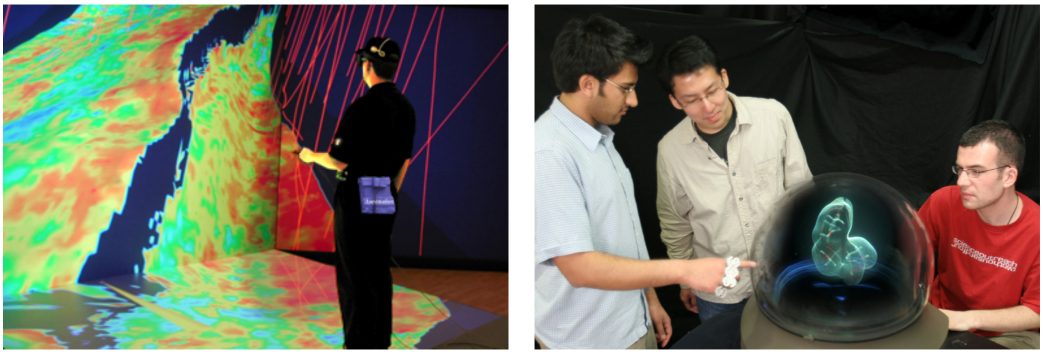
3.1.2 *Human-Computer Interface Design*

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

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

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

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



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

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

Figure 6: 3D Display and Interaction Examples

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

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

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

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

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

1 Image courtesy of Schlumberger and Norsk Hydro

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

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

3.2 AUDITORY PERCEPTION AND PSYCHOACOUSTICS

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

3.2.1 *The auditory System*

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

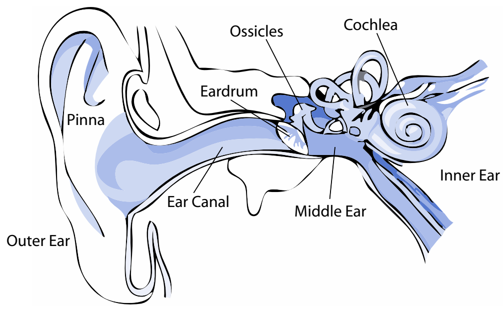
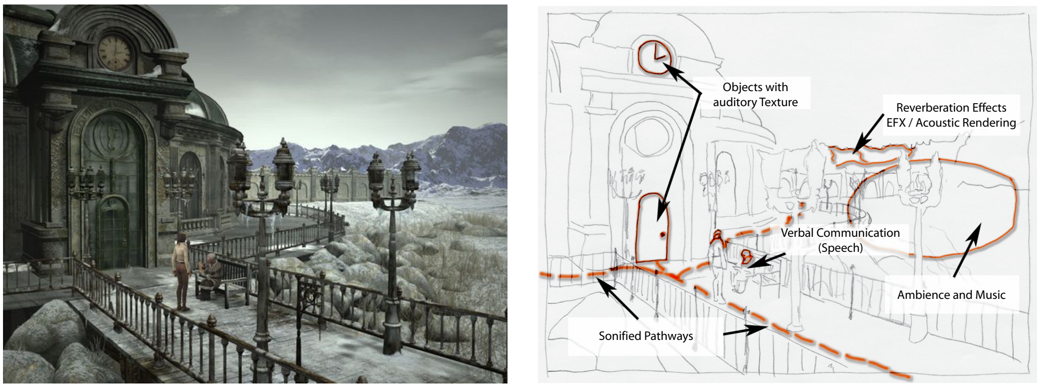


Figure 7: The Human Ear[2].

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

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



(a) Visual Scene Presentation. (b) Auditory Scene Presentation.

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

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

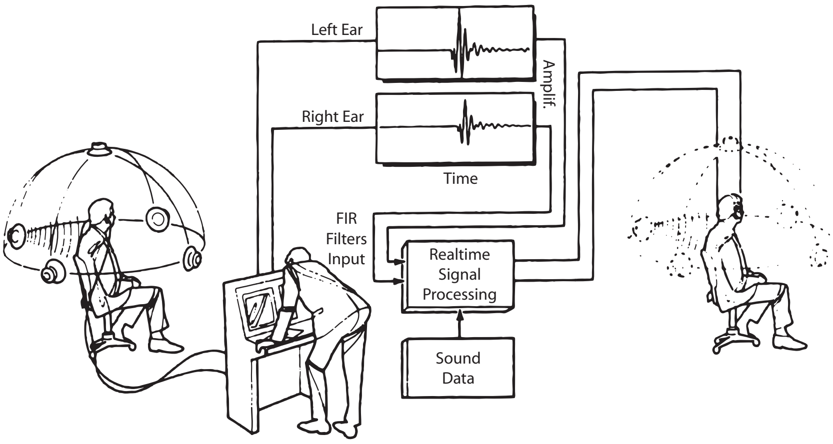
3.2.2 *Perception and Psychoacoustics*

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

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

Psychoacoustics

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



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

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

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

Scale Illusion (Headphones).

Phantom Words (Stereo Speakers).

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

Binaural Beats Demo (Headphones).

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

3D Sound Localization and Environmental Perception

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

3D Sound Demonstration.

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

3D Acoustics Demonstration.

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

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

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

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

3.2.3 *Speech Perception*

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

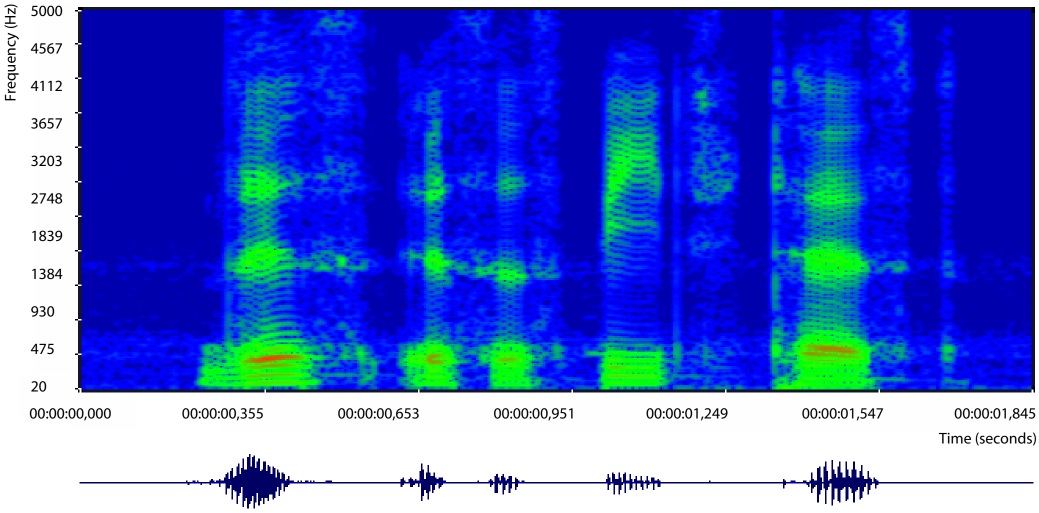


Figure 10: Speech Spectrogram *"This is a Speech Test!"*.

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

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

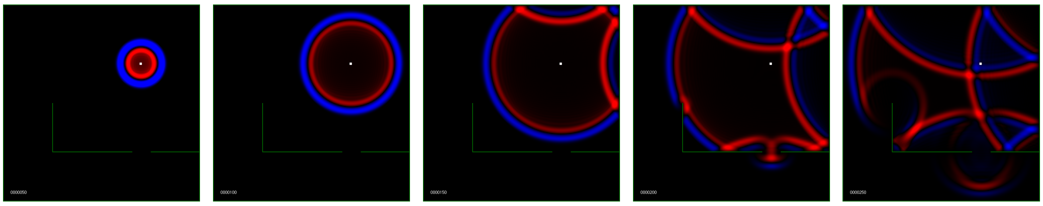
Speech Synthesis Examples.

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

3.3 sound rendering and physical acoustics

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

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



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

Figure 11: Propagation of Sound Waves.

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

3.3.1 *Physical Acoustics*

Sound Wave Propagation.

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

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

3.3.2 *Sound Signal Processing*

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

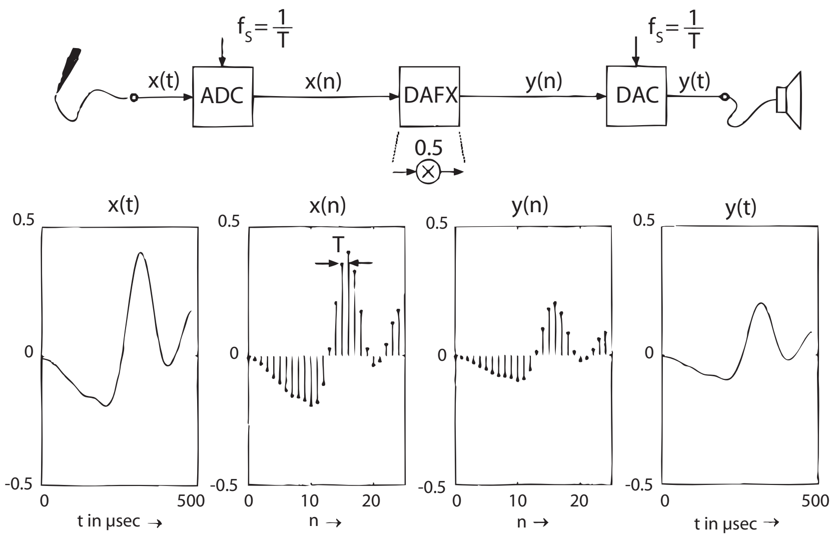


Figure 12: Sound Signal Processing (Zölzer, 2002).

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

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

3.3.3 *Sound Rendering and Audio Hardware*

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

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

4 DSP – Digital Signal Processor

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

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

Implementation

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

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

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

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

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

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

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

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

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

Listing 3.1: Basic OpenAL Example.

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

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

3.4 AUDIO IN ENTERTAINMENT COMPUTING

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

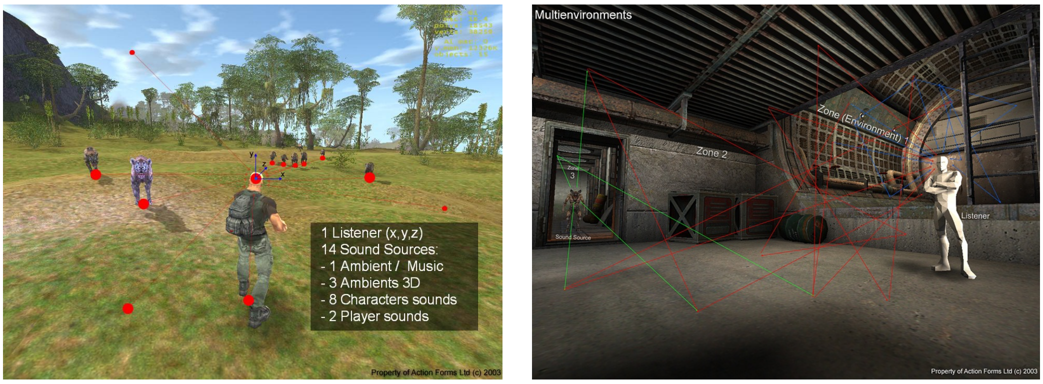
> *"...sound is 50% of a movie experience." (THX Consortium, 2000)*

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

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

Silent Hill Origins Trailer.

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



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

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

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

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

3.4.1 *Music-centered Computer Games*

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

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

(a) REZ (United Game Artists, 2001).

(b) Guitar Hero (Harmonix Music Systems, 2005).

(c) Metronome (Tarsier Studios, 2008).

Figure 14: Music-centered Computer Games

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

REZ Demo Video.

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

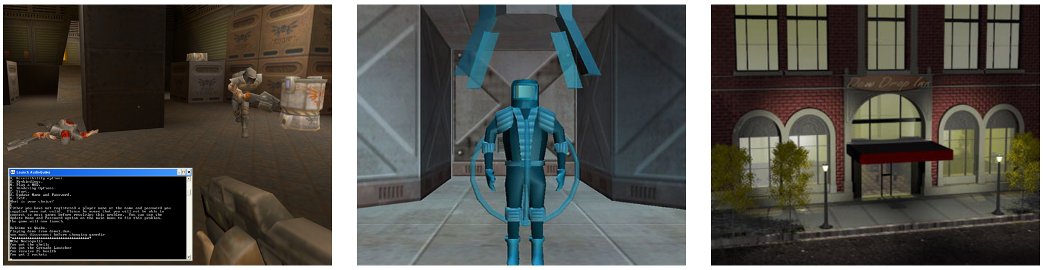
The City of Metronome.

3.4.2 *Audio-only Computer Games*

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

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

Terraformers Demo.



(a) AudioQuake (Atkinson and Gucukoglu, 2008). (b) Terraformer (Pin Interactive, 2003). (c) Seuss Crane (Destiny Media, 1999).

Figure 15: Audio- and auditory Adventure Games

Shades of Doom Demo.

Seuss Crane Detective for Hire.

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

3.4.3 *Audiobooks and Radio Plays*

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

Hans Christian Andersen's "The Nightingale".

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

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

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