Authoring of 3D virtual auditory Environments

Niklas Röber, Eva C. Deutschmann and Maic Masuch

Games Research Group
Department of Simulation and Graphics,
School of Computing Science,
Otto-von-Guericke University Magdeburg, Germany
niklas/masuch@isg.cs.uni-magdeburg.de

Abstract. Auditory authoring is an essential component in the design of virtual environments and describes the process of assigning sounds and voices to objects within a virtual 3D scene. In a broader sense, auditory authoring also includes the definition of dependencies between objects and different object states, as well as time- and user-dependent interactions in dynamic environments. Our system unifies these attributes within so called auditory textures and allows an intuitive design of 3D auditory scenes for varying applications. Furthermore, it takes care of the different perception through auditory channels and provides interactive and easy to use sonification and interaction techniques.

In this paper we present the necessary concepts as well as a system for the authoring of 3D virtual auditory environments as they are used in computer games, augmented audio reality and audio-based training simulations for the visually impaired. As applications we especially focus on augmented audio reality and the applications associated with it. In the paper we provide details about the definition of 3D auditory environments along techniques for their authoring and design, as well an overview of the system itself with a discussion of several examples.

1 Introduction

Many of today’s computer games feature an impressive and an almost photo-realistic depiction of the virtual scenes. Although the importance of sound has moved into the focus of game developers and players, it still does not receive the same level of attention as high-end computer graphics. The reasons for this are manifold, but some of them are already decreasing in a way that sound plays a larger role in certain games and game genre. One niche in which sound is the major carrier for information are so called audio- or audio-only computer games. These type of games are often developed by and for the visually impaired community and are played and perceived through auditory channels alone. Many genre have been adopted, including adventures, action and role-playing games as well as simulations and racing games. To bridge the barrier between visual and auditory game play, some of these games are developed as hybrids, and can be played by sight and ear [4]. For a more detailed discussion on these games we refer to [8], [9] and the audiogames website [11]. Both, audio-visual and audio-only computer games, need to be authored and designed regarding the story and the game play. For this purpose, specially designed authoring environments are often shipped together with the game-engines used. An overview and comparison of some commercially and free available audio authoring environments have been discussed by Röber et.al. [6].

In our research we especially focus on audio-only computer games and augmented audio reality applications in the context of serious gaming. Here we concentrate on techniques for sonification, interaction and storytelling, but also on authoring and audio rendering itself. The methods developed here are not only applicable for entertainment and edutainment, but can also be used in the design of general auditory displays and for training simulations to aid the visually impaired. With the authoring environment and the techniques presented in this paper, we focus on defining a theoretical foundation for 3D virtual auditory environments and the methods necessary to describe and design them. As many auditory environments are currently still programmed using software API’s, this authoring system opens artists and non-programmers a door to design and create (augmented) auditory applications in a very easy and intuitive way. The challenges in the design of auditory environments, which especially applies to the authoring itself, is to provide enough information to the user without overloading the auditory display and to keep the right balance between aesthetics and functionality.

Our authoring system is part of a larger audio framework that can be used for the design of general auditory displays, audio-only computer games and augmented audio reality applications. The system is based on 3D polygonal scenes that form the virtual environment. This description is used for collision detection and to assign sound sources to objects and locations. During the authoring, additional acoustic information is added to the scene. Therefore, for each object an auditory texture is defined and set up to specify the objects later auditory appearance. This includes different sounds and sound parameters per object, as well as story and interaction specific dependencies. The auditory presentation of such sound-objects can be changed by user interaction, time, other objects or an underlying story event. The authoring system is additionally divided into an authoring and a stand-alone player component. This allows an hardware independent authoring and the player to be used independently from the main system in mobile devices.

The paper is organized as follows: After this introduction, we focus in Section 2 on the definition of 3D virtual auditory environments and discuss here especially the concept of auditory textures along the varying possibilities for sonification and interaction. In this section we also motivate and explain the additional changes necessary to support dynamic environments and augmented auditory applications. Section 3 is build upon the previous sections.
and discusses in detail the authoring system using several examples. Here we explain the techniques and concepts used, and provide together with Section 4 additional information regarding the user interface and the soft- and hardware implementation. Section 5 presents and discusses the results achieved using some examples, while Section 6 summarizes the paper and states possibilities for future improvements.

2 Virtual auditory Environments

Vision and hearing are by far the most strongest senses and provide us with all information necessary to orientate ourselves within a real-world environment. Although one perceives the majority of information visually through the eyes, different and often invisible information is sensed acoustically. Examples can be easily found in daily life, such as telephone rings or warning beacons. Though the visual and the auditory environment, which are perceived by the ears respective the eyes, are partially overlapping, the larger portion is dissimilar and complements each other to provide a comprehensive overview of the local surroundings. Virtual environments are computer created worlds, which often resemble a real environment. Depending on the realism of the computer generated graphics and sound, the user might immerse into this virtual reality and accepts it as real. Virtual environments have many applications, ranging from simulations and data visualization to computer games and virtual training scenarios. The most successful implementation are computer games, in which players immerse themselves into a different reality as virtual heroes.

3D virtual auditory environments represent a special form of virtual environments that uses only the auditory channel to convey data and information. As discussed in the last paragraph, the auditory and the visual channel sense different information and form a diverse representation of the users surroundings. This has to be incorporated into the design of virtual auditory environments, if the goal is to visualize a (virtual) real-world-ensemble virtual environment. An advantage of hearing opposed to vision is the possibility to hear within a field of 360 degree and to also perceive information from behind obstacles and occlusions. Difficulties sometimes apply with the amount of data perceivable and the resolution of the spatial localization for 3D sound sources. Furthermore, auditory information can only be perceived over time and only if a sound source is active. For a technical realization, virtual auditory environments are simpler and cheaper to build, as no screens or visual displays are needed. Auditory environments have many applications, including auditory displays and of course audio-only computer games and augmented audio reality.

In order to receive enough information for the users orientation, navigation and interaction, a 3D auditory environment must exhibit certain qualities. These qualities and functions can be described as:

- A 3D (polygon-based) virtual environment managed by a scenegraph system,
- A 3D audio-engine with a non-realistic acoustic design,
- Sonification and interaction techniques,
- Input and interaction devices, and
- User-tracking equipment.

This list extends a little further with the design of dynamic and augmented auditory environments, see also the following paragraphs. The basis of virtual auditory environments is built by a 3D scenegraph system that manages all the 3D polygonal meshes which describe the scenes. This scenegraph is also responsible for collision detection, level of detail and to handle possible time-, position-, object- or user-based dependencies. Every object within the auditory scene must be audible in some way, otherwise it is not detectable and not part of the environment. The objects can be grouped into non-interactable, passage ways and doors and interactable objects [7]. Combined with this scenegraph is a 3D audio engine that is capable of spatializing sound sources and simulating the scenes acoustics. Due to the differences in perception, the acoustic design must not resemble a real-world acoustic environment, instead certain effects, such as the Doppler, need to be exaggerated in order to be perceived. Also, additional information for beacons, earcons and auditory icons to describe non-acoustic objects and events need to be integrated in the auditory description of the scene. In order to interact with the environment and to derive useful information from the scene, the user needs to be able to input information. This is be handled through a variety of sonification and interaction techniques, which have already been discussed in the literature [8], [6]. Difficulties often occur with navigational tasks in which the user needs to navigate from one point to another more distant location within a large scene. Path guiding techniques, such as Soundpipes have here proven to be useful to not get lost [10]. Another technique, which has demonstrated to greatly enhance the perception by imitating natural hearing behaviors, is head-tracking that measures the orientation of the users head and directly applies this transformation to the virtual listener. This enables the user to immediately receive feedback from the system by just changing the heads orientation. Head-tracking can also be used for gesture detections, in which nodding and negation directly transfer to the system. Section 4 presents an actual implementation of such a 3D virtual auditory environment, while the next to paragraphs extend the system towards dynamic and augmented auditory applications.

2.1 Dynamic Environments

While the user can only explore static environments, more interesting, but also more difficult, is the creation of dynamic and through user interaction changing environments. Dynamic classifies here not only animations and loops, but a reaction of the environment to the users interaction. This can be expressed through time-, position- and object-dependencies, which are directly bound to certain objects in the scene.
object-dependencies change an object's state and are induced by other related objects. Figure 1 shows an action graph that visualizes these dependencies, while Figures 2(a) and 2(b) display the later authoring and design of these dependencies using auditory textures. A menu system and soundpipes, which are using mobile sound sources, can be designed as well by using additional object dependencies.

The (inter)action graph that is depicted in Figure 1 is composed by time and user interactions and bound to a specific object in the scene. The edges describe conditions, which, if satisfied, connect to the next possible actions, while the nodes are build by counters and user interactions. All object conditions that are not related through user interaction can be described using time. This allows also the execution of events directly following other events. With some additional control mechanisms, this description can also be used to model a story-engine that controls narrative content and additional control mechanisms. Figure 2(a) shows the authoring of dependencies and their arrangement within the auditory texture after type for faster access. Figure 2(b) displays the final action graph that is constructed from the previous auditory textures.

2.2 Augmented Audio Reality

The term Augmented Reality comprises technologies and techniques that extend and enhance a real-world environment with additional (artificial) information. It is, unlike virtual reality, not concerned with a complete replacement of the real environment, but focusses deliberately on the perception of both worlds that intermingle and blend over. Ideally, the user would perceive both environments as one, and artificial objects and sounds positioned within the real environment [5], [1]. Augmented reality has many applications, ranging from entertainment and visualization to edutainment and virtual archeology [2], [3].

Augmented Audio Reality describes the part of augmented reality that focusses exclusively on auditory perception. The aforementioned time-, object- and user-dependencies change the auditory environment [6]. These auditory textures have now been extended to also control the object states through time, position and user interaction dependencies, and additionally handle also the references to the various sound files along the parameters for their playback.

Figures 2(a) and 2(b) display the authoring and design of dependencies using the concept of auditory textures. Figure 2(a) shows here the different dependencies and their arrangement within the auditory texture after type for faster access. Figure 2(b) displays the final action graph that is constructed from the previous auditory textures.

3 Auditory Authoring

Authoring is the process of designing a document and filling it with content and information from possibly different sources and media. Auditory authoring refers to the design of virtual auditory environments and the specification of dependencies to model the applications behavior. The authoring for audio-only and augmented audio reality applications takes often place directly using programming languages. But this method is neither intuitive nor can the content later be changed easily or adjusted. Together with the development of applications, this was one of the main motivations for this research as the need for more professional authoring systems is growing. A previous publication was already concerned with the authoring of virtual auditory environments, on

![Figure 2: Authoring and Design of Dependencies.](image-url)
which the current work is based along with the development of an augmented audio reality system [6].

Figure 3 shows a screenshot of the authoring environment, that explains the menu and the authoring concept. The center window shows a visual representation of the scene, while the right hand side offers sliders and parameter entries to adjust and fine tune the sound sources as well as the auditory textures. Objects can be selected by either clicking on them or through the list on the left. Basic functionalities that have to be supported by any such authoring system are:

- Select, create and delete sound sources,
- Position and orientation of sound sources,
- Specification of playback parameters, such as attenuation, loudness, rolloff etc.,
- Setup of background and environmental sounds,
- Definition and set up of dependencies, and
- The design of an auditory menu system.

3.1 Sound and Environmental Authoring

The first step for the sound and environmental authoring is to load a VRML file that represents the scene geometry. This data can be modelled with any 3D program, such as Maya or 3D Studio MAX, from which the geometry can be exported as VRML. After this, objects are selected and auditory textures as well as sounds assigned and defined. Several parameters can be adjusted per sound and also vary over time, see Figures 3 and 6(b). The user interface was designed using Qt4 and allows to detach the parameter entry forms and float them over the application to customize the layout.

In Figure 4 one can see a screenshot of the authoring environment with a graphical representation of a virtual scene with one sound source along their parameter visualizations. The cone visualizes the direction of the sound source, while the two wirespheres represent the attenuation and the rolloff space. The sound parameters that are adjustable include position, loudness, direction, inner and outer opening, minimal and maximal loudness, rolloff and many other.

Figure 5 displays the authoring of a ring topology-based menu system using six spheres. The ring menu allows between two
to six objects, which are automatically arranged and evenly distributed around the listener. Every object within the menu can be assigned an auditory texture with all the possible modifications. This system can therefore be easily used to control and adjust parameters inside the virtual environment.

3.2 Dynamic Authoring

After the authoring of the basic parameters, the dynamic authoring starts with the definition of dependencies and auditory textures. For each dependency exists a different input form, that assists the user in the authoring of parameters for the animations.

Figure 6 displays two examples for dynamic authoring. Here Figure 6(a) shows the design and animation authoring of a circle-based soundpath. Other geometries, like polygon lines or splines, can be used as well and are employed later within the environment to assist the player with navigation and orientation. For the animation, an object (sphere) is selected and the time for the animation specified. The start of the animation can also be triggered through any event, like time or user interaction, and repeated as often as required. In Figure 6(b) one can see the visualization of a positional dependency. The two transparent boxes mark the entry, respective the exit event to play a sound file if the user approaches the center box object. The two boxes are due to the low resolution of the user positioning in order to avoid a parameter flipping, see also Section 2.2 for more details.

4 System Design

While the last two sections focused on the theoretical foundations of auditory environments and their authoring for audio-only and augmented audio reality applications, this section provides an overview of the systems design along with some implementation details and a discussion on hardware related issues.

4.1 Software

The authoring system is based on a previous audio framework that was developed and applied to design audio-only computer games and to evaluate sonification and interaction techniques [7], [8]. This framework was built using OpenAL for sound rendering and OpenSG to manage the 3D content of the scenes. The same framework and libraries were used as basis for this authoring system, and extended by Trolltech’s Qt4 as API for the user interface design. Figure 3 shows a screenshot of the final application.

The authoring system was designed to allow an easy authoring of (augmented) 3D virtual auditory environments without the need for special knowledge or programming experiences. Additionally, the system was designed as a universal modeler to serve multiple applications, ranging from entertainment and edutainment environments to training simulations for the visually impaired. The authoring and the presentation of the designed application takes place in two different components. The entire system is therefore divided into two parts: the authoring and a runtime module. The authoring system is used to design the virtual auditory environment, which can also be tested on the fly using a built-in player component. The authored application can then be saved and executed on a mobile platform using the runtime system as well. This division allows a hardware independent authoring, in which the additional tracking and input devices are simulated by the mouse and keyboard.

Figure 7 shows an overview of the system, with the authoring component on top and the player module at the bottom of the figure. The player component also uses the VRML model to visually inspect the scene and to verify the authoring. The evaluation of the scene events are carried out using the authored auditory textures and the information from the tracking and user interaction equipment. The final acoustic presentation using sound spatialization and room acoustics is rendered by OpenAL.

4.2 Hardware

As the main focus of the paper is on the authoring of virtual auditory environments, we will keep the discussion on hardware-related issues very brief. The hardware for our portable augmented audio system consists of a regular Laptop (Dell Inspiron8200), a digital compass that is used as head-tracking device, a gyro mouse for 360 degree interaction, bone-conducting headphones for the acoustic presentation and a W-Lan antenna along several portable W-Lan access points for the user positioning. Although the system is very low cost and cheap, it is still very reliable and achieves good results. The digital compass is a F350-COMPASS reference design from Silicon Laboratories that uses three separate axis of magneto-resistive sensing elements that are tilt compensated. The compass connects to the computer via USB and can be easily polled using a simple API.
The 3D model that was used in the original game has been used several times learning about the history of the city and the cathedral. A tourist visiting the city can unveil several mysteries, while at the same time several sagas and myths have been combined into one story. A serious game takes place in an ancient cathedral of Magdeburg, where in training their orientational skills. The augmented audio application is considered as a serious game that assists visually impaired people. Tests using both applications yielded good results, although some points need to be improved. So far we have tested both applications using sighted users only, but additional tests with visually impaired participants are scheduled for the next month. Although the system worked well, difficulties arose with the accuracy of the positioning due to the highly reflective stone walls that interfered with the W-Lan based user tracking.

5.1 Campus Training Simulation

The other example, which shall be discussed in a little more detail, is an augmented virtual training scenario for the visually impaired. Figure 3 in Section 3 displays an overview of the map used and also shows the authoring of the dependencies using auditory textures. In this simulation, buildings and important places are characterized through certain sounds that are specific to their function. In our campus simulation, this is for example the rattling of plates and cutlery in the cafeteria, the rustling of pages and books in the library and space-like sounds representing the department of computer science. Using this training simulation, the user becomes familiar with the arrangement of buildings and locations in an off-line simulation using the player component. The orientation and position of the user are hereby input using the mouse and keyboard. Later in the augmented version, the user walks through the real environment recognizing the various sounds. The user perceives the same sounds and information, except that the position and orientation are now measured by digital compass / gyro mouse and the W-Lan positioning engine.

The authoring for this training simulation was very straightforward and relatively easy. The 3D model could be designed very fast using 3D Studio MAX as the buildings did not need to be highly realistic. Describing sounds for each building were taken from a sound pool CD-ROM and also created by ourselves by simply recording the auditory atmosphere at these locations. In the final authoring using the system depicted in Figure 3, these sounds were assigned to each building along some object and position dependencies. Figure 8 displays a screenshot from the runtime component (left) and the W-Lan positioning engine (right). It shows the view from the department of computer science towards the library. In the right figure, the users position is marked by a bright red dot in the corner of the middle/right building.

5 Applications and Discussion

The focus of the last sections was to form a theoretical foundation for 3D virtual auditory environments with applications in audio-only computer games and augmented audio reality. The emphasis in this section lies in the analysis of the results and a discussion on the performance of the authoring environment, the system and the initial definition of auditory environments.

As one of the two foci was the design and evaluation of augmented audio reality applications, we have implemented and tested two different scenarios. One is an augmented adaptation of an earlier audio-only adventure game [8], while the other can be considered as a serious game that assists visually impaired people in training their orientational skills. The augmented audio adventure game takes place in an ancient cathedral of Magdeburg, where several sagas and myths have been combined into one story. A tourist visiting the city can unveil several mysteries, while at the same time learning about the history of the city and the cathedral. The 3D model that was used in the original game has been used again, but the story and the story points were slightly adjusted to match the new requirements of the augmented system, especially for the user positioning. The story, the events and the user interaction were encoded within the dependencies of auditory textures. Although the system worked well, difficulties arose with the accuracy of the positioning due to the highly reflective stone walls that interfered with the W-Lan based user tracking.

Tests using both applications yielded good results, although some points need to be improved. So far we have tested both applications using sighted users only, but additional tests with visually impaired participants are scheduled for the next month. Although the entire campus is equipped with many overlapping access points, the positioning algorithm performs better indoors, due to the shadowing effects of the rooms geometry and furniture. In wide open spaces, such as the campus scenario, the signal strength is homogenous over long distances. Advantageous in
outdoor applications is the existing ambient sound environment from the real-world, whereas indoors are more silent. Hence, the authoring for outdoor applications is easier as many sound sources are already present. Additionally, in outdoor augmented scenarios the distribution of event locations is scattered over a larger area, which at the same time allows a better positioning as overlapping effects are easy to avoid. One subject reported that the quality of the bone-phones was too poor and disturbed the perception, while all candidates stated that the system, see also Figure 9 is easy to wear and handle.

![Figure 9: The Augmented Audio System in Action.](image)

Important for all applications is a careful selection of sounds, as some of the sounds used in the campus training simulation were difficult to classify and sometimes even bothersome. Longer acoustic representations performed better than shorter ones.

The next steps to improve the system and the two applications are a refining of the positioning system for outdoor tracking. Here we need more sampling points and a better interpolation scheme for the radio maps. Additionally, some sounds and event distances need to be checked and probably adjusted as well. But the most important part of future work is a detailed user study with sighted and blind users that features also a comparison between users that need to be checked and probably adjusted as well. But the most important part of future work is a detailed user study with sighted and blind users that features also a comparison between users that performed an off-line training with user that did not.

### 6 Conclusions and Future Work

In this work we have discussed virtual auditory environments and their basic qualities that define them. We have motivated this through several applications like audio-only computer games and augmented audio reality, for which the definition of auditory environments was extended. Furthermore, we have presented a system for the multipurpose authoring of various auditory applications together with several user supporting techniques. Finally we have presented and discussed a hardware realization for an augmented audio reality system along two example implementations.

Future work includes, as already outlined in the last section, a detailed user study using sighted and blind participants, as well as a refinement of the positioning engine to improve the resolution and accuracy.

### Acknowledgment

The authors would like to thank Mathias Otto for his help in developing the augmented audio system and especially for his work on the W-Lan positioning engine. Furthermore, we would like to thank Stefanie Barthen for her help in designing the test scenarios, as well as the group of Professor Nett for helpful discussions on W-Lan positioning and for lending us portable W-Lan access points.

### References


