3D AUDIO INTERFACES FOR THE BLIND

Christopher Frauenberger

Institute of Electronic Music and Acoustics
University of Music and Dramatic Arts Graz
Inffeldgasse 10/3
A-8010 Graz, AUSTRIA
frauenberger@iem.at

Markus Noisternig

Institute of Electronic Music and Acoustics
University of Music and Dramatic Arts Graz
Inffeldgasse 10/3
A-8010 Graz, AUSTRIA
noisternig@iem.at

ABSTRACT

Our work is dealing with alternative interaction modes for visually impaired and blind people to use computers. The aim of the proposed approach is to exploit the human hearing capabilities to a better degree than this is done by customary screen-readers. A surrounding, three-dimensional audio interface is potentially increasing the information flow between a computer and the user. This paper presents a virtual audio reality (VAR) system which allows computer users to explore a virtual environment only by their sense of hearing. The used binaural audio rendering implements directional hearing and room acoustics via headphones to provide an authentic simulation of a real room. Users can freely move around using a joystick. The proposed application programming interface (API) is intended to ease the development of user applications for this VAR system. It provides an easy to use C++ interface to the audio rendering layer. The signal processing is performed by a digital signal processor (DSP). Besides the details of the technical realisation, this paper also investigates the user requirements for the target group.

1. INTRODUCTION

This paper presents the results of investigations on using Virtual Audio Reality (VAR) as an interface to computers for visually impaired and blind users. Because of the sequential nature of commonly used technologies, it was found to be essential to develop a new interaction mode to increase the information flow between user and computer. While tactile devices like braille lines can only provide a limited amount of information per time, audio has the capability to provide a lot more information at once if made surrounding and spatial. Furthermore, hearing is a sense which allows different levels of intenseness of perception, where the range reaches from background sound to speech. This allows us to adjust the priority of information to the desired attention of the user.

The system presented realises Virtual Audio Reality using binaural simulation techniques and room acoustics. This VAR is intended to provide sound sources on different places embedded in a virtual environment.Visually impaired people should be able to explore computers as they do explore rooms they do not know in their everyday life. VAR enables visually impaired people to get a first overview of an environment without the necessity to explore it in detail.

The subsequent sections state techniques to transfer visual information into the virtual environment. They investigate the user requirements as well as the system requirements and provide a short description of the proposed system.

2. AUDIO INTERFACE

Introducing spatial sound to audio interfaces changes the strategy of converting visual information to audio interfaces. The most obvious advantage over sequential techniques is the natural perception, humans are using their directional hearing for orientation at all times. Audio interfaces can be modelled in the space allowing multiple information sources.

2.1. Mapping

In the style of screen icons, ”Earcons” [1] can be defined which represent the information acoustically, including position properties. The container in which all these earcons are embedded is modelled as a room in the virtual audio space. The user himself resembles the pointer device which is capable of movements and actions. This allows an auditory representation of a so called window, icon, menu and pointer (*wimp* [2]) style of interface on which most of the commonly used operating systems are based. Visual to auditory mapping can be illustrated by the example of a Windows desktop containing the usual icons. The user as pointer device might virtually “walk” to a desired earcon and perform any action (e.g. execute, rename, view properties) on it.

A desktop application is only one of various possible applications where this mapping can be performed. Every menu, every list or even drawing applications can be mapped by following the same strategy:

1. Identifying the desired information content.
2. Assigning earcons to the information and defining the auditory representation considering the desired level of attention.
3. Grouping earcons to parent earcons to keep the amount of sources low.
4. Defining the desired actions which need to be performed on the earcons.
5. Defining a proper space in the virtual environment as the enclosing virtual room of the interface.
6. Placing the earcons in the VAR.

Spatial audio interfaces provide the possibility to cover most of the needed applications. However, the quality of the result depends on how well the mapping was performed. To improve the mapping, it is necessary to know as much as possible about the target group, what audio representation might achieve the desired goal and which technical possibilities are available.

2.2. User Requirement Specification

The target group is visually impaired or totally blind. This restricts interaction modes to audio and tactile devices. Many studies proved that they have better other senses to compensate for their lack of vision. Auditory compensation might consist of a reorganisation and reallocation at the level of the cortex (structural hypothesis) so that auditory and tactile areas function better or are the result of a better development due to the vision impairment (strategic hypothesis) [3]. However, the consideration of the differences in perception and imagination is inevitable for interface designers [4].

Most of the visually impaired were educated to navigate according to their listening skills in their everyday life. Mobility training is providing them with a highly sensitive warning, alerting and scanning system which allows visually impaired people to avoid the risks they are exposed while doing any physical activity. In general, the target group has a very distinct ability of orientation on the base of hearing.

Individuals who are congenitally blind rely on their remaining senses ever since they were born and therefore have the most distinct hearing and sense of touch among the visually impaired on average. On the other hand, congenitally blind have limited imagination of space and rely very much on their memory when moving around in a room. People who went blind later can build up a better imagination of reality. Although the memory skills of the visually impaired are generally above the average, these facts must be considered when modelling virtual rooms. Too complex rooms and combinations of rooms impose extra load on the user's memory and may confuse users.

Feedback devices like keyboard and mouse also need to be adapted for the target group. While a keyboard is suitable because it is basically tactile, a mouse is hardly usable for visually impaired user. The lack of a clear boundary and the impossibility to check the movements performed simultaneously on screen is confusing. In the system presented, the user himself is the pointer device. Intuitive movements can be achieved by using a joystick, but experiments with the prototype showed that especially congenitally blind people do have problems with a joystick. They had no clear imagination of the implementation of movements when using the joystick. A touch board device might satisfy the requirements better.

2.3. System Requirement Specification

The technical challenge of creating virtual audio reality demands special input-output devices and significant computational power. To be able to provide a system to the people concerned which is accessible not only in terms of usability but also in terms of costs, it was important to use only components which are customary and widely used. This section describes the chosen components and their properties with which the prototype was built.

To be able to compute the algorithms presented in section 3 significant computational power is needed. Current VAR systems either use DSP (Digital Signal Processor) cards to obtain the needed processing power or are entirely software based [5]. The system proposed employs a DSP because a high computational load at the PC due to the interface would restrict the applications the system is intended to present. The hosting PC controls the virtual scene and provides the DSP with the necessary data to compute the simulation.

The pointer device in this system was chosen to be a joystick for the reasons stated in the user requirement specifications. A joystick with three axis is employed to provide the user with the maximum of mobility. Besides the normal X and Y transversal movements, this joystick provides a Z rotation and thus support for bringing the source into the direction of maximum resolution via the joystick.

The localisation of sound sources can be improved significantly, if head movements are considered [6]. Head movements provide the user with two additional methods of improving the localisation: 1) Subliminal head movements solve ambiguous localisation problems, most often with decisions of whether the source is in front or behind. 2) Head rotation can be used to bring the source into a direction where the resolution of directional hearing is the best, to the front. For this reason the presented system is using a headtracker device which is mounted on the headphones.

Figure 1 shows a simple block diagram with all key components.

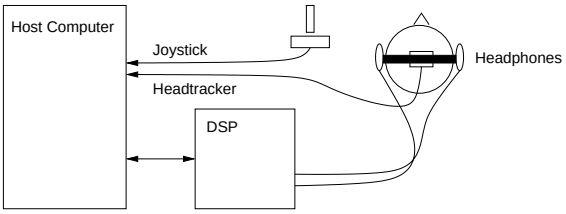

Figure 1: *Block diagram of the prototype*

The value of the prototype is about $6000.- including a non-commercial DSP evaluation board and a highly sophisticated ear-speaker set. If the minimal requirements of this system can be determined and applied to a product, the system would be available at lower costs. This would provide professional computer access to the visually impaired at costs comparable to those of visually unimpaired people.

3. VIRTUAL AUDIO REALITY

The quality of the proposed audio based human-computer interface is ultimately depending on the accuracy of the three-dimensional audio rendering. Key issues of creating virtual audio reality are 1) directional encoding, 2) reflections of the sound in the enclosing room and, 3) reverberation modelling. Algorithms are presented for each of these issues, including some implementation aspects.

3.1. Ambisonic

Ambisonic is a three-dimensional sound field reproduction system. It has the ability to localise sounds to a better degree than stereo or Dolby Surround[1] [7]. Although Ambisonic was originally designed for the use with loudspeaker arrays it was chosen as the directional encoding method in this project because if adapted to headphones it offers a number of advantages like economical resource consumption and effective rotation mechanisms.

The Ambisonic algorithm is based on the matching of interfering sound waves produced by a loudspeaker array and the original sound wave to be recreated at a certain listening point. A

[1]Dolby is a registered trademark of Dolby Laboratories, Dolby Surround is a trademark of Dolby Laboratories

virtual loudspeaker array in the horizontal plane consisting of n loudspeakers which are reasonably far away from the listening point to produce plane sound waves can reproduce a plane sound wave from any direction up to the m^{th} spherical harmonic where $n = (2m + 1)$. Applying HRTFs on these loudspeaker signals result in the desired binaural output. Figure 2 shows an Ambisonic array of third order used with the prototype.

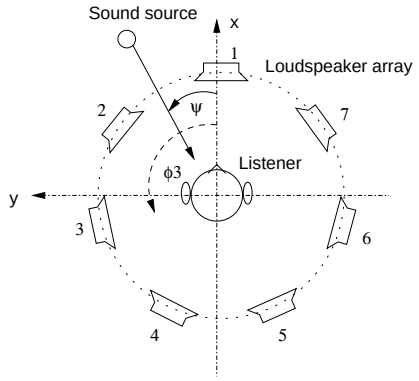

Figure 2: *Two-dimensional Ambisonic array of third order*

The advantages of Ambisonic over direct encoding with HRTFs can be summarised by:

- Accurate directional encoding without the need of interpolations between HRTFs. The loudspeaker array is predefined and located at fixed positions during simulation.
- Simply extendable for multiple sources. Encoding of sources and their reflections can be done economically by summing the Ambisonic signals before applying HRTFs. The most costly part in terms of computational power, the convolution with the HRTFs, is therefore independent of the amount of used sound sources.
- A static set of only a few HRTFs is needed which saves storage space and makes them exchangeable more easily.

More detailed information on binaural reproduction systems using Ambisonic can be obtained from [8, 9].

3.2. Early Reflections, Reverberation

It was determined that early reflections are significantly contributing to the sense of space and sound source localisation [10]. After approx. 50ms of time the density of reflections in a room reaches a value where the listener is only perceiving a decaying reverberation.

The proposed system uses a mirror-source algorithm of 2^{nd} order to produce the early reflections [11]. These mirror sources can be treated as normal sources but with additional attenuation because of the wall absorption. The location of the sound sources implies the time delay and attenuation of the incident sound wave according to the path difference.

The late reverberation does not contribute to the localisation of sound sources but is significant for the sense of space and qualifies the properties of the simulated environment. The system implements a feedback delay network with a unitary feedback matrix (Hadamard) producing a high density reverberation controlled by the loop gain and delay length [12].

4. APPLICATION PROGRAMMING INTERFACE

This section presents two APIs which provide user applications with intuitive functions to use the proposed audio rendering system. The AISound3D API was developed as the interface to the DSP program described in section 3, AIInput provides hardware independent functionality to user applications for input devices like the headtracker and the joystick. Figure 3 illustrates the co-

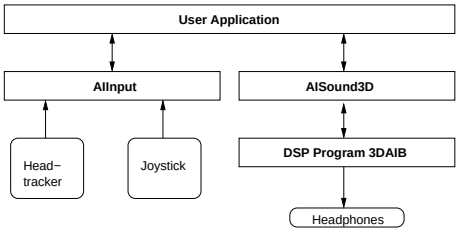

Figure 3: Software Architecture

operation of the APIs with their neighbouring layers. The user application which is modelling a virtual audio reality incorporates the APIs in order to realise the VAR. Setting up the scene is done with methods from the AISound3D API. It controls the DSP program and is responsible for providing it with the necessary audio data. During simulation the user application can obtain information about the current state of the feedback devices from the AIInput API and adjust the scene accordingly.

The design of AISound3D provides similar functionality as the most popular 3D-modelling APIs, extended with methods specific to this application. AISound3D was implemented in C++ and is extensively employing object orientated design patterns. Figure 4 shows the class diagram resulted from the design process which describes the most important components and their relations. In

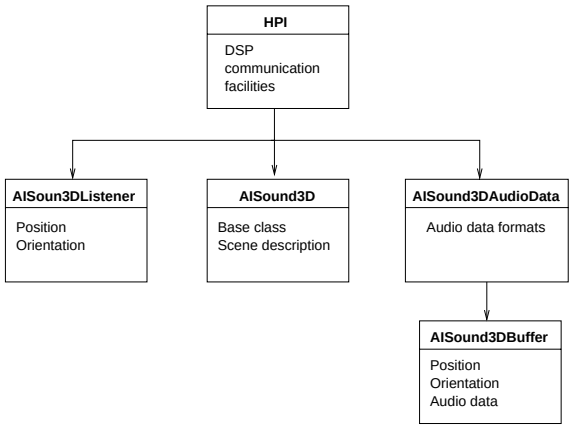

Figure 4: Class diagram

the centre of the design a main class called AISound3D manages the virtual scene. It defines the room in which the scene is taking place and handles references to the scene's content objects like listeners and sound sources (buffers).

The following example is intended to clarify the usage of the proposed APIs. In Listing 1, the core instances are generated and the properties of the room is defined.

Listing 1: Initialisation

```
aiSound = new AISound3D();
aiInput  = new AIInput();

AIS3DRoom *aiRoom;
aiRoom = aiSound->getRoom();
aiRoom->iDimension[0]=10;
aiRoom->iDimension[1]=10;
aiSound->setRoom(aiRoom);
```

After the initialisation process the system is ready for sound buffers to be created.

Listing 2: Buffer management

```
aiBuffer[i] = aiSound->createBuffer();
aiBufferProp = new AIS3DBuffer;
aiBufferProp->fPosition[0]=1;
aiBufferProp->fPosition[1]=2;
aiBuffer[i]->setSource("WindowsWord.wav");
aiBuffer[i]->setAllParameters(aiBufferProp);
```

Here, an earcon is created at a certain location of the room with the "WindowsWord.wav" file assigned as sound source.

5. CONCLUSION

The proposed audio rendering system is capable of producing virtual audio reality in order to model a user interface for people with visual impairment. The current state of development allows user applications to create VAR and provides techniques to manage its content. The first contacts of the target group with the system were encouraging and the approach to exploit the capabilities of the human auditory system to improve the access for visually impaired and blind people to computers is promising. For the time being there has been no experimental evaluation carried out which would proof the concept, but it will be the next step in the development progress.

The key to a high quality VAR is the audio rendering. The proposed spatialisation algorithms have still room for improvements. The Ambisonic algorithm uses very simple HRTFs at the moment. Research has proven that shorter HRTFs, modified by windowing techniques, improve the sound source localisation [8]. Additional distance cues may support users to build up the right mental model of the virtual scene presented. Just-noticeable difference (JND) levels can reduce the necessary number of computed early reflections when considering hearing thresholds.

With the increasing capabilities of customary PCs and sound cards it will also become possible to do all calculations at the host and no longer on a DSP. This would drastically decrease the costs of the system and makes it even more available for the people concerned. Real-time computer music software like Pd by Miller Puckette[2] already provides DSP like capabilities which could be used to compute the simulation.

To be able to evaluate the proposed system it is necessary to define clear attributes in terms of quality and usability. The definition of quality for virtual environments is crucial to be able to compare real world solutions with their virtual replacements. Especially in the field of audio, the simulation of the real world is always restricted by assumptions and approximations. These constraints result from limited processing power, the personalised auditory capabilities and cross-modal interaction. For a more detailed discussion about quality assessment and usability evaluation for this system refer to [13].

[2] `http://crca.ucsd.edu/~msp/`

6. REFERENCES

[1] R. M. Greenberg M. M. Blattner, D. A. Sumikawa, "Earcons and icons: Their structure and common design principles," *Human-Computer Interaction*, vol. 4, no. 1, pp. 11–44, 1989.

[2] A. D. N. Edwards, "The design of auditory interfaces for visually disabled users," in *Proc. ACM Conference of Computer Human Interactions*. May 15–19 1988, pp. 83–88, ACM Press, Washington, D. C., USA.

[3] P. Ghesquiere et.al., "The significance of auditory study to university students who are blind," *Journal of Visual Impairment & Blindness*, pp. 40–45, January 1999.

[4] D. M. Lane B. N. Walker, "Psychophysical scaling of sonification mappings: A comparison of visually impaired and sighted user," in *ICAD Proceedings*. ICAD: International Conference on Auditory Display, July–August 2001, Espoo, Finland.

[5] E. M. Wenzel J. D. Miller, "Recent developments in slab: A software-based system for interactive spatial sound synthesis," in *ICAD Proceedings*, Kyoto, Japan, July 2–5 2002, International Community for Auditory Display.

[6] P. Minnaar et.al, "The importance of head movements for binaural room synthesis," in *ICAD Proceedings*. ICAD: International Conference on Auditory Display, July–August 2001, Espoo, Finland.

[7] J. S. Bamford, "An analysis of ambisonic sound systems of first and second order," M.S. thesis, University of Waterloo, `http://audiolab.uwaterloo.ca/~jeffb/thesis/thesis.html`, 1995.

[8] Piotr Majdak M. Noisternig, "A head position related binaural sound reproduction system," M.S. thesis, Institute of Electronic Music and Acoustics, Graz University of Music and Arts, 2002.

[9] A. Sontacchi et.al, "An objective model of localisation in binaural sound reproduction systems," in *AES: Proceedings*. AES: Audio Engineering Society, June 2002, St.Petersburg, Russia.

[10] H. Järveläinen T. Lokki, "Subjective evaluation of auralization of physics-based room acoustics modelling," in *ICAD Proceedings*. ICAD: International Conference on Auditory Display, July–August 2001, Espoo, Finland.

[11] A. Helmuth L. Cremer, Mueller, *Die wissenschaftlichen Grundlagen der Raumakustik*, vol. 1, Hirzel Verlag Stuttgart, 2nd edition, 1978.

[12] J. M. Jot L. Dahl, "A reverberator based on absorbant all-pass filters," in *DAFx-00: Proceedings*. DAFx: COST G-6 Conference on Digital Audio Effects, December 2000, Verona, Italy.

[13] C. Frauenberger, "Three-dimensional audio interfaces for the blind," M.S. thesis, Graz University of Technology, Department of Communications and Wave Propageation, 2003, `http://iem.at/Members/frauenberger/Publications/thesis/html/`.