THREE DIMENSIONAL AUDITORY DISPLAY: ISSUES IN APPLICATIONS FOR VISUALLY IMPAIRED STUDENTS

Martyn Cooper

Institute of Educational Technology
Open University
Walton Hall
Milton Keynes
MK7 6AA UK
m.cooper@open.ac.uk

Helen Petrie

Centre for Human-Computer Interaction Design
City University
Northampton Square
London EC1V 0HB UK
h.l.petrie@city.ac.uk

ABSTRACT

This paper discusses issues arising from both practical investigations and conceptual work directed towards applications of three dimensional (3D) audio displays for blind students. These applications fall into two distinct categories:
1. The use of 3D audio to enhance interaction with the computer environment generally and
2. The creation of auditory virtual learning environments.

In the first category this paper discusses the issue of what should be the underlying metaphor for interaction, after stating the inappropriateness of the desk-top metaphor now ubiquitous in graphical computer interfaces. In the second area this paper outlines how various learning objectives may be achieved in 3D rendered audio and issues emerging from this are discussed with reference to an illustrative example of a sonic orrery.

1. INTRODUCTION

The authors have been leading a programme of work directed towards applications of 3 dimensional (3D) audio for blind students [1]. To date this work has consisted of technical and psychoacoustic investigations to verify the validity of the approach and the development of tools to enable simple audio environments to be created. This practical work has produced much discussion on issues arising from its envisaged application. This paper opens up these issues for wider discussion.

1.1. Use of Sound by Blind People

Contrary to popular belief blind people do not have an inherently more acute sense of hearing. However, because they are more dependent on this sense, many become well practiced in gathering information about the world around them from analysis of the sonic landscape. These skills can be potentially exploited in blind people's interaction with computer-based environments. One key feature of our sense of hearing that has to date been underexploited in this area is the spatial or 3D aspects of the human perception of sound.

1.2. Technologies for Reproducing Sound in 3D

Various technical approaches exist and are emerging which could be employed as methods to synthesize and replay 3D auditory displays. A brief overview of the possible approaches is given here.

During the 1970s much work was undertaken to try to extend the then established practice of stereo recording and reproduction to the reproduction of sound which would encircle the listener and even to give a true 3D *soundfield* about the listener. Ambisonics was one such development. Most other systems (e.g. Dolby MP encoding, with Dolby Surround or Pro Logic decoding, DTS, Lucasfilm's THX, etc.) have emerged for the creation of 3D sound effects in cinema or home theatre. They produce excellent effects in those contexts but are unable to locate a sound source from the side or the rear of the listener with any spatial stability. They are therefore not suitable for many applications of auditory displays.

Ambisonics is a method of capturing, encoding for recording and reproducing sound in 3 dimensions. Full descriptions of the technique with its psycho-acoustic and mathematical justifications can be found in papers by its founding father Michael Gerzon [2, 3]. Keating has presented a paper outlining the advantages of Ambisonics in audio virtual reality [4], which are equally applicable to auditory displays.

The unsuitability of cinema and home theatre 3D audio techniques to auditory displays has already been stated. The general conclusions of the authors from their work is that Ambisonics B-Format is a good option for the encoding of the 3D audio and that it offers a computationally efficient approach. However, alternative replay methods to the classical Ambisonic multi-speaker arrays must be considered. This should include Head Related Transfer Function (HRTF) approaches replayed through headphones or near field loudspeakers. Significant computational economies can be made if reproducing a virtual environment over headphones, using binaural signals, if Ambisonics is used as the means of encoding the environment [4]. A key reason for this computational economy is that rotations within, or of the soundfield, become simple cosine transforms of the encoded signals.

2. THREE DIMENSIONAL AUDIO IN HCI

One of the major envisaged application areas for 3D audio is in improving the interaction of blind users, in particular, with computer operating systems generally; i.e. within the domain of Human Computer Interaction (HCI). For example, this could be viewed as a sonic replacement for the Windows or Macintosh desktop.

2.1. The Importance of 3 Dimensions in Auditory Display

In complex auditory displays with potentially many sonic events occurring simultaneously, it becomes very difficult to isolate currently important information. If these events are rendered about the listener as if originating from different directions, this separation is made substantially easier. If one considers an auditory user interface based on "auditory icons' [5], or "earcons" [6], analogous to icons in a graphical user interface (GUI), then by separating the auditory icons spatially it would be easier to gain an overview of the interface and to home in on the desired application or tool. Taking this approach seems then to make feasible the creation of a *Sonic User Interface* to a computer operating system or application.

2.2. Auditory Metaphors

The now ubiquitous desktop metaphor in GUIs is fundamentally inappropriate when seeking to render the same information in sound. Objects on a desk in the real world rarely emit any sound at all. Therefore we propose that a radically different metaphor is required and suggest that this could be a *conversational metaphor*. In this metaphor the different objects a user might wish to interact with are considered as *conversation partners*.

Such conversation partners would each have a characteristic auditory signature. As a user navigates their way through a Sonic User Interface they would become increasingly aware of the auditory signatures around them, those closest to them being louder and acoustically rendered to be perceived as being closer. When they find the partner they want they would start a conversation with it. This might mean opening an application or executing some other command which could be styled as initiating a conversation. The user would be able to construct a mental representation of the world in such a way that they could readily locate the conversation partners they use. These would be organized in 3 dimensions rather than just the 2 dimensions of the GUI. Further, one could envisage a nested approach with worlds within worlds grouping together related conversational partners.

There is a key question that will only be answered when appropriate Sonic User Interfaces are constructed and user evaluations conducted. That is, are there significant benefits to blind users of this approach over and above the current approaches of making GUIs accessible, which are based on keyboard navigation and speech output? However there is reason to believe this may be so.

3. AUDITORY VIRTUAL ENVIRONMENTS

The other principal application of 3D audio the authors are working towards is audio virtual environments (AVEs) and in particular educational applications of such environments. This is best thought of as being akin to applications of virtual environments that are currently principally graphically based. Examples of these include interactive simulations of physical worlds to illustrate scientific principles or the use of virtual worlds in computer-based collaborative work.

3.1. An Example Educational AVE

The following example is offered as a potential AVE, targeted at blind students, to provide a context for discussion of the emerging issues in this area.

Figure 1. *A classical orrery.*

An orrery (named after Charles Boyle, the Earl of Orrery) is a mechanical model of the solar system (See Figure 1). These have been used to teach about the relative motions of the planets and their satellites since the 17th century. One can envisage an analogous computer based model of the solar system rendered as an AVE. The user perceives each heavenly body by its characteristic auditory signature. These appear to come to the user from the directions determined by the relative positions of the heavenly body within the computer model and the specified user position within this world. The user is then able to move through the solar system and explore it sonically. It is thus suggested that they will be able to build up a detailed mental model of the solar system. Beyond what is possible with a mechanical orrery it is suggested that the user could "visit" in turn the different objects within this virtual world and access detailed information about them.

3.2. Issues in User Perception in AVEs

There is a fundamental issue in determining whether AVEs have a realistic potential in education and that is: is human perception of worlds presented in such a way sufficiently true and consistent to achieve the learning objectives? Much of the authors' work to date has been directed towards determining this. The detailed results from this work are in preparation for publication elsewhere but some key points are summarized here.

The spatial acuity of individuals' perception of the position of sonic objects in virtual environments has been shown to be less precise than in the real world; however, it is probably not so much reduced as to invalidate the proposed approach. Further, although individuals may not be able to determine absolute position of sonic objects quite as accurately as in the real world, the relative position of sonic objects appears to be consistently interpreted. It is this relative position of sonic objects that is judged of prime importance for educational AVEs.

A key issue that will determine how widespread this approach might find application within education is the degree to which human perception enables mental models of complex spatial relationships of objects in AVEs to be constructed. In the orrery example above, would the user be able to build up a detailed picture of the relative positions and motions of the planets? Another envisaged application that might prove even more challenging would be an audio representation of molecular structures. Here an important leaning objective might be to understand particular alignments of atoms within a given molecule. Can these alignments be perceived from just a 3D sonic representation? This is not analogous to the ways which we normally use our sense of hearing. We can readily determine when we are exactly between two sounds but can we determine when three sounds are aligned when we move around them? Exploring these questions is an objective for empirical investigation in the next stage of the authors' work.

The orrery example described above raises another issue. It has been suggested that the user would be able to "visit" a planet and find out detailed information about it. For a blind user this information would be probably rendered in sound either as synthesized or recorded speech. Thus the user would in effect make a transition between the audio environment and what might be called an *information space*. Now what happens to the user's perception of where they are in the virtual world when they return from the information space? Will they retain the metal model they had previously built up by navigating through the AVE or will they have to reconstruct it by further exploration? What will be the effect of changes in the virtual world that have occurred while they have been in the information space?

4. CONCLUSIONS

Sound rendered in 3 dimensions has great potential for improving the ease of interaction with computer based environments, especially for blind users. There is still substantial work to be done both in developing appropriate technologies for realizing such environments and in exploring how they may be best used. This paper has raised some of the issues that impact on this ongoing development process. The key issues discussed and opened for wider discussion in this paper are:

- What metaphors are most appropriate when sonically rendering computer environments
- What are the benefits of rendering audio in 3D?
- What are the perceptual issues in determining spatial relationships and when moving between different sonic representations in virtual worlds?

The authors would welcome further discussion of these issues from interested parties and approaches suggesting possible collaboration towards realizing the potential of 3D auditory display for the benefit of blind students.

5. REFERENCES

[1] Cooper M., Pearson M.D., and Petrie H. *The Computer synthesis and reproduction of 3D sound environments - research towards an implementation for blind students*, Proc. Audio Engineering Society 16th International Conference on Spatial Sound Reproduction, Rovaniemi, Finland, April 1999

[2] Gerzon, M.A. *Practical periphony, the reproduction of full sphere sound.* Pre-print of the 65th Audio Engineering Society Convention, (1980), London.

[3] Gerzon, M.A. *General metatheory of auditory localisation,* Pre-print of the 92nd Audio Engineering Society Convention, (1992), Vienna.

[4] Keating, D.A. *The generation of virtual acoustic environments for blind people,* Proc. 1st European conference on Disability, Virtual Reality and Associated Technologies, (1996), 201-207, Maidenhead, UK

[5] Gaver, W. The SonicFinder: an interface that uses auditory icons. *Human Computer Interaction 4*, (1989), 67 – 94.

[6] Blattner, M., Sumikawa, D. and Greenberg, R. *Earcons and icons: their structure and common design principles*. Human-Computer Interaction 4 (1989), 4 – 11.