AUXie: Initial Evaluation of a Blind-Accessible Virtual Museum Tour

Aram Dulyan
Creativity and Cognition Studios
University of Technology, Sydney
PO Box 123, Broadway, NSW 2007, Australia
aram@it.uts.edu.au

Ernest Edmonds
Creativity and Cognition Studios
University of Technology, Sydney
PO Box 123, Broadway, NSW 2007, Australia
ernest@ernestedmonds.com

ABSTRACT
Remotely accessible audio-based virtual tours can offer great utility for blind or vision impaired persons, eliminating the difficulties posed by travel to unfamiliar locations, and allowing truly independent exploration. This paper draws upon sonification techniques used in previous implementations of audio-based 3D environments to develop a prototype of blind-accessible virtual tours specifically tailored to the needs of cultural sites. A navigable 3D world is presented using spatially positioned musical earcons, accompanied by synthesised speech descriptions and navigation aids. The worlds are read from X3D models enhanced with metadata to identify and describe the rooms and exhibits, thus enabling an audio modality for existing 3D worlds and simplifying the tour creation process. The prototype, named AUXie, was evaluated by 11 volunteers with total blindness to establish a proof of concept and identify the problematic aspects of the interface. The positive response obtained confirmed the validity of the approach and yielded valuable insight into how such tours can be further improved.

Author Keywords
Sonification, Audio-based 3D Environments, Inclusive Design, Accessibility, X3D

ACM Classification Keywords
H.5.2 User Interfaces: Auditory (non-speech) feedback, H.5.1 Multimedia Information Services: Artificial, augmented, and virtual realities, K.4.2 Social Issues: Assistive technologies for persons with disabilities.

INTRODUCTION
The acceptance of inclusive design principles has been rising steadily over the past few decades, with affordances to accommodate the needs of disabled people becoming increasingly ubiquitous. However, despite these advances, the importance of providing inclusive access to the world's cultural heritage, as offered by sites such as museums and galleries, has often been neglected (United Nations, 1994, s.10.2). In addition, the historic nature of some cultural sites can make providing inclusive physical access impractical. In such cases, local disability rights laws may advise that alternate access be provided using an auxiliary aid or service, such as a virtual tour, thus recognising such tours as viable alternatives to physical access (DDA, 1995). In addition, many cultural sites now offer virtual tours purely as a means of attracting new visitors.

However, the virtual tours currently being implemented are visual in nature, overlooking their potential utility to blind and vision impaired persons. Since such tours can be made remotely accessible, they can eliminate the need for the user to travel to an unfamiliar location, a task that blind persons find especially difficult (Heuten et al., 2006, p.155), as well as allow truly independent exploration, thus providing a more enjoyable and motivating experience (Lumbreras & Sánchez, 1999, p.323). The system presented in this paper, named AUXie, offers a means for blind users to take advantage of this utility offered by virtual tours by providing a remotely accessible audio-based representation of a 3D environment tailored to the needs of cultural sites.

An additional consideration addressed by AUXie is the process of creating an audio-based virtual tour. Since cultural sites are often limited in funds and technological resources, care was taken to ensure that the tour creation process is as economical and effortless as possible, particularly by providing the option to easily convert an existing visual virtual tour into an audio-based one.

We begin by providing a background on the sound perception theories and spatial sound reproduction technologies that have afforded the means of presenting navigable 3D environments using sound alone. We then discuss the techniques developed in related previous research and their applicability in the context of cultural sites. The implementation and testing process of the AUXie system is then described, followed by a discussion of the findings obtained from the evaluation of our prototype by 11 volunteers with total blindness.

BACKGROUND
The most fundamental concept at the core of this research is *sonification*, the process of representing objects, events, and other information using non-speech audio (Kramer et al., 1998, p.3). One of the most integral aspects of the sonification process is the choice of *earcons*, the sounds used to represent objects, actions, or events (Blattner et al., 1989, p.13). These choices are generally characterised by the nature of the relationship between an object or event and the earcon it is associated with. These associations can be placed on a scale ranging from

OZCHI 2010, November 22-26, 2010, Brisbane, Australia.

OZCHI 2010 Proceedings ISBN: 978-1-4503-0502-0

representational to abstract, where representational earcons, also referred to as *auditory icons* (Gaver, 1986, p.171), have direct and intuitive associations (e.g. the sound of flowing water representing a fountain), while abstract earcons can be completely unrelated sounds, such as musical melodies (Blattner et al., 1989, p.22).

An obvious disadvantage of abstract earcons is the fact that they require the user to learn what they are associated with. However, they are also capable of representing a much wider range of concepts than auditory icons. This is especially pertinent in the context of cultural sites, which may feature a highly diverse set of exhibits that need to be sonified. Furthermore, abstract earcons, especially musical ones, have been consistently preferred by listeners over auditory icons in terms of both pleasantness and appropriateness (Sikora et al., 1995).

Though the earliest examples of sonification predate the 20th century, the first non-auxiliary use of auditory displays was described in the early 1960s (Speeth, 1961), with the first uses of auditory displays as accessibility tools for vision impaired users appearing in the early 1980s (Lunney & Morrison, 1981). It would be some time before sonification could be used to effectively convey 3D environments, due to the advances in digital reproduction of spatial sound that were required (Begault, 1994, p.171), as the first commercial sound cards capable of spatially positioning a reasonable number of sound sources were only introduced at the end of the 1990s.

With the technology in place, the first implementations of fully sound-based navigable 3D environments soon followed. Some of the most advanced implementations were audio-based video games, developed with the support and feedback of thriving Internet communities of blind gamers. The most notable of these are AudioQuake by the AGRIP project (Atkinson et al., 2006) and Terraformers by Pin Interactive (Westin, 2004). A number of implementations were also developed within academia, including the efforts of Röber and Masuch (2004), Jaime Sánchez and her colleagues (Lumbreras & Sánchez, 1999; Sánchez & Saenz, 2006), and the Tactile Interactive Multimedia (TiM) project (Friberg & Gärdenfors, 2004).

Both recreational and academic implementations invariably follow the first-person shooter (FPS) paradigm, where the user hears through the ears of a virtual avatar, commanding it to step in one of the four cardinal directions, and performing actions such as shooting or examining on objects directly in front of the avatar. The following aspects of presenting a sonified environment were also notably consistent:

Navigation: discrete turning angles, i.e. yawing the avatar's viewport by a preset number of degrees (at least 22.5°) with each issued "turn" command; and discrete movement increments, corresponding to avatar steps, are used. This improves the accuracy of sound localisation by ensuring that the angle changes are large enough to account for both localisation blur (between 1° and 5°) (Blauert, 1989, p.39), and the minimum audible movement angle (often greater than 3°) (Begault, 1994, p.40). It also reduces the ambiguity of changes in the avatar's viewpoint, allowing users to easily return the viewport to a previous state or turn around to face the opposite direction, if desired.

Orientation: the absence of prominent landmarks, which can lead to disorientation with sighted and blind users alike, is addressed by compass-like aids, which output either the spoken name of the direction the user is facing or a continuous sound from a pre-set compass direction.

Objects of interest: as expected, spatially positioned earcons are used. Due to the limited variety of types of objects encountered in the existing implementations, the use of auditory icons is prevalent.

Headphone use: like most sonification research involving sound localisation, headphones are the preferred means of sound output, as the use of speakers generally results in an unacceptable loss of control over the auditory soundscape (Begault, 1994, pp.176-177).

IMPLEMENTATION

Interface

The implementation of AUXie generally follows the guidelines set by the previous implementations as described above. One of the crucial differences, however, is the use of abstract musical earcons for the representation of exhibits, which are the primary objects of interest within a cultural site context. Since the associations between the melodies and the corresponding exhibits have to be learned by the user, whenever the user's avatar crossed the threshold into another room, a "description" of the area is triggered. This description uses synthesised speech to list the exhibits located within the area. As each exhibit is named, the musical piece associated with it is played from a location directly in front of the user. Once the name of the exhibit has been spoken, its earcon is gradually moved from this location to the position of the exhibit within the room. The movement of the sound helps reduce confusion over its direction (Begault 1994, p.39), as well as its distance relative to other exhibits (Blauert 1983, Begault 1994). The musical pieces used are all played using distinct instruments and have been selected from various classical compositions. The pieces are reused between separate areas, so a particular piece may be associated with different exhibits depending on the room the user is in.

After listing the exhibits within a room, the description proceeds to list the available exits from the room in a similar manner, with the sound of blowing wind used to represent those exits. The final position of the exit sound is placed some distance past the threshold of the exit into the area beyond. The sound is then occluded by the walls of the room, as proposed by Andresen (2002) and reiterated in Menshikov's (2003) recommendations for the effective use of sound occlusion.

At any point during navigation within AUXie, the user may request the room description to be played again. In addition, when the user's avatar is less than two steps away from an exhibit, the user may request a detailed description of the exhibit to be given, which is played back using synthesised speech.

Technology

As previously mentioned, one of the main goals of AUXie is to ensure that the virtual tour creation process requires as little effort as possible. This was the main factor behind the choice of X3D as the 3D modelling framework used in AUXie. X3D is an open standard developed by the World Wide Web Consortium (W3C), designed to succeed the ill-fated Virtual Reality Modelling Language (VRML), and remains backwards-compatible to a large extent with VRML. This means that X3D is able to take advantage of the many modelling and conversion tools that have been developed for VRML over the years, and thus places very little restriction on the choice of 3D modelling packages that can be used to create X3D worlds. Furthermore, X3D was designed with remote access in mind, and due to the fact that it is eXtensible Mark-up Language (XML) based, allows metadata such as ontologies and semantics of objects within a 3D world to be defined with ease.

The codebase of AUXie itself is an extension of Xj3D, the official and most complete open source implementation of the X3D standard. Since it is written in Java, it allows AUXie to be distributed as an applet, thus ensuring both cross-platform compatibility and ease of remote access. The supporting libraries used in AUXie include OpenAL for 3D audio, and the Java Speech API for text-to-speech synthesis, with FreeTTS as the default synthesiser. All of these packages are unhindered by restrictive licences, ensuring that AUXie can be distributed freely.

An existing 3D world stored in the X3D format can be easily modified to allow AUXie to provide an audio modality in addition to the existing visual one. This is done by adding MetadataSet and MetadataString tags (both part of the official X3D specification) to the X3D document, which can be accomplished by any sufficiently advanced X3D editing tool such as Vivaty Studio, X3D-Edit, or even a simple text editor. Some of the currently supported metadata includes the identification and naming of rooms and exhibits, as well as textual descriptions of the exhibits.

EVALUATION

For the evaluation process, a 3D world was created based on the layout of the "Middle East" section of the British Museum in London. The world was fully modelled and textured to mimic the real-life virtual tour creation process, as well as to ensure that AUXie is visually interesting to any sighted users that may choose to use it. Exhibit descriptions from the British Museum website were used as the descriptive text for the exhibits. The AUXie prototype was then exhibited in an area known as Beta_Space inside the Powerhouse Museum in Sydney, which is a "living laboratory" of the Creativity and Cognition Studios (CCS) research group, providing the means for researchers and artists to exhibit prototypes of their work to the public and conduct user evaluations. User interaction with AUXie during the exhibition was enabled through a keyboard mounted on top of a podium facing a large screen. Only a limited set of keys on the keyboard were exposed to the user, and a set of headphones was attached to the podium.

Via a liaison with Vision Australia, 11 volunteers with total blindness were invited to evaluate the AUXie prototype. The selection of volunteers was very diverse, with a high variance in computer experience, age, and choice of accessibility aids. This is in marked contrast with most previous research, where the subject sampling was usually much more homogenous, such as only young schoolchildren (Lumbreras & Sánchez, 1999) or only computer-savvy gamers (Westin, 2004).

Each volunteer was instructed to interact with the system for as long as they felt comfortable, up to a maximum of 45 minutes. All of their actions during this time were logged by the system. The interaction was followed by a semi-structured interview on their experience. The questions guiding the interview sought to establish the subject's previous experience with computer-based auditory interfaces, identify aspects of the AUXie interface that were helpful or problematic, and prompt for suggestions on improving the experience.

RESULTS

The subjects' response to the system was overwhelmingly positive, validating this project as a successful proof of concept. As expected, the subjects with more previous experience with auditory interfaces were the most comfortable with the system, exploring more areas, and reporting higher levels of satisfaction. A few of the more elderly subjects struggled with the concept of virtualisation, and required some time before grasping the notion of hearing through the avatar's ears and controlling the avatar's movement.

There were a number of specific aspects of AUXie that proved to be problematic. The most widely reported issue was the difficulty of exit localisation, which was raised by four of the test subjects. This was likely due to the nature of the whooshing wind noise used to represent exits, whose high bandwidth, low frequency, and noise-like quality made it difficult to localise.

Three of the subjects also found the room transitions disorienting, often stepping back and forth across the room threshold repeatedly. Again, the likely cause of this was the poor conveyance of exit locations by the wind sounds used.

Two subjects reported issues with front-back confusion, indicating that the default spatial sound transformation may benefit from being augmented by additional muffling of sounds emanating from behind the user.

It should also be noted that the majority of the subjects did not take advantage of the compass aid, though the few that used it systematically found it very helpful.

One of the more surprising observations was the tendency of users of the system to walk backwards for extended periods, a behaviour that was somewhat unexpected given its rare occurrence in real life. This suggests that users may not be sufficiently comfortable with turning without becoming disoriented.

In an attempt to gauge the efficacy of audio-based virtual tours as accessibility tours, one of the interview questions was "Do you think that if the virtual space represented a real-life location, you would find it easier to navigate after using the program". Six of the subjects answered with a resounding "yes", while a further two answered "maybe". A notably encouraging response was from a subject who added that they "would not feel the need to go" to the physical place after exploring it using AUXie.

FURTHER RESEARCH

The next stage of this research is to analyse the log of user actions recorded by AUXie during the evaluation. This will be done by applying the principles of game metrics theory to examine the paths taken by users through the virtual environment. We expect that this will reveal further insight into how users approach an audio-based tour and help identify further aspects of the system in need of improvement.

The area in most apparent need of improvement, as identified by the evaluation results, is the means of informing users about exit locations. This can be addressed by the use of earcons that are easier to localise. Subsequent iterations of AUXie will also take into account the suggestions offered by the test subjects when asked for ideas for improving the system. Some of these suggestions include the use of different genres of music for different rooms; identifying the type of object that the user has bumped into by varying the "bump" sound; a "carpeted" path with muffled footstep sounds to help guide users through the room; lowering the volume of examined exhibits; and a panic button to transport the user to the entrance they had used to enter the room.

The evidence of a significant learning curve for elderly and inexperienced subjects also indicates that a tutorial may be necessary to accommodate the diverse range of users that may choose to use an audio-based virtual tour. A key component of this tutorial should be ensuring that the user is fully comfortable with turning. Finding effective ways to present such a tutorial would certainly also be worthy of further investigation.

ACKNOWLEDGMENTS

Acknowledgments extend to all of our colleagues at Creativity and Cognition Studios, especially Professor Yusuf Pisan. Many thanks to Deborah Turnbull for her continuing support and patience; the Powerhouse Museum for their support of Beta_Space; Vision Australia for their enthusiastic and helpful volunteers; and the British Museum for the use of their exhibit descriptions and floor layout.

REFERENCES

Andresen, G. Playing by Ear: Creating Blind-Accessible Games, http://www.gamasutra.com/resource_guide/20020520/andersen_01.htm (2002).

Atkinson, M. T., Gucukoglu, S., Machin, C. H. C. and Lawrence, A. E. Making the Mainstream Accessible: Redefining the Game, In Proc. SIGGRAPH 2006, ACM Press (2006), 21–28.

Begault, D. R. 3-D sound for virtual reality and multimedia, AP Professional (1994).

Blattner, M. M., Sumikawa, D. A. and Greenberg, R. M. Earcons and Icons: Their Structure and Common Design Principles, Human-Computer Interaction 4, 1 (1994), 11–44.

Blauert, J. Spatial Hearing: The Psychophysics of Human Sound Localization, MIT Press (1983).

Disability Discrimination Act, United Kingdom (1995).

Friberg, J. and Gärdenfors, D. Audio games: new perspectives on game audio, In Proc. SIGCHI 2004, ACM Press (2004), 148–154.

Gaver, W. W. Auditory Icons: Using Sound in Computer Interfaces, Human-Computer Interaction 2, 2 (1986), 167–177.

Heuten, W., Wichmann, D. and Boll, S. Interactive 3D sonification for the exploration of city maps, In Proc. Nordic conference on HCI 2006, ACM Press (2006), 155–164.

Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J., Miner, N. and Neuhoff, J. Sonification Report: Status of the Field and Research Agenda, National Science Foundation (1998).

Lumbreras, M. and Sánchez, J. Interactive 3D sound hyperstories for blind children, In Proc. SIGCHI, ACM Press (1999), 318–325.

Lunney, D. and Morrison, R. C. High Technology Laboratory Aids for Visually Handicapped Chemistry Students, Journal of Chemical Education 58, 3 (1981), 228–31.

Menshikov, A. Modern Audio Technologies in Games, http://www.digit-life.com/articles2/sound-technology/index.html (2003).

Röber, N. and Masuch, M. Auditory game authoring: from virtual worlds to auditory environments, In Proc. CGAIDE 2004, University of Wolverhampton (2004), 114–121.

Sánchez, J. and Saenz, M. Three-Dimensional Virtual Environments for Blind Children, CyberPsychology & Behavior 9, 2 (2006), 200–206.

Sikora, C. A., Roberts, L. and Murray, L. T. Musical vs. real world feedback signals, Conference companion on Human Factors in Computing Systems, ACM Press (1995), 220–221.

Speeth, S. D. Seismometer Sounds, The Journal of the Acoustical Society of America 33, 7 (1961), 909–916.

United Nations, The standard rules on the equalization of opportunities for persons with disabilities, United Nations General Assembly (1994), s.10.2.

Westin, T. Game accessibility case study: Terraformers - a real-time 3D graphic game, In Proc. International Conference on Disability, Virtual Reality and Associated Technologies 2004, University of Reading (2004), 95–100.