An Outdoor Spatially-Aware Audio Playback Platform exemplified by a Virtual Zoo

Graham Healy
CLARITY: Centre for Sensor Web Technologies
Dublin City University, Glasnevin
Dublin, 9, Ireland
ghealy@computing.dcu.ie

Alan Smeaton
CLARITY: Centre for Sensor Web Technologies
Dublin City University, Glasnevin
Dublin, 9, Ireland
alan.smeaton@dcu.ie

ABSTRACT

Outlined in this short paper is a framework for the construction of outdoor location-and direction-aware audio applications along with an example application to showcase the strengths of the framework and to demonstrate how it works. Although there has been previous work in this area which has concentrated on the spatial presentation of sound through wireless headphones, typically such sounds are presented as though originating from specific, defined spatial locations within a 3D environment. Allowing a user to move freely within this space and adjusting the sound dynamically as we do here, further enhances the perceived reality of the virtual environment. Techniques to realise this are implemented by the real-time adjustment of the presented 2 channels of audio to the headphones, using readings of the user's head orientation and location which in turn are made possible by sensors mounted upon the headphones. Aside from proof of concept indoor applications, more user-responsive applications of spatial audio delivery have not been prototyped or explored. In this paper we present an audio-spatial presentation platform along with a primary demonstration application for an outdoor environment which we call a *virtual audio zoo*. This application explores our techniques to further improve the realism of the audio-spatial environments we can create, and to assess what types of future application are possible.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Artificial, augmented, and virtual realities

General Terms

Algorithms, Design, Human Factors

Keywords

Audio-spatial applications, spatial sound, HRTF, virtual zoo application

MM'09, October 19–24, 2009, Beijing, China.
Copyright 2009 ACM 978-1-60558-608-3/09/10 ...$10.00.

1. INTRODUCTION

In the presentation of 2-channel audio through stereo headphones, certain processing techniques are often used to give a user a sense or perception of the direction from where a sound source is originating. The approach of "panning" sounds between earphones is often used to create the perception of a sound source moving linearly, and there are more complex processing techniques such as head-related transfer function (HRTF) [3] synthesis. However, as the user who is wearing the headphones moves his/her head from side to side, or turns around or changes location, the spatial delivery of the audio through the stereo headphones will fail to compensate for the wearer's movements and the wearer will not have any perception of the *source* of the audio. Mainstream audio delivery paradigms do not take account a user's physical location, movement or orientation in the playback of audio through headphones. Thus, audio is usually presented as a medium with a static perception whereby in the real world, for us, it is intuitive for us to recognise audio as a 3D phenomenon. As we listen to naturally-occurring sounds in the real world our senses attune us to the source of the sound(s) and so we perceive natural sounds to originate from specific locations around us.

When a person moves their head or changes location, their internal model of perceived sound sources expects to perceive sounds as still coming from their last perceived position relative to themselves, unless we expect the sound source to be moving, such as a car on a busy street. However with current mainstream audio delivery technologies this is not achieved; any internal model of sound source location which we may create when we use headphones for example, is destroyed as soon as we move or turn our head, and this is essentially discarding our in-built ability for spatial awareness. To remove us from the limitations imposed by current audio presentation technologies, we have constructed a platform capable of real-time spatial augmentation of sound so as to allow a user to freely move about and turn around within an outdoor environment and to perceive sounds as originating from specific spatial locations.

Previous work in this area has been conducted by [4] and by [6], [7], [5], and these have all been primarily proof of concept applications for indoor environments with some initial demonstration applications. Early indoor applications of sound source replication techniques in a sensor-equipped environment have shown that the method of preserving sound source localisation in generated sound is both possible and viable for audio-spatial information presentation. However, following the earlier work, testing was needed within a larger spatial area than had been available in previous work in order to assess the possible types of application which could be developed without the inherent physical constraints of a cluttered indoor environment (chairs, walls, tables, etc.). In the work we report here we have extended the prior work to allow an application testbed to begin assessing the viability of applications which would benefit from the integration of an audio-spatial presentation modality, in our case in an outdoor environment. The chosen outdoor arena was a small segregated park area within our University which is is 33×33 meters in dimension. The area has a grass surface, is fenced in with an ornamental hedge and does not suffer from GPS shadow from nearby buildings.

The rest of the paper is organised as follows. In §2 we present an outline of our system in terms of its components and how they work together. §3 overviews the virtual zoo demonstration application and the organization of the underlying components of its implementation. §4 concludes the paper and outlines possible avenues for future work.

2. SYSTEM OVERVIEW

To spatially enhance audio that is presented to a listener, real-time feedback is required of the listener's head orientation and movement, as well as their physical location and movement within a given environment. This applies whether the environment is indoor or outdoor. Acquiring the location, head direction and movement of the wearer in our case, is accomplished by equipping the wireless headphones with head vector tracking sensors, namely a tilt-calibrated compass module and a GPS module. Readings from these sensors are transmitted wirelessly back to a base station in real-time. These sensor values provide the parameters for our 3D audio production framework to continuously calculate the 2 audio channel sounds needed to create and maintain the perception of the sounds played back through headphones as actually coming from specific spatial points within the environment.

A simple demonstrator application to describe the functionality of the core operation of our system is the guitar demonstration. In this demonstration, we choose a specific 3D position in an outdoor area with GPS tracking availability, from which a user should hear a guitarist continually playing music even as s/he moves about the area. For simplicity purposes, this point is at the center of a segregated square park on the University campus surrounded by low hedges and bushes. A user wears the sensor-equipped headphones and is invited to move about the park area. As the user walks around within the boundaries of the park they continually perceive (regardless of their head movements) that the audio sound (the guitarist) is originating from the same spatial point, namely the center of the park. The basic technique of being able to virtually "place" sound sources in specific physical points can be combined with various ontologies to realize a broader set of applications. An example of this would be the context and location-aware information needed for a visitor taking an outdoor audio tour of a university campus, whereby a narrative, presented spatially, can inform the user of the relevance of physical places (shops, buildings, monuments, etc).

Our current outdoor hardware platform is comprised of off-the-shelf sensors mounted upon a standard pair of Phillips wireless headphones with good wireless range of approximitely 100m (accomplished using a small reflector antenna). The HMC6343 compass module was chosen because its interface is able to provide tilt calibrated yaw, pitch, and roll values of the user's head orientation at 10hz. The LS20031 GPS 5hz GPS module was chosen due to its high positional and temporal accuracy. Both of these sensors are mounted directly on the headphones, and connected back to a small hand-held unit that can also clip onto a belt. This small hand-held device consists of a 9v battery, a Parallax Propeller micro controller to interface with the sensors, and a long range ZigBee-Xbee RF unit capable of transmitting these sensor values to a base station in real time. To improve the accuracy of the GPS localization, we use a form of differential GPS with GPS readings from one fixed point whose coordinates are known, used to adjust and correct, if necessary, the readings from the user on the move. In previous work done by ourselves in [4] with a prototype indoor system, tracking head movements at 10hz appeared to be a sufficient update rate. However, fast movements of the head did cause noticeable temporary artifacts in the presented audio as with our current outdoor prototype. However, with a much increased update rate achievable by the hardware being provided by our partners in Tyndall [2] this will no longer be a problem in future versions of the system.

In Figure 1 we can see the wireless headphones labeled as point A. Points B and C show the compass and GPS sensors respectively, mounted on the headphones. Point D shows the small transceiver carried on the user's belt.

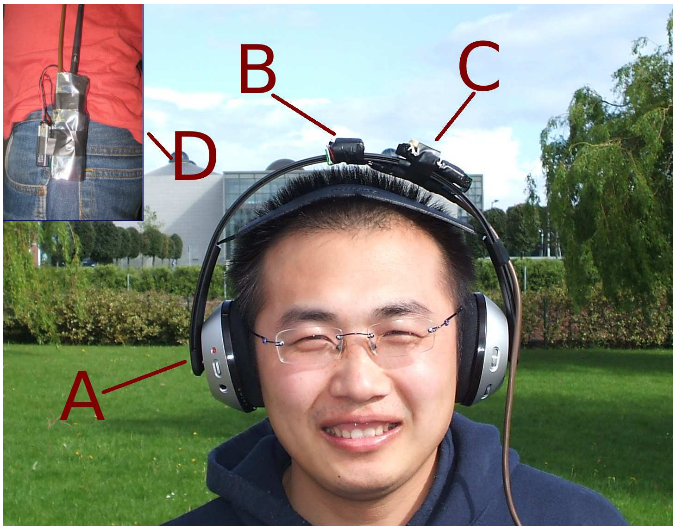

Figure 1: Augmented Headphones showing compass and GPS sensors

Sound generation and processing of the sensor values to adjust the sound to maintain the perception of sound source localization takes place on a computer using the Fmod 3D [1] audio production library in combination with an HRTF (head related transfer function) assisting sound card to more accurately model the psycho-acoustical cues which give us our perception of audio in 3D and because this is done in hardware it is comparatively fast to process. The technique of using HRTF extends the basic idea of "panning" by trying to reproduce sound as we actually hear it, taking into account the contribution of the pinna (outer part of the ear) and its effect upon our perception of 3D source locations. HRTF is concerned with modeling the distortion caused to sound waves by the shapes of our ears and head by the fact that sound waves have to wrap themselves around our head in order to reach both ears. The high level interface to the FMod library essentially comprises a set of functions to set the listener's position and orientation (based upon real-time sensor data), and similarly a set of functions to place sound sources (audio files) in spatial locations. FMod then manages the production and DSP operations preformed upon the audio by interfacing with this API. The result is 2-channel audio output on the sound card, which can then be transmitted wirelessly to the headphones.

3. THE VIRTUAL ZOO DEMONSTRATION APPLICATION

Early applications of the technique for generating virtual audioscapes in a sensor-equipped environment have been almost always indoor and have shown that the method is both possible and viable for audio-spatial information presentation. However, because of the restriction to indoor spaces using technologies like UbiSense for location tracking [8], development and testing of applications for larger areas has not been possible. Indoor location tracking has limitations not only because of space size but there are also few indoor locations which are completely free of physical artefacts like chairs, walls, tables, etc., and so the development of possible application types which would use larger spaces has not been done.

We chose to work on an outdoor area and because we use GPS for location-tracking we do not have space boundary limitations. Our applications are also far more portable as the actual location we use can be "moved" by using different GPS coordinates and the only infrastructure we need in place is an antenna for the streaming of audio to the wireless headphones and a desktop computer. The outdoor arena we used for the application described here was a segregated park area within the University which is 33 × 33 meters in dimension and it is pictured in Figure 2.

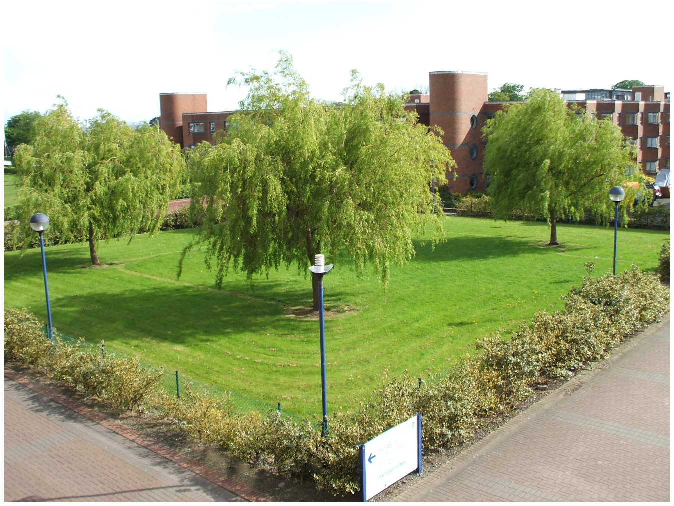

Figure 2: Outdoor audio-spatial area where the virtual zoo is based

An audio virtual zoo environment was chosen as the primary demonstration application. In this audio zoo a user can move freely around the outdoor environment, spatially perceiving a variety of animal and environmental sounds as though coming from particular locations around the park. As the user walks towards certain animal locations within the provided space in the environment, the sounds affiliated with these locations (virtual animal cages) will change accordingly to reflect this. For instance, if the user walks near the virtual monkey cage, s/he will hear an ensemble of monkey noises that reflect the monkeys' curious reaction to the wearer's presence near their cage. However, if the user is beyond a threshold distance from the virtual monkey cage s/he will only hear the odd noise reflecting normal sounds which monkeys would commonly make amongst themselves undisturbed. This is an example of the way in which the environment can be programmed to be dynamically responsive to wearer's movements. Our virtual audio zoo contains 7 animal types, and for each animal there are 8+ sound clips to reflect their various states of excitement or activity. Multiple audio clips may be associated with a given state. A variety of sounds for the same animal and the logic which dictates which sound should be played at a given time can be thought of as an entity tied to a spatial location. These entities respond to environmental factors including the wearer's location, the direction in which the user is facing, and the excitement state of other nearby entities. To clarify, the person will always spatially perceive any presented sound from an entity. However, if the user turns and faces that entity's general direction, similarly to the reactions we would get in a real world zoo with some really responsive animals, this should induce some state of change for that entity. For example, if you look at the virtual monkey cage, and are within a certain distance, the monkey entity at that point will change the sound it makes to attract your attention just as a real monkey might do. Each entity has a unique configuration defined by the sounds it uses and in which states it uses them. It responds to the user's location and to the various other entities around it. An entity can be thought of to be in a state, where transition between its states is determined by a rule set dependent upon environmental events.

For each state the entity enters there may be any number of audio sounds for that animal which may be played from that spatial location. Each audio clip for a particular state has a duration and a probability of being played. This method allows for the definition of entities which sound more natural than what would be achieved with the simple repetition of one particular sound representing the state of the animal at that point. As an example, a monkey in a state labeled "undisturbed" has 5 audio files, associated with that particular state, with respective probabilities for the likelihood of each specific audio clip being played. These are ("curious", 0.09, 3.4s), ("talking", 0.11, 2.2s), ("eating", 0.05, 1.2s), ("crying", 0.05,7s), ("bickering", 0.06, 6s), ("blank", 0.32, 2s) and ("blank", 0.32, 3s)[1]. Once the entity is in this state, it will randomly play one of the audio clips, weighted by the probabilities. Omission of an audio clip (no sound) is defined as blank. The last field of the tuple indicates the duration of that audio segment.

Entities define events of which they should be informed from the environment, so that when an event is detected they can change their state. Events may be generated by the user entering particular regions that are a given (physical and virtual) distance from the entity, facing towards the

[1] There are two "blanks" or silences of different durations

entity, or by a change of state by another nearby entity.

States defined for the animals would include normal (the default), looking for attention, excited, angry, fearful and others. An animal can transition from any state to any other state unless there is minimum duration defined between certain state transitions. A state transition for the pig entity from any state to the "scared" state is defined explicitly as the lion entity being in a state of "angry" for example in our outdoor audio virtual zoo. This event would cause the entity to arrange a sequence of audio clips associated with that sate, and spatially play those sounds as coming from that entitie's location.

4. FUTURE AND ONGOING WORK

Our future and ongoing work is in two directions. At the time of writing, we will shortly receive delivery of 10 additional audio-spatial headphones manufactured by the Tyndall Institute in Ireland. A re-design has packaged the sensor components into a more compact and discreet framework and will open the possibility of interaction not just between the wearer and virtual, sound-generating entities with given locations (animal cages in the case of the virtual zoo) but will allow interaction between users as well. In the zoo interaction this could take the form of animals reacting to different users in different ways, such as getting more excited on seeing the zookeeper (feeding time) than seeing a regular visitor.It is with the integration of this much improved hardware that a realistic subjective evaluation may be carried out on the aforementioned technique and its application to an outdoor space.

The second direction we are pursuing is to incorporate other on-body sensors such as a modular Wireless Inertial Measurement Unit (WIMU) [2] in order to parse simple biomechanical movements from the user, such as raising an arm to a lions' cage, or making a kicking gesture at a nearby cage. The availability of the data from these WIMU devices makes it possible to detect the wearer pointing at real, or virtual artifacts and hear a spatially directed narrative in an outdoor museum/guide type application. Their integration and use as additional sensors will be first be explored in the virtual zoo environment so as to determine and clarify the required functionalities of the underlying API.

5. CONCLUSIONS AND SUMMARY

In this paper we outlined an operational system for the spatially-aware playback of sound through wireless headphones in an outdoor environment which uses real time differential GPS and an on-board tilt-compensated compass as the core localization method. Previous outdoor applications of the audio spatial technique have relied on the user to carry a laptop in a backpack (or similar) to allow for the real time processing of the audio. Our system since it works over a wireless connection allows the user to freely roam without the ergonomic constraints inherent in previous application and demonstrations of the technique to an outdoor space. This technique in itself also allowed for the hardware accelerated HRTF audio synthesis on a PCI based sound card on the computer connected to the wireless base station transceiver. Similarly, since we now have a much larger physical application space, namely an outdoor environment, than in previous work, we were able to implement a comprehensive interactive audio application, a virtual zoo, where the user walks around the space and can interact with various animals at different locations.

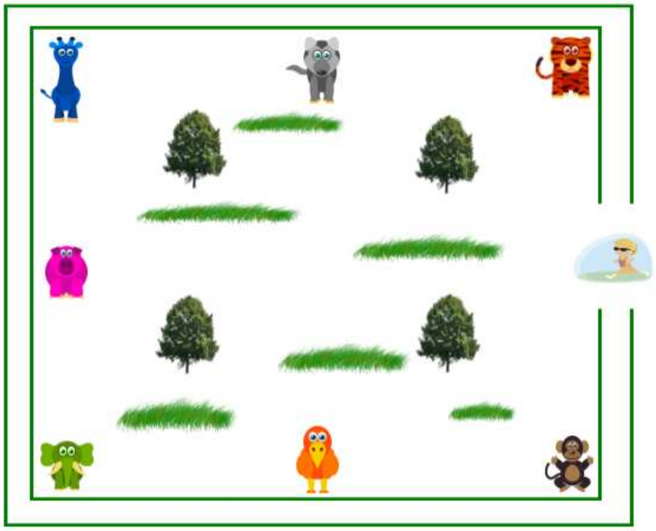

Figure 3: Graphical overlay of animal locations

Acknowledgment

This work is supported by Science Foundation Ireland under grant 07/CE/I1147, the CLARITY CSET. We thank colleagues from the Tyndall Institute in Cork, Ireland, for help in developing the hardware used in this paper.

6. REFERENCES

[1] FMod 3D audio production library. In *http://www.fmod.org*, 2008.

[2] The Tyndall Wireless Inertial Measurement Unit (WIMU). In *http://www.tyndall.ie/mai/wim.htm*, 2009.

[3] D. Begault. 3-D sound for Virtual Reality and Multimedia. In *NASA Ames Research Center, Moffett Field, Calif., USA*, April 2000.

[4] G. Healy and A. F. Smeaton. Spatially augmented audio delivery: applications of spatial sound awareness in sensor-equipped indoor environments. In *ISA 2009: First International Workshop on Indoor Spatial Awareness*, Taipei, Taiwan, May 2009. Institute of Electrical and Electronics Engineers.

[5] N. Röber. Interaction with sound: Explorations beyond the frontiers of 3d virtual auditory environments. In *PhD thesis at the Department of Simulation and Graphics, Otto-von-Guericke-UniversitLt Magdeburg, Germany*, 2008.

[6] N. Röber, E. C. Deutschmann, and M. Masuch. Authoring of 3D virtual auditory environments. In *Proceedings of Audio Mostly Conference*, 2006.

[7] S. Sandberg, C. Hakansson, N. Elmqvist, P. Tsigas, and F. Chen. Using 3D audio guidance to locate indoor static objects. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 50(4):1581–1584, 2006.

[8] P. Steggles and S. Gschwind. The Ubisense smart space platform. In *Adjunct Proceedings of the Third International Conference on Pervasive Computing*, 2005.