Spatially Augmented Audio Delivery: Applications of Spatial Sound Awareness in Sensor-Equipped Indoor Environments

Graham Healy and Alan F. Smeaton
CLARITY: Centre for Sensor Web Technologies
Dublin City University, Glasnevin, Dublin 9, Ireland.
{ghealy, asmeaton}@computing.dcu.ie

Abstract

Current mainstream audio playback paradigms do not take any account of a user's physical location or orientation in the delivery of audio through headphones or speakers. Thus audio is usually presented as a static perception whereby it is naturally a dynamic 3D phenomenon audio environment. It fails to take advantage of our innate psycho-acoustical perception that we have of sound source locations around us. Described in this paper is an operational platform which we have built to augment the sound from a generic set of wireless headphones. We do this in a way that overcomes the spatial awareness limitation of audio playback in indoor 3D environments which are both location-aware and sensor-equipped. This platform provides access to an audio-spatial presentation modality which by its nature lends itself to numerous cross-dissiplinary applications. In the paper we present the platform and two demonstration applications.

1. Introduction

True stereo sound played back on high fidelity stereo speakers or "surround sound" can be used to create an illusion of there being a specific location in space for the different sound sources being played, such as different instruments within an original music recording for example. This illusion of localisation of sound source gives a truer playback experience for the listener, though in practice many of us use either a pair of headphones or a pair of ear buds to listen to music and these can not accurately replicate the sound source localisation that good stereo playback can simulate.

In the current paradigm for playback of personal audio the sound coming from each speaker is the same and no account of the listener's head movements or location in space is used as part of the playback. "Panning" of sound between the headphone speakers is generally the approach taken to give the user a perception of at least some sound coming from a particular direction, but this perceived directionality is relative to the direction in which the user is already facing anyway and the perceived direction of the source will change as the user turns around, thus defeating the purpose of simulating sound source localisation. When we move around or turn our head we thus destroy whatever internal model we may have created in our minds of where the sound is coming from. Without panning of sound between speakers, if we move our head, then there will be no change in the audio and thus no change in our perception of the sound source's direction. This is essentially a static presentation to a listener of what is a 3D audio environment which fails to take advantage of the innate psycho-acoustical perception that we have of sound source locations around us.

To investigate and then develop applications which could be augmented naturally with some kind of spatial elements as part of the playback, we constructed a hardware platform with spatial augmentation of sound and we developed a set of applications to illustrate its potential.

Our prototype spatially augmented audio playback unit comprises off-the-shelf sensor components with which we have combined a 3d tracking technology to allow us to create immersive audio environments and to develop applications for these environments. Equipping a pair of standard wireless headphones with sensors to track the listener's head movements and the listener's location as s/he moves around, untethered, allows us to spatially enhance the audio played back in real-time so the user has the perception of sounds coming from specific and fixed points in their space, enhanced and re-enforced by their head movements and their actual physical location when moving around a room. A user can thus walk around our environment wearing these augmented headphones and perceiving sounds to come from specific fixed points within the room. Initial applications realised and demonstrated using this system range from blind navigation through to music presentation and delivery.

While there has been some previous work conducted in this area, this previous work used hardware which either failed to take into account the listener's actual location leading to positionally static applications, was ergonomically unsatisfactory as an application and/or failed to use user localisation technologies which had enough precision and accuracy so as to be useful for any large range of applications [4], [3]. These qualities each served to reduce the effectiveness of these early platforms in their own research environments which in turn made them unsuitable for adaption into cross-dissiplinary areas and applications. We believe our own work has gone beyond these limitations and that we have developed an ergonomically satisfactory platform which can be used to prototype and evaluate any number of integrated audio-spatial applications which rely on modular pervasive sensing technologies. We describe our system in this paper.

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 briefly presents a validation study we completed to confirm that our prototype was working effectively and in § 4 we present some initial applications of our platform which serve as illustrations of how it can be used rather than limitations on what is possible. § 5 presents our conclusions and future work.

2. System Overview

To spatially enhance audio as it is delivered to a listener, real-time feedback is required of the listener's head orientation and their current actual physical location within a given environment. For our work we assume the playback environment is indoor, though there is no reason that this would not work in an instrumented outdoor environment. The required location, direction and movement feedback is accomplished in our case by equipping wireless headphones with head vector tracking sensors, namely a digital compass, an ultrasound range detector and an accelerometer, which wirelessly transmit their readings back to a base station. This data, along with data on the listener's actual physical location in 3D, is input to a 3D audio production system which compensates for both the listener's head movements and their position in the environment so as to continuously create the perception of sound(s) played back through the headphones actually coming from given spatial point(s) within the environment. Our 3D audio production system continuously calculates parameters for this and then creates the actual 2 channels of sound which are then transmitted back to the wireless headphones and played back. The 3D positional tracking system we use is UbiSense 2.0 [5]. UbiSense equiped areas are fitted with several wideband sensors which each recieve the signals transmitted by small tags which are carried or worn by users. The UbiSense sensors then triangulate their readings from the tags and the system is thus capable of tracking small tags and providing their 3D coordinates within an instrumented area. The UbiSense system provides accuracy of approximately 15cm in $x - y - z$ dimensions.

A simple demonstrator application to functionally describe the core operation of this system is the guitar demo. For this demonstration we choose a specific point in three dimensions within an UbiSense equipped indoor environment and from this virtual location, a listener should hear a guitar continuously being played. For simplicity the sound source location is in one corner of the room. A listener wears the sensor equipped headunit and is then invited to move about the room. As the user moves about the room freely, the audio file (guitar) which is being played will continuously be modulated in order to make it appear to come from that specific corner point of the room regardless of the listener's head or physical movements.

The basic technique of being able to virtually "place" sound sources at particular given physical points can be combined with various ontologies to realise a broader application set. An example of this would be context and location aware information needs in a museum environment whereby the presented spatially directed narrative accomodates a user's need for contextual information which they may have missed by not taking a particular route through the museem.

Our current hardware platform is comprised of numerous sensors mounted upon a set of standard Sennheiser wireless earphones. Data values collected from the compass, acceleromator, and ultrasonic range distance module are transmitted through RF using the ZigBee transceiver mounted above the right earphone.

Sound generation and processing takes place on a computer using the FMod 3D [1] audio production library in combination with a HRTF (head related transfer function) assistive sound card to more accurately model the psychoacoustical cues which give us our perception of audio in 3D. This technique of using HRTF extends upon the basic idea of panning by trying to reproduce sound as we actually hear it taking into effect the contribution of the pinna (outer part of the ear) and its effect upon our perception of 3d sound source locations. HRTF is concerned with modelling the distortion caused to sound waves by the shapes of our ears and head and by the fact sound waves have to "wrap" themselves around our head in order to reach both ears. It leads to very realistic distortions of sound in terms of replicating what a human would hear. The high level interface to the FMod library essentially comprises of functions to set the listeners position and orientation (based upon realtime data), and similiarily a set of functions to place sound sources(audio files) at spatial locations. The production and DSP operations preformed upon the audio is then managed by FMod by interfacing with this API. The result is 2 channel audio output on the soundcard to be transmitted to the headphones. [2] provides a more complete explanation of HRTF in the context of this paper.

An illustration of the prototype we developed is shown in Figure 2. This diagram shows the circuit bread board (A) which houses the accelerometer, digital compass and pic micro controller, the ultrasonic sensor module (B), the XBee RF Module for wirelessly streaming sensor data back to the base station (C) and the Sennhesier wireless headphones (D).

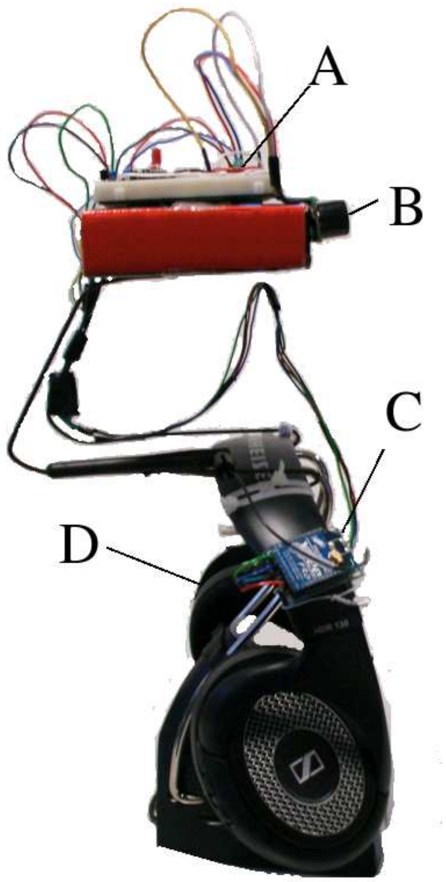

Figure 1. Prototype audio spatial augmentation headphones

3. Validation of Prototype

The basis upon which any prototype application for our spatially enhanced audio playback platform is reliant upon is the listener's ability to discern the directionality of each of the artificial spatially placed sounds. Outlined in this section is an experiment and a set of key metrics to validate that our platform achieves this.

3.1 Accuracy of determining static point sound directionality

In the first test we place a sound source at a random position about the listener and at one of three predefined distances. The listener was then asked to determine what direction the sound source was coming from, whilst standing at the same point on the floor, i.e. not being allowed to move around, and to do so as accurately and quickly as possible, with accuracy being more important. When the listener believed they were facing the sound source, they indicated so, and a reading of the observed direction and the time taken was recorded. The 3D tracking of the listener's actual position in the room was disabled for this experiment since we were only concerned with the listener's ability to locate the sound source direction. The sound origin points were randomly placed about the user so as to ensure an even distribution of their likelihood in a particular direction. The test subjects were given a 5 minute introduction and familiarisation with wearing the headunit which included hearing each of the sample sounds used in the experiment. Prior to the test, no subject had had any previous experience with the device.

Each user was required to locate the direction of 10 sounds with each placed at 1, 3 and 5 metres. The reasoning for choosing 3 different distances was an attempt to replicate the expected sound source distance range that would occur in real applications and to ensure that a user could still localise these sounds. The testing was done on 10 users, over a sample of 10 sounds, at 3 different distances. The users were undergraduate students from our University, all in good overall health, all aged in their early 20s and none had any specific hearing difficulties. The two primary metrics recorded were each users' time to locate the direction of the sound source, and the number of angular degrees of error between their perceived direction and the actual direction of the virtual sound source. This means we recorded these 30 times for each user.

Having tested a number of easily localisable sound in the development of the prototype we chose to use 10 sounds described in Table 1 which we found to be easily localisable.

Presented in Table 2 are the average error in angle and overall time taken to find sound source values for each subject:

These results indicate that the head unit device performs acceptably for tasks which require sound based navigation. The average of 11.5° with standard deviation of less than 7° comes from a fairly even performance across users except for user 1 who was off-target by more than 30° on average. Even this, though, is reasonably good as 30° corresponds to the angle of one hour on a 12-hour analog clock or watch face. Also, the time taken to perform sound source localisation is less than 10 seconds on average across almost all

Table 1. Sounds used in validation tests

Name	Description
Tap	Sound of a running tap
Guitar	Continuous Spanish acoustic
Dog	Dog barking
Cat	Cat meowing
Lamb	Lamb bleating
Blackbird	Blackbird whistling in an outdoor environment
Violin	Continuous source of violin music
Hooves	Continuous sound of horses hooves
Slurping	Sound of a man continuously slurping while eating
Street	Street noise, variety of sounds including people, cars, etc.

Table 2. Performance of sound source localisation tests

Subject	Final Angle (°) Error	Time to find sound source (secs)
1	31.15°	6.4s
2	6.9°	13.2s
3	7.3°	4.8s
4	9.1°	6.4s
5	8.9°	9.8s
6	11.3°	9.1s
7	8.2°	6.2s
8	15.6°	7.2s
9	9.7°	10.2s
10	7.2°	8.3s
Average	11.5°	8.18s
Std. Dev.	6.67°	2.3s

users.

It should be kept in mind that these values are for a user in a fixed point in space. One feedback comment from our users was that being able to move about gave them a higher degree of confidence as to which direction the sound was originating from, meaning that applications making use of the UbiSense component could yield even more accurate sound source direction. It should be also taken into consideration that in any real application the sound will generally be persistent in a particular direction, meaning that the initial or bootstrap time incurred by the user getting the general direction of the sound would be removed.

4. Applications of a Spatially Enhanced Audio Platform

Two proof of concept applications were created and tested using this headunit. These two very different applications were intended to show applicable diversity of this system to two differnet problems.

4.1 Blind Navigation

To show the applicability of this technique to blind navigation, we constructed a demonstration application to guide a user around a UbiSense equiped space, along a predetermined route. Guiding the user along this along this route involved drawing an overlay pathway on a map of the UbiSense area. This path was then divided into multiple segments where at the border of each segment a localised sound would be played to direct the user as to the next direction to follow.

Perceptually, the user had the impression of a sound originating from a particular point within the room, around which they could move freely whilst correctly perceiving the sound to continue to come from that point. When the user came within a pre-defined distance of that sound, or in other words walked into that point, the virtual location of the sound source would “move” to the next location as specified on the map to be navigated. This simple changing of localised sound sources proved to be an adequate system for guiding a blinded user following a pre-defined map route.

4.2 Virtual Band

For this experiment we acquired individual recordings for each instrument of a song, specifically the vocals, bass, guitar and drums. Having each instrument split into separate audio files allowed us to virtually “place” instruments at given spatial points accomplished using a GUI assisted application showing a map of the UbiSense equiped area. For the user this allowed them to walk among a virtual band perceiving instruments to come from particular spatial location thus enriching the presentation of the music and making the user seem to be part of the band, or at least be located among the band. Figure 2 illustrates the concept behind this.

Figure 2. Virtual band using spatially augmented headphones

5. Conclusion and Future Directions

Outlined in this paper is a platform we have built which is capable of augmenting mobile audio playback applications with an audio-spatial awareness. This can be used in combination with a 3D positional tracking technology such as UbiSense in order to create audio applications which use some physical space to allow users to move around and listen, wirelessly. Validation tests and some demonstrator applications, described in the paper, show it is a robust and working platform onto which we can now extend further applications.

At the time of writing there are four aspects to our future work with this platform. Firstly, we are developing other applications using the present platform. These include a museum guide application where the wearer would receive an audio commentary on what s/he was viewing in real time. This requires us to work with a 3D model of the instrumented space so we can account for occlusions and perspective from the viewer's line of sight for scenarios where the museum space is cluttered with multiple exhibits.

The challenge in this, apart from integration with the 3D model, is to author the audio material so it forms cohesive sense in terms of the listener's experience and isn't just a series of audio clips naming the object being viewed.

The current platform is a prototype and we are in the process of having the electronics miniaturized and encased in a smaller, neater housing as well as producing multiple platforms. This will allow us to develop applications with multiple simultaneous users wearing the headphones. One of the areas we have instrumented with UbiSense is an indoor area the size of a tennis court and we plan to develop game applications where position in the space relative to others, either playing alone or as a team, scores points . . . a kind of physical version of the board game ”Four-in-a-row”, sometimes known as ”Connect”.

Our third area for future work is to use the ultrasound sensor for more than just warning the wearer that they are close to a wall. This could lead to applications where we create 3D soundscapes in a physical environment.

Finally, we have started work on a version of our platform which works outdoors and where the location-awareness is provided by GPS. This will allow us to develop positional audio applications which can easily be set up to run in an outdoor space.

Acknowledgment

This work was supported by Science Foundation Ireland as part of the CLARITY CSET, under grant number 07/CE/I1147.

References

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

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

[3] N. Rober, E. C. Deutschmann, and M. Masuch. Authoring of 3D virtual auditory Environments. *Proceedings of Audio Mostly Conference, Pitea, Sweden*, 2006.

[4] 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*, 50:1581–1584(4), 2006.

[5] P. Steggles and S. Gschwind. The Ubisense Smart Space Platform. *Adjunct Proceedings of the Third International Conference on Pervasive Computing*, 191, 2005.