Auditory display design for exploration in mobile audio-augmented reality

Yolanda Vazquez-Alvarez, Ian Oakley and Stephen A. Brewster

Y. Vazquez-Alvarez and S.A. Brewster: Glasgow Interactive Systems Group, School of Computing Science, University of Glasgow G12 8QQ, UK

+44 (0) 141 330 8430

+44 (0) 141 330 4913

E-mail: {yolanda,stephen}@dcs.gla.ac.uk

www.gaime-project.org

I. Oakley: Madeira-ITI, University of Madeira, Funchal 9000-390, Portugal

+351 291 70 5117

E-mail: ian@uma.pt

www.m-iti.org

This work was partly presented at the Workshop on Multimodal Location Based Techniques for Extreme Navigation at Pervasive 2010, Helsinki, Finland.

Abstract. In this paper we compare four different auditory displays in a mobile audio-augmented reality environment (a sound garden). The auditory displays varied in the use of non-speech audio, Earcons, as auditory landmarks and 3D audio spatialization, and the goal was to test the user experience of discovery in a purely exploratory environment that included multiple simultaneous sound sources. We present quantitative and qualitative results from an initial user study conducted in the Municipal Gardens of Funchal, Madeira. Results show that spatial audio together with Earcons allowed users to explore multiple simultaneous sources and had the added benefit of increasing the level of immersion in the experience. In addition, spatial audio encouraged a more exploratory and playful response to the environment. An analysis of the participants' logged data suggested that the level of immersion can be related to increased instances of stopping and scanning the environment, which can be quantified in term of walking speed and head movement.

Keywords: *Sound garden, spatial audio, auditory displays, eyes-free interaction, mobile audio-augmented reality, exploratory environments*

1. Introduction

Recent advances in mobile technologies have made it possible to create location-aware audio-augmented spaces almost anywhere. A sound garden [1] is an example of such a space and it consists of a virtual audio environment superimposed on a *real* urban park featuring a set of precisely situated sounds

surrounding the user. In contrast to a simulated virtual reality environment in which participants are abstracted from the reality they are interacting with, in a mobile audio-augmented reality environment participants interact with the virtual audio mixed with *real* vision and motion.

A sound garden is usually intended for users to explore and experience casually rather than navigate via predefined paths. The unstructured nature of this activity presents unique challenges for the design of audio feedback to support exploration. Fundamentally, individual landmarks need to advertise themselves both to attract the user's attention and support subsequent targeting. This is typically achieved through a combination of user tracking technology (e.g. Global Positioning System (GPS)) and auditory beacons– sounds that activate when a user is within a specific distance from a landmark, typically within a *capture radius* [2]. Two concentric levels of audio feedback are often used, the first in a wide proximity zone and the second in a narrower activation zone [3]. The goal of audio cues in the proximity zone is to provide unobtrusive audio guidance, which enables a user to move towards the activation zone. Once this inner zone is successfully reached, additional content is made available to the user, either to indicate that a landmark has been found or to provide structured information describing it. Any error provided by the positioning system used will tend to require an increase in the size of these zones. Furthermore, the more unstructured and exploratory the environment, the more important the proximity zone becomes as a means of advertising landmarks. In a real environment, there is a likelihood that proximity zones may overlap if landmarks are located relatively close to each other.

One way to manage the presentation of overlapping audio landmarks is using spatial or 3D audio. Spatial audio refers to a set of techniques and algorithms, which allow audio delivered via a pair of speakers or headphones to appear to originate from different locations [4]. Stereo panning reliably positions sound to the left or right of a listener, while variations in intensity can indicate distance. However, binaural 3D audio algorithms allow more accurate localization of a sound source around the user, including the front and back. They work through the use of headphones and specific filters, or HRTF's (Head Related Transfer Functions) [5], through which monaural sounds are altered to appear to originate in particular spatial locations. Although much work has examined the use of

spatial audio for audio-augmented reality (see [6] for an overview), less work has compared different audio feedback strategies [7,8], and no work has investigated the use of 3D audio HRTF techniques in an exploratory mobile audio-augmented reality environment, especially dealing with the problem of overlapping proximity zones. Furthermore, a quantified approach to evaluating the implementation of a discovery environment is complicated by the open nature of the task. Unlike a conventional task-based experiment, we cannot equate speed of completion with success, and in many cases the idea of completion itself is inappropriate for what is in effect more an example of 'play' than 'work'. In this paper we analyse both movement sensor data and informal user feedback in order to describe user experience in an exploratory mobile audio-augmented reality environment. Ultimately, the research questions that this paper seeks to address are:

RQ1. What is the most appropriate auditory display configuration for an exploratory mobile audio-augmented reality environment?

> **a.** Non-speech audio is a commonly used feedback strategy in audio-augmented environments. However, to what extent do audio cues such as Earcons add to the experience in an exploratory mobile audio-augmented reality environment? To what extent do Earcons interact with other auditory display features?
>
> **b.** Given that the proximity zones surrounding the landmarks can overlap when these are located close to each other, should this overlapping be avoided or embraced?
>
> **c.** To what extent does spatial audio feedback, including distance and direction cues, affect the user experience when compared to limited spatial audio feedback including only the distance cue?

RQ2. Given the little amount of systematic assessment of user behaviour in this type of exploratory environment, what metrics and methods of analysis are best applied in a mobile audio-augmented reality environment?

2. Background

Early applications demonstrating the concept of mobile audio-augmented reality environments include Here&There [9]. The Hear&There system was able to determine the location and head position of the user using the information from GPS and a digital compass. This system used 'audio imprints' at the points of

interest. Audio imprints were "customizable collections of sounds that [could] be placed in the space" and consisted of "a single primary sound, with other audio braided in the periphery. These braids overlap the imprint, with each braid of audio shifting into and out of prominence". Users could listen to these imprints by walking into the area that the imprint occupied which was triggered by proximity. However, no further details on how these imprints were implemented or formal evaluation was provided for this work. More recently, Reid *et al*'s. Riot! 1831 [10] used similar techniques to recreate the Bristol riots of 1831 as a location-based audio drama in the streets of modern day Bristol. Users walked around one of the squares in the city equipped with a small backpack containing an iPAQ PDA, a GPS receiver and a pair of headphones; user position was used to trigger a variety of non-overlapping sound effects and script files based on real events that took place in the square. The Riot! 1831 system was found to provide a deep level of immersion within this exploratory experience.

Route finding applications such as Holland *et al.*'s AudioGPS system [11], Carter *et al.*'s Mediascapes [12], Audio Bubbles [13] and Soundcrumbs [14], have used abstract sounds as an auditory beacon to support navigation tasks and guide users to points of interest. These beacons alert users of their proximity to a location of interest through a brief repeating sound such as an Earcon [15] or an Auditory Icon [16]. An Earcon is a structured non-verbal audio message which uses an abstract mapping to provide information to the user (e.g. a trumpet sound to indicate the discovery of a location). On the other hand, an Auditory Icon is a familiar sound mapped onto an event to which it clearly relates (e.g. water noises to indicate the presence of a river). Auditory beacons are generally presented within proximity and activation zones around the landmarks. For instance, the proximity zone was 250 meters in the Audio Bubbles study, 55 meters in the Mediascapes, 20 meters in the Soundcrumbs study and not reported for the AudioGPS. The activation zone was 10 meters in the Audio Bubbles study, 5m in the Mediascape implementation, not used in the Soundcrumbs study and no information was provided for the AudioGPS application. Other applications like Stahl's [3] Roaring Navigator estimated the position and orientation of the listener's head by means of a GPS receiver and magnetometer, and also used stereo panning to indicate the direction of a navigational goal, i.e. animal sounds, located at the various enclosures in a zoo both in a navigational and an

exploratory scenario. This implementation is similar to AudioGPS and Mediascapes in that the landmarks were spatialised using stereo panning and more complex than the Audio Bubbles and Soundcrumbs implementations. Audio Bubbles did not spatialise the landmarks and only used distance mapped to the repetition rate and volume of a short 'click' sound to indicate that the user was near a point of interest (replicated the Geiger counter principle implemented in AudioGPS). Similarly, in the Soundcrumbs system the proximity of a "crumb" was mapped to a linear increase of the sound's volume. In addition, Stahl's system allowed for the simultaneous playback of five spatial sound sources but no detailed investigation was carried out into how this affected the user experience. Apart from the use of non-speech audio such as abstract or animal sounds for navigational tasks, other studies have explored the use of music in a similar manner. Examples include the Tactical Sound Garden (TSG) [1], Mobile Immersive Music [17], the Melodius Walkabout project [18], ONTRACK [19], and gpsTunes [20], a system in which users' own music from playlists was spatialised through the panning of the sound across the stereo sound stage as though it was coming from the specified destination or point of interest. Except for the TSG application and the Melodius Walkabout project, all the other systems logged heading data using magnetometer sensors supported on the mobile device. However, no heading data analysis was provided in the ONTRACK or the Mobile Immersive Music study, and in the gpsTunes system, heading data were used to identify at what point users were trying to locate the direction of targets by rotating around and pointing the device at each target. Other applications, such as those by Lyons *et al.*'s [21] and more recently Heller *et al.* [22] made use of ambient sound and narration to construct their sound environments. Interestingly, Heller *et al.* tracked head orientation in a non-realistic Wizard-of-Oz experience by mounting a compass sensor on the headphones worn by the user and, although no user experience evaluation was carried out, they observed that turning the head was the key to navigation by ear in this kind of mobile audio-augmented reality environment. The importance of head-turning data was also highlighted in Mariette's experimental work on outdoor navigation performance [23] in which he examined the impact of source capture circle radius and head-turn latency on performance measures of distance efficiency and head-turn latency rating. He concluded that the activation zone should be 3 meters or more for better user

navigation performance and also, as previously found by Brungart *et al.* [24], that the degradation of head-turn latency damages objective and subjective participant performance.

3. Sound garden implementation

Our study took place in a sound garden set in the Municipal Gardens in Funchal, Madeira. The sound garden ran on a Nokia N95 8GB mobile phone using software adapted from the Mobile Trail Explorer[1] application together with the HRTFs in the JAVA JSR-234 Advanced Multimedia Supplements API to position the audio sources. The location of the user was determined using an external Qstarz BT-Q1000X Travel Recorder GPS receiver[2] connected to the mobile phone via Bluetooth. The head orientation (compass heading) of the user was determined using a JAKE[3] sensor pack also connected via Bluetooth. No pre-determined route or visual aids such as maps were provided, but users held the N95 in their hands in order to press keys and make system input. They listened to the sounds planted in the garden using a pair of Beyerdynamic DT231 headphones. The GPS receiver was placed on the headphone's left ear-cup and the JAKE on the crown of the head, in the middle of the headphone's headband. Both sensors were mounted using Velcro tape. Figure 1 shows the final system setup.

[1] http://code.google.com/p/mobile-trail-explorer/

[2] http://www.qstarz.com/Products/GPS%20Products/BT-Q1000X-F.htm

[3] http://code.google.com/p/jake-drivers/

Figure 1 Experimental setup. 1) JAKE sensor, 2) GPS receiver (both mounted on headphones) and 3) mobile device.

3.1 User location tracking reliability

Location inaccuracy is always a concern in studies relying on GPS user tracking. Therefore, we ensured that, at all times the GPS data were as accurate as possible. In the *design phase* of the sound garden it was noted that the sensitivity and reliability of the in-built GPS receiver on the Nokia N95-8GB were not good enough for the requirements of this study, at least on the island of Madeira. Hence, the Qstarz BT-Q1000X external GPS receiver was tested and found more reliable and consistent for the purpose of our study. Also, *before* the start of each trial and as a training exercise for each participant, we checked GPS accuracy by asking users to find a virtual audio landmark situated outside the park. The application running the sound garden logged the GPS signal accuracy and printed it to the screen so the experimenter could confirm the GPS signal was good enough before asking the user to enter the park and start the experiment. *During* each trial, the experimenter closely shadowed the participant at all times. As all participants had been instructed beforehand to 'think aloud' while they walked through the park, the experimenter was able to detect whether the GPS had stopped tracking the user location.

The GPS resolution proved to be sufficient as participants were demonstrably able to find the virtual audio landmarks. However, if at any point the GPS stopped updating and it was not recoverable, the experimenter made a note of it, restarted the application, the participant was asked to go back to the last landmark they had successfully discovered and the data were discarded from the analysis. *After* the study was completed and while analysing the GPS data, all the trajectories recorded for each participant were plotted and confirmed GPS tracking reliability.

3.2 Audio content and system configuration

Five different Earcons in the form of recordings of animal sounds (an owl, goose, cricket, nightingale and frog) were created to alert the user of the presence of five physical landmarks: the Rua Sao Francisco; a Coat of arms of Saint Francis convent; the Statue of Joao Reis Gomes; the café and the pond. An illustrative map of the garden is shown in Figure 2. Animal sounds were used to identify landmarks because they seemed a good fit to the natural environment. Otherwise, the mapping between sounds and landmarks was abstract and symbolic; there was no pre-existing relationship between the sounds and the information they were representing. Furthermore, for each landmark brief speech audio clips were synthesized using Cereproc's (www.cereproc.com) British English male RP voice. These clips provided basic factual information about the sites. Synthesis made the setup of the sound garden easier by offering consistent and well-enunciated recorded speech without the need for a voice talent and a studio. Both the animal sounds and the audio clips were mono, 16-bit and sampled at 16 kHz. They were adjusted to a conversational volume (approx. 60-70dB).
Two circular zones surrounded each landmark: activation (radius 10m) and proximity (radius 25m) zones, in which different audio feedback could be enabled. Due to the size of the garden (82m x 109m), only three landmarks had overlapping proximity zones while the other two were isolated. Figure 2 shows the audio landmark configuration.

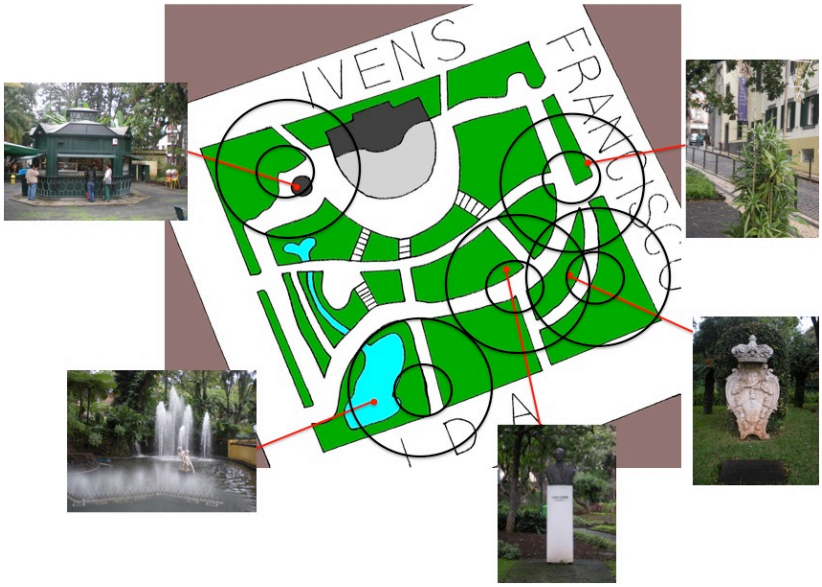

Figure 2 Municipal Gardens in Funchal, Madeira. Still images of the landmarks and illustration of proximity and activation zone per landmark.

4. Experimental design

4.1 Conditions

In order to answer our research questions, we need to evaluate the absence or presence of the following auditory display features: non-speech sounds (Earcons), proximity zone and spatial 3D audio. Some combinations of these features are inappropriate. Without a proximity zone we cannot use spatial 3D audio, as there would be no area for spatialization. Given we wish to investigate the overlapping of proximity zones and that previous work has shown that users can find concurrent speech streams frustrating and difficult to understand [25], Earcons are a requirement for these conditions. These restrictions result in four separate conditions, which vary in their complexity (see Table 1 for a summary):

1. *Baseline*. No Earcons or audio spatialization: When the user entered the activation zone, only the audio clip with information corresponding to that landmark was triggered and played once. The proximity zone was not used.
2. *Earcons*. Earcons but no audio spatialization: Whilst the user was within the activation zone, the Earcon (animal sound) corresponding to that location played continuously. The audio clip containing information about the location could be played (and the animal sound stopped) by pressing the central navigation button on the mobile phone. The proximity zone was not used.

3. *Spatial*. Basic proximity zone with Earcons and limited audio spatialization (distance): When the user entered the proximity zone, the Earcon, i.e. animal sound, corresponding to the location was triggered to alert the user of its presence (see Figure 3). The animal sound increased in loudness as the user walked closer to the physical landmark. The original sound level of the animal sound (60-70dB) dropped normally over distance (approx. 6dB per doubling of the distance to the sound source) making the quietest sound at the edge of the proximity zone 36dB. Once the user entered the activation zone, the audio clip could be played (and the animal sound stopped) by pressing the central navigation button on the mobile phone.
4. *Spatial3D*. Earcons and audio spatialization: Behavior similar to Condition 3, with the difference that the animal sounds in the proximity zone were played using full spatialization, varying not only in amplitude but also by direction of the sources.

Table 1. Summary of auditory display features per condition.

	Earcons	Proximity zone	Spatial 3D audio
Baseline	×	×	×
Earcons	√	×	×
Spatial	√	√	×
Spatial3D	√	√	√

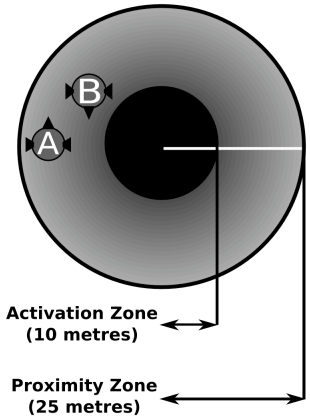

Figure 3 Audio landmark - gradient indicates volume. In the Spatial3D condition, User A (looking up in figure) hears a quiet sound to the right; User B (looking down) hears a louder sound front left.

4.2 Methodology and procedure

Very little previous work has carried out systematic and repeatable user experience evaluations in mobile audio-augmented reality. In addition, there is a lack of formal methodology on how to analyse and interpret user data that is not just qualitative, especially in an outdoor mobile audio-augmented exploratory environment. Thus, in this paper we set out to design and carry out an initial pilot study to explore these issues by focusing on user performance both quantitatively and qualitatively over a number of different audio displays. This resulted in a between-subjects design that allowed us to test our approach and offered rich and detailed results by participant but at the expense of controlling for cross-subject variation.

Eight users (6 male, 2 female, from 24 to 39 years in age) participated in the study. They were all students and members of staff at the University of Madeira and were familiar with the Municipal Gardens in Funchal. They all reported normal hearing and were right-handed. Five of these users had used GPS-based systems before. None were paid for their participation. Two different participants tested each of the four auditory display conditions described in the previous section. The experiment lasted no more than half an hour.

First, users were asked to familiarise themselves with the system by finding a landmark situated outside the park. This procedure served to check the system had GPS signal prior to starting the test and also provided participants with the chance to ask questions. They were then asked to enter the park and explore it freely whilst looking for the audio landmarks. They were all given a maximum of thirty minutes to walk around the garden. Half were directed to start at the part of the park with the isolated landmarks, while the others started where the landmarks were clustered together. Participants were instructed to verbalise their thinking process (a ‘think aloud’) while they walked through the park, and this information was noted down. As they encountered each audio landmark, the users were asked to listen to the corresponding audio clip before continuing their search. At the end of each trial for each different condition, participants filled in a questionnaire and provided informal feedback about their experience. In addition to participants’ comments and opinions, detailed logs (including distance covered, time spent, user location coordinates and head orientation) were collected on the mobile device to later perform an in-depth analysis of participant behaviour.

5. Results

We measured objective and subjective data on user performance. The objective measures we investigated were the time taken to complete the sound garden experience, the distance walked in meters, walking speed in meters per second, time spent stationary and head-turning data collected from participants exposed to spatial audio feedback. For subjective measures, we present feedback from the participant questionnaire.

5.1 Quantitative analysis (Time spent and distance covered)

The logged data showed that participants completed the experiment on average in 16 minutes and 15 seconds and the average distance covered by each subject was 692 meters (see Figure 4 and Figure 5 for more details per participant). The inclusion of spatialization in the audio feedback resulted in participants spending more time walking through the park and covering more distance.

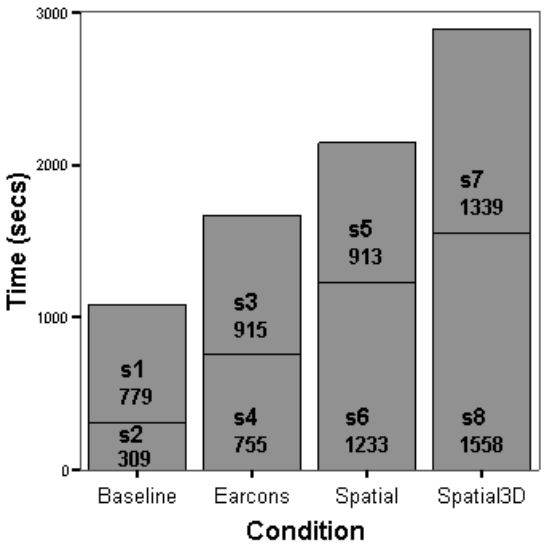

Figure 4 Time spent exploring for each participant (s1-8), stacked to show time spent per condition.

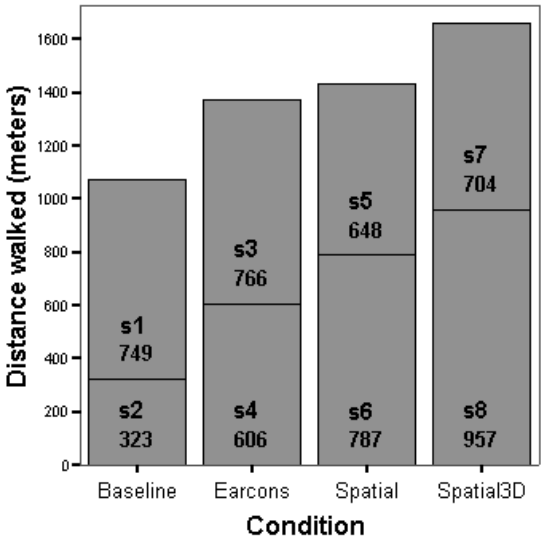

Figure 5 Distance walked for each participant (s1-8), stacked to show distance walked condition.

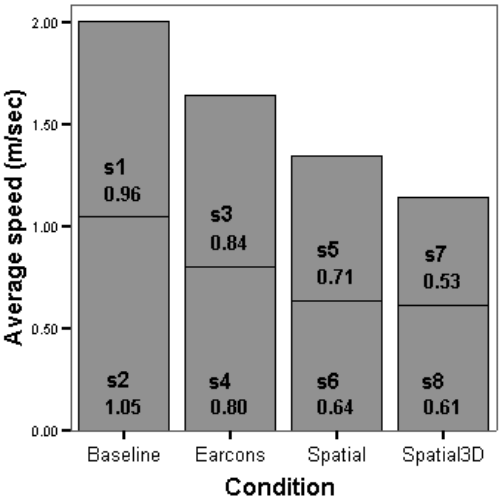

Figure 6 Average walking speed for each participant (s1-8), stacked to show walking speeds per condition.

In addition, participants' average speed dropped with increasing audio feedback complexity (Figure 6). The distribution of speed by non-spatial and spatial conditions (Figure 7) showed a significant main effect for condition type (t-test on log10 transform, to reduce skew, of speed values: $t(2874)=13.662$, $p < 0.001$).

Participants walked at a significantly lower speed during the spatial conditions (mean= 0.62 m/sec., SD= 0.51) than during the non-spatial conditions (mean= 0.90 m/sec., SD= 0.79). Looking more closely at the distributions, we can see that this drop in average speed was caused less by the participants walking more slowly but rather by an increase of the time they spent stationary (note the peak at 0 for spatial conditions compared to non-spatial conditions).

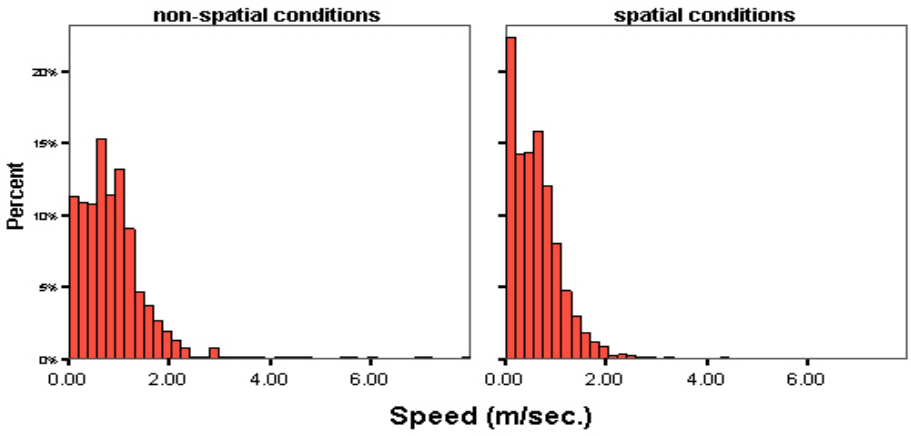

Figure 7 Histograms showing the distribution of walking speed by non-spatial (Baseline and Earcons) and spatial (Spatial and Spatial3D) conditions. Speed was calculated by dividing the distance walked by the time taken between each data point logged approximately every 2 seconds (mean = 2.28secs, SD = 0.29).

In this paper a threshold of less than 0.25 m/sec. (0.9 km/h) was used to identify stationary periods to allow for error in GPS readings. Error from the GPS readings means that subsequent positions are rarely identical even when the participant is completely stationary. Thus, in order to quantify stationary periods, the threshold was set based on the observation of the distributions in Figure 7. Histograms for both the spatial conditions show a bimodal log distribution. As we regard a participant to be either stationary or moving, we fitted these two distributions to these two states. Given an average human walking speed is 4.3 km/h, it is reasonable to regard 0.9 km/h as slow enough to be stationary. Using this threshold, Figure 8 shows the differences in the percentage of time participants were stationary. A Chi-square test showed that the percentage of time participants remained stationary significantly differed by condition ($\chi^2(3, N=3025) = 85.565$, $p < 0.001$). The effect of providing proximity information and full spatial audio feedback was that participants appeared to stop more often.

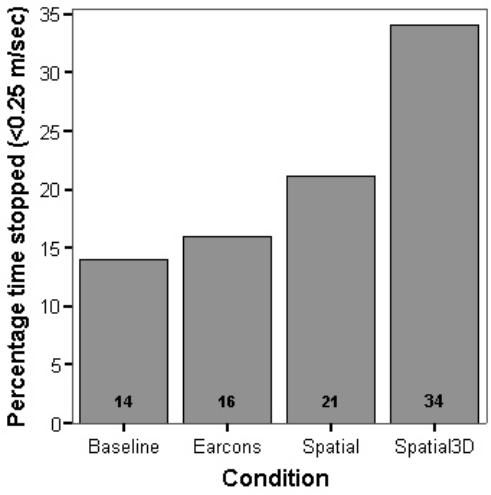

Figure 8 Percentage of time stopped for each condition. A threshold of less than 0.25m/sec was used to process user data identifying stationary periods.

The number of overlapping proximity zones for audio landmarks also had an effect on the percentage of time participants stayed stationary. Figure 9 shows percentage of time participants were stationary per number of nearby audio landmarks for the spatial conditions[4]. A Chi-square test showed that the percentage of time participants remained stationary significantly differed by number of overlapping proximity zones for audio landmarks ($\chi^2(7, N= 842) = 100.273, p < 0.001$). Participants exposed to full 3D audio feedback (Spatial3D condition) stopped more often as more proximity zones for the audio landmarks overlapped. In contrast, participants in the Spatial condition show a constant percentage of stopping as overlapping increased (see section 5.3 Figure 10a and 10b for an example of illustration of user behaviour).

[4] Only data from within the proximity zone were considered and data points while in the activation zone were excluded as we were only interested in user behaviour while exploring and not once they had reached the activation zone.

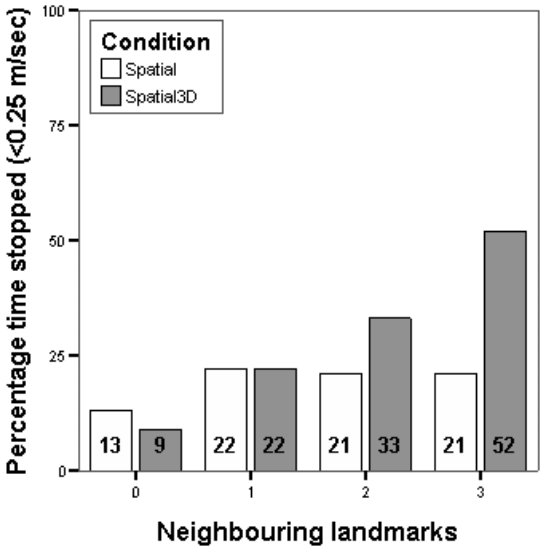

Figure 9 Percentage of time stopped for different numbers of overlapping proximity zones for audio landmarks. A threshold of less than 0.25m/sec was used to process user data identifying stationary periods.

5.2 User feedback

Based on the user feedback, the extra time spent stationary and the extra distance covered when audio spatialization was used, did not lead to frustration, rather it appears to be related to the enjoyment and sense of discovery of the participants. In contrast, for the conditions lacking audio spatialization, participants behaved more like in a navigation environment setting themselves the task of finding all the landmarks by systematically walking through the park. This behaviour emerged despite participants in all conditions being given the same set of instructions before starting the exploration of the garden. They were all told to walk through the park in their own time and without rushing or walking too fast and that audio landmarks would be triggered as they got closer to them. Overall, sound levels were reported to be appropriate and the speech was clear and intelligible. Informal user feedback is presented for each of the four auditory display conditions.

5.2.1 Baseline

In the first auditory display condition (no Earcons or spatialization), the audio clips were simply triggered when users entered the activation zone. Consequently, the users tended to systematically explore to find the audio clips. Once they were located, users reported being pleased with locating the landmark but remarked the sound was “a bit abrupt when triggered”. The value of the information in the audio clips was found to be appropriate, but especially directed towards tourists. The material in the audio clips was found “appropriate and informative” mainly due to the physical landmarks and because “if you were walking around the garden you wouldn’t like to read it”. One user suggested that the content of these audio clips “would potentially trigger a conversation” if walking with a friend or partner. The users highlighted that “the experience of moving around to get the information was good” and the “novelty of coming across the message like stumbling across something on your way. Serendipity and wonder”. Users found navigating the park to find the audio clips “very easy, just walking around” as it was “not a big space”. However, the instability of the GPS information sometimes resulted in the user overshooting the physical landmark by the time the audio clip was triggered. Users sometimes found that “the sound was triggered after walking past” or “it was quiet and thought I was on the wrong path”. One of the participants failed to find one of the audio clips, reflecting the difficulty of successfully exploring such a sound environment.

5.2.2 Earcons

Earcons were present in three of the conditions. Participants reported they “liked the sound of the animals” and described them as “lively”, “clear”, “natural”, “crisp” and “interesting”. They reported enjoying the fact that “you just walk around and the sounds get triggered”. Despite the background noises in the park, the animal sounds successfully indicated the presence of information at particular locations. One user remarked: “I liked that I realized that it [the animal sound] was prompting me to press the button. Maybe if it had been too realistic I would have missed that”. In the third condition (which adjusted volume based on distance to the landmark in the proximity zone) one of the participants reported that the animal sounds “blended very well. Made it more seamless”. The other participant felt that the echoing (reverb) in the animal sounds made him feel “like

being in a quiet place in the forest. Reminded me of a place close to home". A participant in the fourth, fully spatialized condition suggested that it "helped that they [the sounds] were different from the ones already in the park". Both participants in this final condition enjoyed the animal sounds, stating they were the "best part" and "especially nice for a garden like this one". They did not expect these animal sounds to blend so well and also found them "just playful in themselves".

5.2.3 Audio spatialization

Spatial

Participants experiencing the Spatial condition, in which amplitude of the Earcons varied with distance to target, reported this to be useful and appropriate. The intensity was reported to remain at a comfortable level throughout. However, users experienced difficulty determining the distance to particular landmarks. One stated: "guessing how close I was from a location was based on distance travelled when I first heard it and intensity combined. Not proportional" and reported that the alterations to volume were not physically accurate. The other user noted that the variations in volume were a bit "jumpy", something probably due to noise in the GPS position sensing. He also noted, that "it took time to get used to the distance distinction near/far. Once I found the first one [landmark] it was easier to find the others because I already knew what I was looking for".

Spatial3D

During the Spatial3D condition, participants reported a sense of "discovery" and that the sound garden was "quite immersive". The participants in this condition liked the experience because "you rely only on your hearing" and often closed their eyes in order to listen to the Earcons. They found the system curious because "you know sounds come from headphones but it sounds like it is coming from the outside". The variation in loudness used to represent distance away from the landmark gave "a good indication of distance" but it was also reported that "going from far away to closer was too quick". One participant stated that even in situations with multiple sound sources "overall the localization was easy" but became harder in the area of the park where three animal sounds overlapped. However, when the user walked away from this area and only two animal sounds

overlapped, heading helped. This was echoed by the opinion that while two overlapping sounds were understandable, three were "a bit chaotic". Overlapping sounds also conveyed benefits as "hearing sounds at a distance that [I] have already heard gave familiarization with the surroundings". One of the users admitted: "it would be difficult to find them [landmarks] without spatialization. If it doesn't point you in the right direction it would be harder".

5.3 User behaviour

A more detailed analysis of the logged data for each participant revealed a tendency for participants in the Baseline condition to walk at a steadier pace, in straighter lines, while looking in the direction they were going, when compared to participants in the Spatial3D condition. Figure 10a shows an example of subject 1 in the Baseline condition walking from the stone coat of arms to the statue of Joao Reis Gomez. The solid line is the direction of travel and the short splines illustrate the participant's head orientation approximately every two seconds. Figure 10b shows a contrasting path from a participant in the Spatial3D condition. The gray rings 1&2 highlight two points where the participant stopped and began looking around, probably trying to ascertain the direction of the audio being played in the proximity zone. This type of behaviour was typical of the Spatial3D condition where the head movement while stationary appears to characterise a 'searching behaviour'. If we examine the distributions of head orientation change for the spatial conditions (see Figure 11), it can be observed that the Spatial3D condition encourages this type of head movement (lower percentage of 0° data points and broader distribution) compared to the Spatial condition showing a more peaked distribution, i.e. a different kurtosis[5]. The mean and SD of both distributions are similar (Spatial: mean= 0.038, SD= 42.77; Spatial3D: mean= -1.612, SD= 50.630), however the kurtosis is quite different (Spatial: Kurtosis= 3.470; Spatial3D: Kurtosis= 1.648). This means that head change within the regions 36 degrees to 108 degrees contains more data than angles closer to 0 and wider

[5] *Kurtosis* is the name of a statistical measure used to describe the distribution of observed data around the mean. A normal distribution has a kurtosis 0, a peaked (tall and skinny) distribution has a positive or high kurtosis and a flat distribution has a negative or low kurtosis.

angles. Wider angles are likely to be caused by changes in body position. This would fit our observation that participants moved their heads from side to side in the 3D spatial condition to gauge the direction of sounds heard. Although there is no formal statistic test to compare Kurtosis, a Chi-square test on observed counts across five bins (as shown in Figure 11), showed that observed counts from the Spatial3D condition significantly differed from expected counts matched based on likelihoods calculated on observed values in the Spatial condition ($\chi^2(4, N= 1160) = 73.764, p < 0.001$). If we compare logged information from participants with limited spatial information we see they did stop as in the Spatial3D condition, but they seemed to keep their head much closer to their direction of travel (Figure 12a). Finally, Figure 12b shows one of the participants in the Spatial 3D condition within the three overlapping proximity zones. This participant shows an extreme case example of amount of head-turning to ascertain direction, which frequently occurred in the spatial conditions. This user in particular spent a substantial time walking and altering his head position in order to determine the direction of one of the landmarks. Far from frustrating, as user feedback showed, this searching process was enjoyable and added to the sound garden experience.

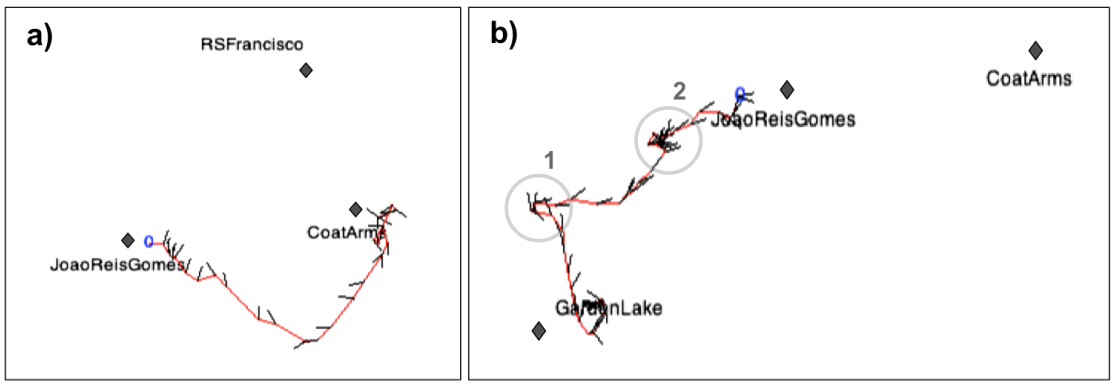

Figure 10 a) Route taken by one user from the stone coat of arms to the statue of Joao Reis Gomez during the Baseline condition. **b)** Route taken by one user from the Garden Lake to the statue of Joao Reis Gomez during the full 3D audio spatialization (Spatial3D) condition. Gray circles indicate stationary periods along the route with greater amounts of head-turning. Short splines illustrate the user head direction approx. every 2 seconds (mean= 2.28 secs, SD= 0.29).

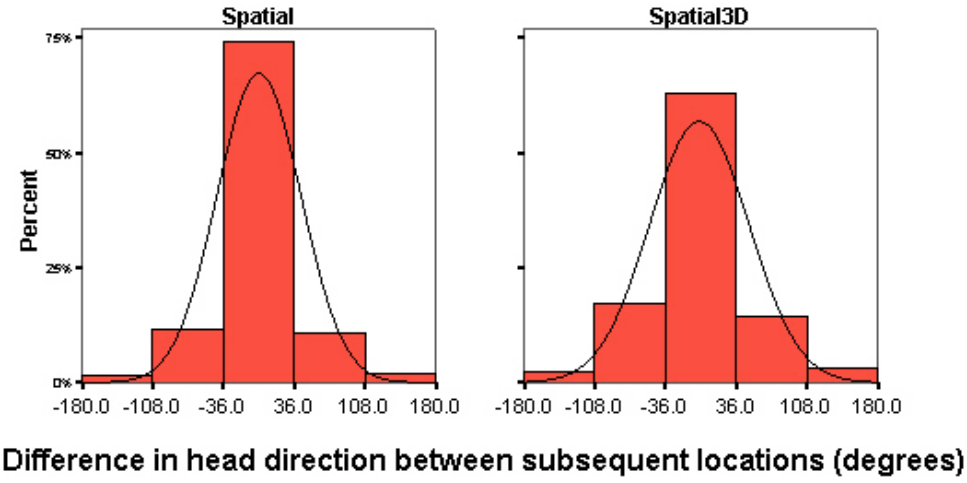

Figure 11 Histograms showing the distribution of the total amount of head-turning for the spatial conditions. Head-turning audio feedback was only provided in the Spatial and Spatial3D conditions.

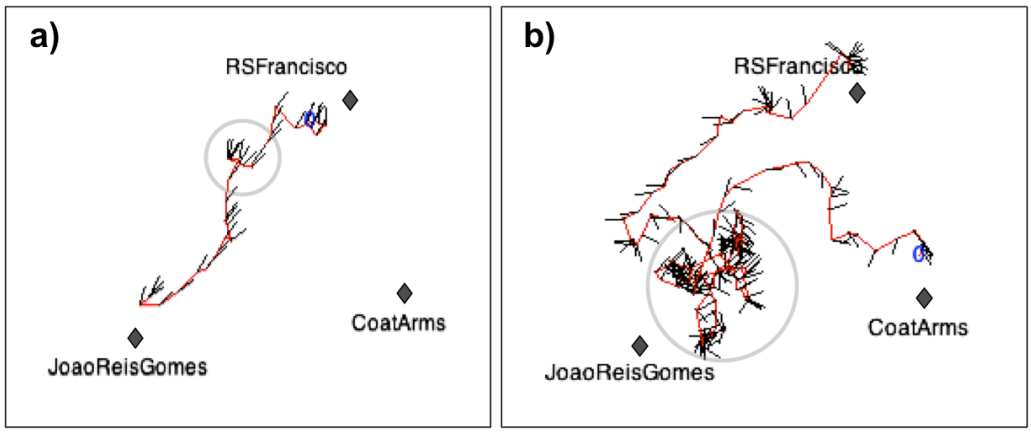

Figure 12 a) Route taken by one user from the statue of Joao Reis Gomez to the Rua Sao Francisco during the limited audio spatialization (Spatial) condition. Head direction fits much closer to the direction of travel (short splines illustrate the user head direction). **b)** Route taken by one user from the Rua Sao Francisco to the stone coat of arms during the full 3D audio spatialization (Spatial3D) condition. Head direction changes greatly in order to determine the direction of one of the landmarks as illustrated by the route data within the gray circle.

6. Discussion

In this paper we have presented an initial user study evaluating four different auditory displays in a mobile audio-augmented reality environment (a sound garden). We have compared, quantified and qualitatively described how users behavior and experience varied when exposed to the different configurations of proximity zones, non-speech sound and spatial 3D audio available in our sound garden. In addition, we have examined head-turning data and its relation to user behaviour in exploratory environments, with particular attention devoted to situations when multiple audio landmarks overlap. Although this study did not examine a large user sample, the wide range of measurements recorded were able to support a rich, detailed and informative analysis.

To answer our research questions, the results show that when users were provided with spatial audio feedback within the proximity zone, they spent more time in the park, walked more and spent more time stationary while turning their heads searching for landmarks. When distance away from the landmark was the only spatial audio cue available, some participants reported it to be useful while others were confused by the relationship between virtual and physical distance. GPS error also appeared to interfere with the overall experience. However, in the Spatial3D condition, participants reported that the audio feedback gave a good indication of distance and was more immersive.

Users reported they were able to resolve two overlapping sounds easily but when three sounds overlapped, although heading information was a great help, localizing the sounds became harder. As a result, participants' average speed dropped more when they were provided with spatial 3D audio feedback, as they had to stop to search and ascertain the direction of the audio, which was not the case when distance was the only cue available. However, far from frustrating users, they appeared to take their time to enjoy the sense of discovery (on average 21 minutes when spatialized compared to 11.49 when not spatialized), and immerse themselves in this mobile audio-augmented reality environment. We also found that Earcons played an important role as a playful element successfully indicating the presence of information at a particular location. In contrast, when users were not provided with spatial audio feedback, they systematically explored (as in a navigation task) at a steadier pace in straight lines mainly looking in the direction they were walking. In the non-spatial conditions, only the animal sounds

were reported to provide a touch of playfulness to the exploration. However, users remarked on the "abruptness" of walking right into the audio clips and GPS error had a worse effect on the user experience in this case.

The quantitative data presented in this paper aimed at describing users' exploratory behaviour. We found that head position data and the details of the participants' movement in a mobile audio-augmented reality environment are of critical importance to fully understand their behaviour in such environments. Using blunt averages of speed or task completion times are less likely to show meaningful differences in user behaviour. In this study we tracked head orientation using a magnetometer attached to the middle of the headphone's headband. By doing this, it was difficult to differentiate between head turn and body turn. In future work, tracking body turn with the aid of an additional JAKE sensor mounted on the shoulder would help when analysing this kind of data.

A number of technical limitations affected this study. Firstly, as known from previous work, GPS can be problematic when seeking to situate audio precisely in space. However, GPS technology is increasingly present in smartphones and becoming ever more popular in mobile applications making use of geo-tagged data. Participants in this study did complain that the audio garden was jerky and unpredictable at times due to variance in the position reported by the GPS unit. Despite this system limitation, a high level of immersion was reported by users when exposed to spatial 3D audio and the combination of proximity and activation zones around the landmarks helped minimize GPS error.

Four separate devices were used in the system: a GPS unit, a magnetometer unit, a mobile phone and a pair of headphones. This was a somewhat overwhelming collection of devices and there would be many benefits to creating a more integrated solution. However, as the sensors were all situated on the headphones, one key advantage of this solution is that it enabled true 3D audio interaction based on head position and orientation. It is not clear the sound garden would be as compelling if all sensing was integrated into a handheld device, but further work is required to explore this issue.

Our results build on previous work by extending and evaluating the complexity of the audio spaces used previously for exploration in audio-augmented environments [9,10,11,12,13,14]. Moreover, this study offers an initial qualitative and quantitative insight into overlapping spatialized sounds in a realistic

environment, a design feature first implemented by Stahl [3] but never evaluated. In particular, our findings on the critical importance of head position data in spatialized mobile audio-augmented environments confirm and complement those by Heller *et al.* [22] and Mariette [23]. Ultimately, this work follows up on recent studies describing the design of purely exploratory audio-augmented environments such as Heller et al.'s CORONA [22] and Magnusson et al.'s Soundcrumbs [14], rather than on navigational tasks [e.g. 11,12,13]. As in Heller and Magnusson's work, the non-speech sounds used to identify the landmarks created an enjoyable and "playful" experience, despite increasing the audio feedback complexity due to their spatially overlapping nature.

A number of practical lessons were also learned regarding the creation of audio-driven sound gardens. For example, although the circular activation zones used in this work are simple and easy to understand, they are a poor fit for the complexities of a space with paths, hedges and trees. There is a clear tension between situating sounds at the correct geographical location and situating them at a place where it is possible to ensure that users can observe the target item. With activation radii of 10 or more meters, users can easily encounter sounds from behind barriers such as walls or dense plants, a potentially confusing situation. One clear way to address this is through developing non-circular activation regions, but this may also cause problems, as the realism of the metaphor connecting the virtual sounds to physical spaces may break down. Other solutions may include dynamically adjusting activation zones, or calculating optimal solutions, which maximize the size of all zones (as in the bubble cursor [26]). Exploring richer interactions with the sound sources would also be beneficial. In this work, users were able to explore a physical space and press a button to start an audio clip. By allowing other interactions such as silencing, moving, adjusting or otherwise interacting with audio in a sound garden, it may be possible to create denser audio environments, which remain simple, effective and engaging. Although further work is required, the initial findings and methods presented in this paper provide a valuable framework for the analysis and description of user behaviour in mobile audio-augmented reality environments.

7. CONCLUSIONS

In conclusion, the combination of 3D spatial audio techniques together with Earcons was the most effective auditory display. In addition, capturing user position and head orientation has been shown to be an effective means of describing participants' exploratory behaviour in an audio-augmented reality system, such as the one presented in this paper. This work suggests that the location and orientation sensing technologies now present in commercially available smartphones can be used to create rich and compelling outdoor audio-augmented environments.

Acknowledgements

This work was supported by the Ken Browning Travelling Scholarship (University of Glasgow, UK), Nokia and EPSRC research grant EP/F023405 "Gaime". We would like to express our gratitude to the members of the Madeira-ITI group at Madeira University who participated in this research project.

References

1. Shepard, M. Tactical Sound Garden [TSG] Toolkit. 3rd International Workshop on Mobile Music Technology, Brighton, UK (2006).
2. Walker, B. N. and Lindsay, J. Navigation Performance With a Virtual Auditory Display: Effects of Beacon Sound, Capture Radius, and Practice. *Human Factors*, 48, 2 (2006), 265-278.
3. Stahl, C. 2007. The roaring navigator: A group guide for the zoo with shared auditory landmark display. In Proceedings of MobileHCI 2007. ACM Press, New York, NY, USA (2007), pp. 282-386.
4. Blauert, J. Spatial Hearing: The psychophysics of human sound localization. The MIT Press, (1999).
5. Begault, D. R. 3D sound for virtual reality and multimedia. Boston, MA: AP Professional, (1994).
6. Mariette, N. From Backpack To Handheld: The Recent Trajectory Of Personal Location Aware Spatial Audio. PerthDAC 2007: 7th digital arts and culture conference, Proceedings, A. Hutchison. Curtin University of Technology, Perth, Australia, (2007), pp. 233-240.
7. Mynatt, E., Back, M. and Want, R., Baer, M., and Ellis, J. B. Designing Audio Aura. In Proceedings of CHI 1998. ACM Press, New York, NY, USA (1998), pp. 566-573.

8. Marentakis, G.N. and Brewster, S.A. Effects of Feedback, Mobility and Index of Difficulty on Deictic Spatial Audio Target Acquisition in the Horizontal Plane. In Proceedings of ACM CHI 2006. ACM Press, New York, NY, USA, pp 359-368.
9. Rozier, J., Karahalios, K. and Donath. J. Hear & There: An augmented reality system of linked audio. In Proceedings of the International Conference on Auditory Display – ICAD 2000.
10. Reid, J., Geelhoed, E., Hull, R., Carter, K. and Clayton, B. Parallel worlds: Immersion in location-based experiences. In Proceedings of CHI 2005. ACM Press, New York, NY, USA (2005). vol. 2, pp. 1733-1736.
11. Holland, S., Morse, D. R. and Gedenryd, H. AudioGPS: spatial audio in a minimal attention interface. Personal and Ubiquitous Computing (2002). Volume 6(Issue 4): 253-259.
12. Cater, K., Hull, R., O'Hara, K., Melamed, T. and Clayton, B. The potential of spatialised audio for location based services on mobile devices: Mediascapes. SAMD: Workshop on Spatialised Audio for Mobile Devices, MobileHCI 2007.
13. McGookin, D., Brewster, S. and Priego, P. Audio Bubbles. Employing Non-speech Audio to Support Tourist Wayfinding. M.E.Altinsoy, U.Jekosch, andS.Brewster(Eds.): HAID 2009, LNCS 5763, (2009), pp. 41-50.
14. Magnusson, C., Breidegard, B. and Rassmus-Gröhn, K. Soundcrumbs – Hansel and Gretel in the 21st century. In: HAID 2009, Springer LNCS (2009).
15. Blattner, M. M., Sumikawa, D. A., and Greenberg, R. M. Earcons and icons: Their structure and common design principles. Human Computer Interaction (1989), 4(1): 11-44.
16. Gaver, W. W. Auditory interfaces. M. G. Helander, T. K. Landauer, and P. V. Prabhu, editors, Handbook of Human-Computer Interaction. Elsevier, Amsterdam, 2nd edition, (1997), pp. 1003-1041.
17. Lemordant, J. and Guerraz, A. Mobile Immersive Music. In Proc. of the 2007 Int. Computer Music Conference, ICMC 2007. ICMA, San Francisco (2007), pp. 21-24.
18. Etter, R. and Specht, M. Melodious Walkabout: Implicit Navigation with Contextualized Personal Audio Contents. Sankt Augustin, Germany: Fraunhofer Institute of applied Information Technology. (2005).
19. Jones, M., Jones, S., Bradley, G., Warren, N., Bainbridge, D. and Holmes, G. ONTRACK: Dynamically adapting music playback to support navigation. Personal and Ubiquitous Computing (2008). Volume 12(7): 513-525.
20. Strachan, S., Eslambolchilar, P. and Murray-Smith, R. gpstunes - controlling navigation via audio feedback. MobileHCI 2005. ACM, New York, NY, USA (2005), vol. 1, pp. 275-278.
21. Lyons, K., Gandy, M. and Starner, T. Guided by Voices: An Audio Augmented Reality System. Intl. Conf. on Auditory Display (ICAD), Atlanta, GA, (2000), pp. 57-62.
22. Heller, F., Knott, T., Weiss, M. and Borchers, J. Multi-user interaction in virtual audio spaces. CHI Extended Abstracts of ACM CHI 2009, ACM Press (2009), pp. 4489-4494.

23. Mariette, N. Navigation Performance Effects of Render Method and Head-Turn Latency in Mobile Audio Augmented Reality. CMMR/ICAD 2009, LNCS 5954, (2010), pp. 239-265.
24. Brungart, D. S., Simpson, B. D. and Kordik, A. J. The detectability of headtracker latency in virtual audio displays. In Proceedings of the 11th International Conference on Auditory Display (ICAD), Limerick, Ireland, (2005), pp.37-42.
25. Vazquez-Alvarez, Y. and Brewster, S. A. Designing Spatial Audio Interfaces to Support Multiple Audio Streams. In Proceedings of MobileHCI 2010. ACM Press, New York, NY, USA (2010), pp. 253-256.
26. Grossman, T. and Balakrishnan, R. The bubble cursor: enhancing target acquisition by dynamic resizing of the cursor's activation area. In Proceedings of CHI 2005. ACM Press, New York, NY, USA (2005), pp. 281-290.