Multilevel Auditory Displays for Mobile Eyes-Free Location-Based Interaction

Yolanda Vazquez-Alvarez
GIST
School of Computing Science
University of Glasgow,
Glasgow G12 8QQ, UK.
Yolanda.Vazquez-Alvarez@glasgow.ac.uk

Rocio von Jungenfeld
Edinburgh College of Art
University of Edinburgh
Edinburgh EH8 9DF, UK.
rocio.von-jungenfeld@ed.ac.uk

Matthew P. Aylett
CereProc, Ltd.
University of Edinburgh
Edinburgh EH8 9LE, UK.
matthewa@inf.ed.ac.uk

Antti Virolainen
Nokia Research Centre
Helsinki, Finland.
antti.virolainen@gmail.com

Stephen A. Brewster
GIST
School of Computing Science
University of Glasgow
Glasgow G12 8QQ, UK.
Stephen.Brewster@glasgow.ac.uk

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s).
CHI 2014, Apr 26 - May 01 2014, Toronto, ON, Canada
ACM 978-1-4503-2474-8/14/04.
http://dx.doi.org/10.1145/2559206.2581254

Abstract

This paper explores the use of multilevel auditory displays to enable eyes-free mobile interaction with location-based information in a conceptual art exhibition space. Multilevel auditory displays enable user interaction with concentrated areas of information. However, it is necessary to consider how to present the auditory streams without overloading the user. We present an initial study in which a top-level exocentric sonification layer was used to advertise information present in a gallery-like space. Then, in a secondary interactive layer, three different conditions were evaluated that varied in the presentation (sequential *versus* simultaneous) and spatialisation (non-spatialised *versus* egocentric spatialisation) of multiple auditory sources. Results show that 1) participants spent significantly more time interacting with spatialised displays, 2) there was no evidence that a switch from an exocentric to an egocentric display increased workload or lowered satisfaction, and 3) there was no evidence that simultaneous presentation of spatialised Earcons in the secondary display increased workload.

Author Keywords

Eyes-free interaction; auditory displays; spatial audio

ACM Classification Keywords

H.5.2 [User Interfaces]: Interaction styles, evaluation.

Introduction

Audio-Augmented Reality (AAR) enables users to interact with location-based information purely through sound while on the move. This is particularly useful when the users' visual attention is already being compromised by real visual objects in the surrounding environment. Consider the following scenario: *There is a conceptual art exhibition in London and art lover, David, has arranged a visit with his friend Rocio. Before they enter the gallery, they download an application onto their mobile phone that will enable them to listen to information about the art pieces using their headphones while walking around the exhibition. As they get close to an audio-augmented location, different sounds allow users to browse the audio information available. This varies between comments left by visitors, the artist herself and an art critic. At one artifact, Rocio selects a comment left by a previous visitor that says the piece reminds him of a circulatory system. David selects a comment left by the artist, which describes how the frame squeezes wool of different colours to contrast the 2D nature of the photo frame with the 3D element of materials. David and Rocio have a lively discussion based on these comments. They agree that comments provided by the artist helped them appreciate the ideas in the work, while the opinions left by other visitors mentioned things they would never thought of themselves. Overall, the result is a personalised museum experience, which has responded to the individual user interests and encouraged them to appreciate and enjoy the art work in more depth in their own way.*

As illustrated in our example, location-based information can be presented using AAR. When using such an eyes-free auditory interface, each location being augmented requires the use of an audio stream which means it may be necessary to discriminate between them, especially when they could overlap if locations are close to each other. Spatial audio has been used in previous research [8, 12] as a successful technique to segregate multiple audio streams by placing each audio stream at a different location around the user's head, mirroring how humans perceive sounds in real life. When designing auditory displays for mobile audio-augmented environments, choices have to be made on both the audio presentation and the spatial arrangement of the audio streams. A multilevel auditory display allows concentrated areas of information to be structured in a location-based system. However, given a top-level spatial auditory display, should a secondary display also be spatialised and if so, how? Should information be provided sequentially or simultaneously? While simultaneous presentation is important to create a rich immersive audio environment, high levels of workload may affect exploration and selection between different locations and also the exploration and selection of the various amounts of information provided at each location.

Background Work

In 1993 AAR was proposed as the action of superimposing virtual sound sources upon real world objects [4]. The key idea is that users can explore an acoustic virtual environment augmenting a physical space solely by listening as they walk [6, 12]. The majority of indoor AAR systems have been developed for museums, exhibitions or historic sites in order to replace linear keypad-based audio tour guides that constrained users to a linear access to information and could pull the visitor's attention away from the actual exhibits and disturb the overall user experience. Bederson's automated tour guide [1] was an early example of an exploratory non-linear playback system. Since this early prototype, indoor AAR applications have grown in complexity providing a much greater amount of information about the audio-augmented locations [13, 6].

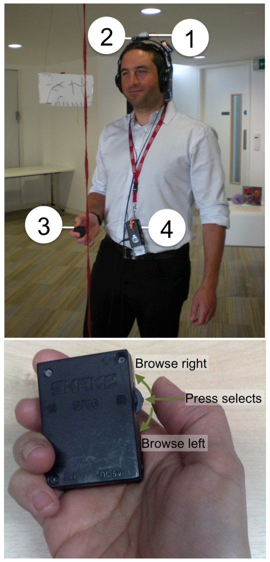

Figure 1: Experimental setup. Top: 1) IR tag, 2) JAKE sensor (both mounted on headphones), 3) SHAKE SK6 sensor pack and 4) mobile device. Bottom: Detail of SHAKE SK6 navigation switch interface.

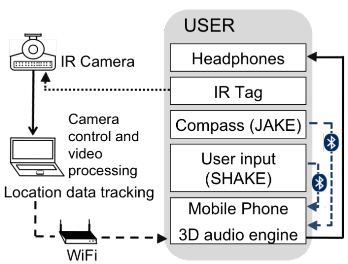

Figure 2: Schematic of the system architecture.

User interaction with location-based information in a exploratory indoor mobile AAR application usually takes place within a proximity and an activation zone [9]. The proximity zone advertises an audio-augmented location and the activation zone presents more detailed information. The amount of information presented within the activation zone in previous systems has varied greatly, from one [1] to multiple [13, 6] audio streams. As the amount of information presented increases, more complex auditory displays are required. Spatial audio techniques enable user interaction with multiple audio streams by providing orientational information that aids segregation and attention switching between the audio streams [10]. However, how should a spatial auditory display be designed in order to support increasing amounts of information? Previous work has used spatial audio to design mobile auditory displays presenting information sequentially in an egocentric [5] or exocentric [11] display; or simultaneously in an egocentric [8] or exocentric display [3].

However, the usability of these designs has not been compared against each other as part of an interactive audio-only environment. In an egocentric display, elements are always in a fixed position relative to the user, which can be particularly useful for mobile users as changes in orientation when moving are inevitable. In an exocentric display, on the other hand, display element positions have to be updated real-time according to the user orientation as they appear to be fixed to the world. This can aid user navigation but can be computationally intensive for a mobile device. Previous research has shown that interacting with an egocentric display when mobile is faster but more error prone than an exocentric display [7] and that simultaneous presentation in an exocentric display allows for faster user interactions [3]. However, to our knowledge, no previous research has compared the use of combined ego- and exocentric designs within the same spatial hierarchical auditory display. This paper presents an initial evaluation of a complex multilevel spatial auditory display including egocentric and exocentric designs and sequential and simultaneous presentation, tested within a mobile AAR environment.

Audio-augmented Art Exhibition

A conceptual art exhibition was used as the setting for all conditions in this study. The virtual audio environment was run on a Nokia N95 8GB and the built-in head related transfer functions (HRTFs) were used to position the audio sources. User position was determined using an Infrared (IR) camera tracking an IR tag powered by a 9V battery and mounted on top of a pair of headphones. Coordinate information was fed to the mobile phone over an Internet connection and was used to activate the zones associated with the art pieces.

User orientation was determined using a JAKE[1] sensor pack connected to the mobile phone via Bluetooth. No visual aids were provided on the mobile device and, to ensure a full eyes-free experience, the phone was placed on a lanyard around the user's neck (see Figure 1 - top). A SHAKE SK6 sensor pack[2] was held the user. It was also connected via Bluetooth and the navigation switch provided input. This navigation switch allowed users to activate, browse, select and deactivate audio content (see Figure 1 - bottom). The audio was played over a pair of DT431 Beyerdynamic open headphones with the aim to reduce the isolation of the listener from the surrounding environment. The IR tag and JAKE sensor were placed in the middle of the headphone's headband. Figure 2 shows the final system setup. The augmented exhibits

[1]http://code.google.com/p/jake-drivers
[2]http://code.google.com/p/shake-drivers

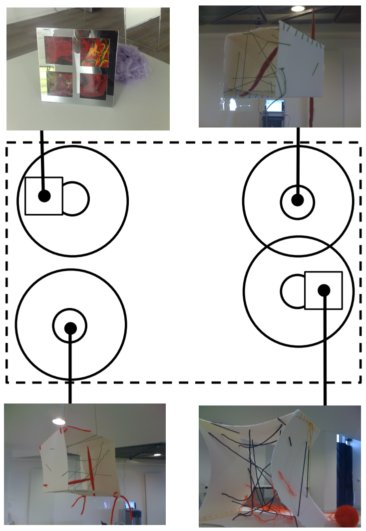

Figure 3: Illustration of the top-level sonification layer showing the location of the circular proximity (radius 1.25m) and activation zones (radius 0.75m). Small squares with a dot at its centre identify art pieces placed on a table.

consisted of four different art pieces from the *Weaving the City* project (www.weavingthecity.eu) situated in an exhibition space measuring 3m by 3.85m see Figure 3).

Top-level sonification layer
This layer used an exocentric design to present a chattering voices sound within the proximity zone advertising content about each art piece. The sound increased in volume when the user reached the activation zone in which a secondary interactive layer could then be activated/deactivated (see Figure 3 for an illustration of the setup) by pressing down (long press, $>$ 2 secs) on the SHAKE navigation switch (see Figure 1 - bottom).

Secondary interactive layer
For each art piece, different Earcons were used in an audio menu. Earcons provide an abstract and symbolic relationship between the sounds and the information they are representing [2]. The Earcons identified 1-3 different audio menu items as follows: "water waves": for the artist's comments, "open crackling fire": for positive non-expert reviews, and "stormy wind": for negative (cold) non-expert reviews. To select one of the menu items the SHAKE navigation switch was pressed down (short press $<$ 2 secs). Once the menu item was selected, an approx. 25 secs long audio clip was played containing the comment or review. User interaction with the audio menu items varied for the different experimental conditions:

Baseline: Each Earcon was *always* played sequentially at each push of the navigation switch either right or left. There was no spatialisation of the audio items so they seemed to originate from within the user's head. The aim was to recreate a traditional audio guide style interaction in which users triggered the audio content by the press of a button in a sequential order (see Figure 4a).

Egocentric Sequential: Each Earcon was presented in a radial menu (virtually located around the user's head to the right, left or in front of the user's nose) and played one at a time when selected by pushing the navigation switch for the sequential presentation group. Selection was performed by pushing the navigation switch either right or left and the Earcons were located at 0°, -90° and +90° azimuth (see Figure 4b).

Egocentric Simultaneous: Similar to the egocentric sequential condition except that all of the Earcons were played simultaneously. When a menu item was selected, the volume increased for that selected item to bring it into focus and decreased for the rest (see Figure 4c).

Study
Thirty-two participants (21 males, 11 females, aged 18 to 39 years) were recruited, all were studying or working at the University. They all reported normal hearing, were right-handed and paid £6 for participation, which lasted just over an hour.

Experimental Design and Measures
Participants were split equally into two groups: sequential and simultaneous presentation, in a between-subjects design. The baseline condition was used as a control in both groups and the order of conditions was randomised. Three dependent variables were analysed: perceived workload, overall user satisfaction and time spent while interacting with the secondary display. In addition, user location coordinates and head orientation data were also collected to investigate user behaviour. The following hypotheses were tested:

H1: A spatialised secondary display will increase exploration. We define increased exploration as users spending significantly more time without a significant drop in user satisfaction or a significant increase in perceived workload.

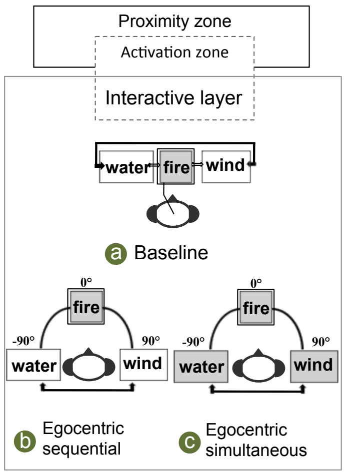

Figure 4: Schematic of the interactive auditory displays tested.

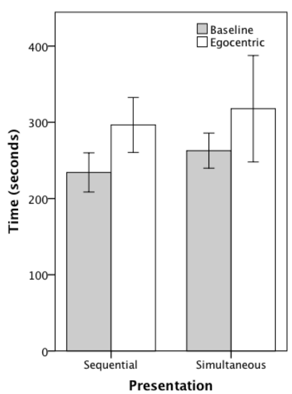

Figure 5: Mean time taken interacting with the secondary interactive layer per condition and presentation group. Baseline was used as a control for both egocentric conditions.Total audio content: 9 audio clips x 25 secs = 225 secs. Error bars show Standard Error of Mean ± 2.0.

H2: Changing between the exocentric and the egocentric layers will increase perceived workload.

H3: Simultaneous presentation of Earcons in the egocentric secondary interactive layer will increase perceived workload.

Procedure

The experiment started with a training session before the test conditions to familiarise the participant with the multi-level auditory displays around one of the art pieces in the exhibition space. For each test condition, participants were asked to explore the exhibition space and find as much information as possible about the art pieces by interacting with the different auditory displays. Participants were given a maximum of 10 minutes of exploration time for each test condition. There was no minimum time and participants could choose to stop whenever they wanted. All the participants had time to explore the art pieces in the allocated time. After each test condition, participants were asked to complete a NASA-TLX subjective workload assessment, a satisfaction questionnaire that was modified from that used in Wakkary and Hatala [13] and also provide some informal feedback.

Results

To test hypotheses 2 and 3, a two-way mixed-design ANOVA was performed on the overall perceived workload and overall user satisfaction mean scores with condition type as a within-subjects factor and presentation group as a between-subjects factor. Overall workload was calculated across all six subscales to a maximum of 120 mean score and overall user satisfaction was calculated as an average over 62 satisfaction questions rated on a five-point scale where 5 was 'best'. No significant results are reported for the different conditions, Baseline *versus* Egocentric display, or for Simultaneous *versus* Sequential presentation; or for interactions between these variables, either for perceived workload or overall satisfaction. Thus, hypotheses 2 and 3 can be rejected. Across all conditions and presentations mean overall user satisfaction was high at 3.9/5 and mean overall workload was low at 32/120.

A two-way mixed-design ANOVA showed significantly more time was spent in the spatialised Egocentric conditions than in the baseline condition ($F(1,30)=8.21$, $p=0.008$), see Figure 5. As there was no evidence that spatialisation increased perceived workload, extra time spent can be attributed to an increase in user exploration and therefore hypothesis 1 can be accepted.

The logged user behaviour data showed a much simpler pattern of exploration for participants in the Baseline than in the Egocentric conditions. Figure 6 (top) shows an example of one of the participants in the Baseline condition walking in a straighter trajectory between the art pieces and then staying mainly stationary once the secondary interactive layer was activated. Figure 6 (bottom) shows the same participant taking more time to explore around the art pieces once the secondary interactive layer was activated in the Egocentric simultaneous condition.

Conclusions and Future Work

This paper investigated the usability of complex spatial auditory displays designed to enable user interactions with concentrated areas of information in an exploratory mobile audio-augmented reality environment. Both egocentric and exocentric designs were combined in the multilevel auditory display to test whether these 3D audio techniques would encourage an exploratory behaviour. Informal feedback suggests that the egocentric secondary interactive layer allowed for a more exploratory experience. Although overall the baseline condition was reported as "easy to use", it was also found to be "less im-

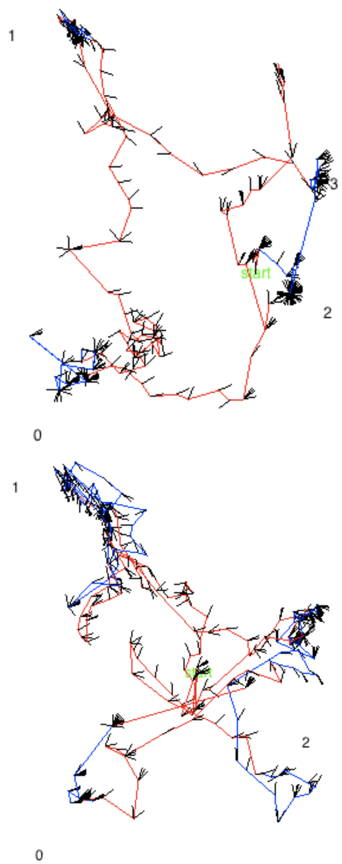

Figure 6: Route taken around the art pieces (0-3) by one participant from the *simultaneous* presentation group in the Baseline (top) and Egocentric (bottom) conditions. Solid red and blue lines show path of exploration in the sonification layer and the secondary interactive layer respectively. Short splines illustrate the participant's head direction every 0.5 second.

mersive". Users liked the control over the interaction with the location-based information provided by the egocentric design and remarked on how this "spatialised interface was more fun than simply scrolling through sounds". Performance results supported user feedback and helped characterise the egocentric secondary display as an exploratory experience, where interaction times increase without an increase in workload and a decrease of user satisfaction. The egocentric design performed well across presentation type with no evidence that the transition from the top-level exocentric layer into the egocentric secondary interactive layer had a negative impact.

These results show that spatial audio can encourage an immersive experience and an exploratory behaviour. Future work will investigate more closely how this sense of immersion could be further encouraged, for example with the use of an exocentric secondary display, and how such an immersive experience might be shared across multiple users interacting with the same display. We believe that a deeper understanding of the extent to which novel and more complex auditory displays can impact the user experience will allow designers to make more informed decisions when designing eyes-free auditory interfaces for mobile audio-augmented reality environments.

References

[1] Bederson, B. B. Audio augmented reality: A prototype automated tour guide. In *CHI'95*, vol. 2 (1995), 210–211.

[2] Blattner, M. M., Sumikawa, D. A., and Greenberg, R. M. Earcons and icons: Their structure and common design principles. *Human-Computer Interaction 4*, 1 (1989), 11–44.

[3] Brewster, S. A., Lumsden, J., Bell, M., Hall, M., and Tasker, S. Multimodal 'eyes-free' interaction techniques for wearable devices. In *CHI'03* (2003), 463–480.

[4] Cohen, M., Aoki, S., and Koizumi, N. Augmented audio reality: Telepresence/vr hybrid acoustic environments. In *2nd IEEE Intern. Wrkshp. on Robot and Human Comm.* (1993), 361–364.

[5] Dicke, C., Wolf, K., and Tal, Y. Foogue: eyes-free interaction for smartphones. In *MobileHCI'10* (2010), 455–458.

[6] Eckel, G. Immersive audio-augmented environments: The listen project. In *IV'01*, IEEE Computer Society Press (2001), 571–573.

[7] Marentakis, G., and Brewster, S. Effects of feedback, mobility and index of difficulty on deictic spatial audio target acquisition in the horizontal plane. In *CHI'06* (2006), 359–368.

[8] Sawhney, N., and Schmandt, C. Nomadic radio: Speech and audio interaction for contextual messaging in nomadic environments. *Trans. on Computer-Human Interaction 7*, 3 (2000), 353–383.

[9] Stahl, C. The roaring navigator: A group guide for the zoo with a shared auditory landmark display. In *MobileHCI'07* (2007), 282–386.

[10] Stifelman, L. The cocktail party effect in auditory interfaces: a study of simultaneous presentation. Tech. rep., MIT Media Laboratory, 1994.

[11] Terrenghi, L., and Zimmermann, A. Tailored audio augmented environments for museums. In *IUI'04* (2004), 334–336.

[12] Vazquez-Alvarez, Y., Oakley, I., and Brewster, S. A. Auditory display design for exploration in mobile audio-augmented reality. *Personal and Ubiq. Computing 16*, 8 (2012), 987–999.

[13] Wakkary, R., and Hatala, M. Situated play in a tangible interface and adaptive audio museum guide. *Personal and Ubiq. Comp. 11*, 3 (2007), 171–191.