Designing Interactions with Multilevel Auditory Displays in Mobile Audio-Augmented Reality

YOLANDA VAZQUEZ-ALVAREZ, University of Glasgow
MATTHEW P. AYLETT, CereProc, Ltd. & University of Edinburgh
STEPHEN A. BREWSTER, University of Glasgow
ROCIO VON JUNGENFELD, Edinburgh College of Art
ANTTI VIROLAINEN, Nokia Research Centre

Auditory interfaces offer a solution to the problem of effective eyes-free mobile interactions. In this article, we investigate the use of multilevel auditory displays to enable eyes-free mobile interaction with indoor location-based information in non-guided audio-augmented environments. A top-level exocentric sonification layer advertises information in a gallery-like space. A secondary interactive layer is used to evaluate three different conditions that varied in the presentation (sequential *versus* simultaneous) and spatialisation (non-spatialised *versus* egocentric/exocentric spatialisation) of multiple auditory sources. Our findings show that: 1) Participants spent significantly more time interacting with spatialised displays; 2) using the same design for primary and interactive secondary display (simultaneous exocentric) showed a negative impact on the user experience, an increase in workload and substantially increased participant movement; and 3) the other spatial interactive secondary display designs (simultaneous egocentric, sequential egocentric, and sequential exocentric) showed an increase in time spent stationary but no negative impact on the user experience, suggesting a more exploratory experience. A follow-up qualitative and quantitative analysis of user behaviour support these conclusions. These results provide practical guidelines for designing effective eyes-free interactions for far richer auditory soundscapes.

Categories and Subject Descriptors: H.5.2 [**User Interfaces**]: Evaluation, Interaction Styles

General Terms: Design, Experimentation, Human Factors

Additional Key Words and Phrases: Eyes-free interaction, auditory displays, spatial audio, mobile audio-augmented reality, exploratory behaviour

ACM Reference Format:
Yolanda Vazquez-Alvarez, Matthew P. Aylett, Stephen A. Brewster, Rocio von Jungenfeld, Antti Virolainen, 2014. Designing interactions with multilevel auditory displays in mobile audio-augmented reality. *ACM Trans. on Computer-Human Interaction* , , Article A (January YYYY), 31 pages.
DOI: http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

In 1993 AAR (Audio-Augmented Reality) was proposed as the action of superimposing virtual sound sources upon real world objects [Cohen et al. 1993]. The key idea is

This work was jointly funded by Nokia and the "Gaime Project" through the EPSRC research grant EP/F023405.
Author's addresses: Y. Vazquez-Alvarez and S. A. Brewster, Glasgow Interactive Systems Group, School of Computing Science, University of Glasgow, Glasgow, G12 8QQ, UK; M. P. Aylett, CereProc, Ltd. & University of Edinburgh, Edinburgh EH8 9LE, UK; R. Von Jungenfeld, Edinburgh College of Art, University of Edinburgh, Edinburgh EH8 9DF, UK; A. Virolainen, Nokia Research Centre, Helsinki, Finland. Contact's email: Yolanda.Vazquez-Alvarez@glasgow.ac.uk.

DOI: http://dx.doi.org/10.1145/0000000.0000000

that users can explore an acoustic virtual environment augmenting a physical space solely by listening as they walk [Eckel 2001; Vazquez-Alvarez et al. 2012; McGookin and Brewster 2012; Betsworth et al. 2013]. 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, indoor 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. This is especially relevant in indoor environments where audio-augmented locations are often situated closer to each other. Spatial audio techniques aid segregation and attention switching between 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 [Bronkhorst 2000], and thus facilitating user interaction with multiple audio streams [Stifelman 1994]. However, how should a spatial auditory display be designed in order to support increasing amounts of information?

A multilevel auditory design can be used to address the problem of structuring larger amounts of information. While there is currently no strict definition of an auditory display, it can be defined as "the use of sound to communicate information about the state of an application or computing device to a user" [McGookin 2004]. Such displays have also been called *auditory interfaces* [Gaver 1997]. A multilevel auditory display can thus be defined as an auditory display in which information is presented as a tree-like structure with several levels. Different designs of a multilevel auditory display will have to take into account both the audio presentation and the spatial arrangement of the audio streams. Gaver [1997] proposed three dimensions to define a design space that encompassed auditory interfaces: 1) the choice of sounds: simple multidimensional tones, to musical streams, to everyday sounds; 2) the way sounds are mapped to information: completely arbitrary mappings on the one hand to metaphorical and literal ones on the other; and 3) the kinds of functionality that sounds have provided. He later discusses the potential of spatial audio and the importance of auditory interfaces in portable handheld devices. Nowadays, with the availability of real-time spatial audio on mobile devices, the use of spatial audio adds a fourth design factor to the design space.

The approach this work takes in the face of this very large possible design space is driven by the requirements of mobile audio-augmented environments. In these environments sounds must be embedded in physical spaces by tracking the position of the user, thus conditioning the design of any auditory interface we can explore. Previous work [Vazquez-Alvarez et al. 2012] has looked at how best to design such interfaces using a limited amount of audio content. The desire to increase the amount of content we can use to augment the space is the main motivation behind the exploration of multi-level auditory displays presented here. This is realised by experimenting with a secondary level display in a number of configurations.

In a location-based system, a top level functions as a sonification layer, where virtual sounds are associated with real locations. A proximity zone is often used in a sonification layer to surround audio-augmented locations in order to provide unobtrusive audio guidance that enables a user to move towards an activation zone. An activation zone can then be used to access a secondary level containing additional audio information linked to that location. Previous research by Vazquez-Alvarez et al. [2012] showed that a spatialised top level was able to deliver a more engaging and immersive user experience than a number of other non-spatialised alternatives and aided users' exploration when audio-augmented locations overlapped. However, interactions with multiple information items within the secondary level were not investigated. If multiple information items must be supported in the secondary level of the auditory display, how should such an interface be designed?

In order to design an interactive secondary level containing multiple information items, we need to consider how to present the auditory streams without overloading the user. Given a top-level spatial auditory display, should the secondary display also be spatialised and if so, how? Should we mirror the presentation arrangement displayed in the top layer or would other designs, such as a combination of exocentric[1] top level and an egocentric[2] secondary level, be more usable and efficient? A homogeneous design across levels in the auditory display would follow the design principle of consistency. Consistency is a widely used principle in user interface design [Helander et al. 1997; Shneiderman 1998; Nielsen 1994] and it has been found to impact both usability and cognitive load [Lund 1997]. In visual interfaces "consistency allows users to transfer existing knowledge to new tasks, learn new things more quickly, and focus more on tasks because they need not spend time trying to remember the differences in interaction. By providing a sense of stability, consistency makes the interface familiar and predictable" [Microsoft 1995]. When designing visual interfaces that display large amounts of data, multiple visual levels have been suggested in order to improve usability and reduce cognitive load [Lam and Munzner 2010]. A Zoomable User Interface (ZUI) is an example of such an interface. ZUIs have been defined as *"systems that support the multi-scale and spatial organisation of and magnification-based navigation among multiple documents or visual objects"* [Bederson 2011]. In a ZUI, multiple visual information can be presented simultaneously using a *consistent* multilevel layout in which the user can navigate to different zoom levels by zooming up close to interact with detailed content or zooming out for an overview. In an auditory display, this overview+detail structure could be supported in a multilevel auditory display implemented using spatial audio techniques on both levels in order to present information simultaneously. In this way, zooming between the sonification top (overview) level and the interactive secondary (detail) level would remain consistent.

[1]Auditory display elements appear to be fixed to the world so their positions have to be updated real-time according to the user orientation.
[2]Auditory display elements are always in a fixed position relative to the user.

Presenting information simultaneously in a multilevel and spatialised auditory display can help create a rich immersive audio environment, however 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. Vazquez-Alvarez and Brewster [2011] showed that spatialisation techniques are not as effective when users are under high cognitive load but they can offer an effective means of presenting and interacting with multiple audio streams simultaneously when cognitive load is kept low. Alternatively, presenting auditory streams sequentially will prevent the sources from competing with each other but this could result in a more lengthy interaction when switching between sources, poorer recall of earlier information, and irritation caused by continuous interruption. In this article we will investigate how these different presentation styles affect performance and user behaviour when using a consistent multilevel auditory display configuration and other mixed designs.

Spatial auditory displays for mobile audio-augmented environments can be designed to support navigational tasks but also more exploratory or wandering situations. A navigational system can be assessed by the user's success or failure at reaching a navigational goal, but this can also result in a system which prioritises efficiency over the exploratory and playful nature of the user experience [McCarthy and Wright 2004; Morrison et al. 2007]. On the other hand, evaluating *exploratory behaviour* in an audio-augmented environment presents challenges due to the implicit open-ended nature of exploration. In this article, an exploratory audio-augmented indoor environment is used to test a number of different design choices for implementing spatial auditory displays that support multiple auditory streams. It was felt that such a design made fewer prior assumptions concerning user behaviour, allowing users more freedom in their interaction with the auditory displays.

This study constitutes the first detailed examination of the potential of different spatial audio configurations using a multilevel auditory display design in order to understand how they affect the user experience in eyes-free and mobile audio-augmented interactive environments.

2. BACKGROUND

2.1. Situated interaction in indoor Mobile AAR systems

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 [Bederson 1995] was an early example of an exploratory non-linear playback system. This prototype system relied on a non-linear playback system and codes locally broadcasted by small infrared transmitters installed above every exhibit. The visitor had to carry a random access audio device, a modified Sony MiniDisc player, and a custom infrared receiver that would track the location of the visitor. As the visitor came close to an exhibit, the associated comment would automatically start and then stop if the visitor walked away. Similarly, the Audio Aura system [Mynatt et al. 1998] was designed to provide serendipitous information (i.e. information not actively asked for) in the periphery via audio cues based on the motion of the user in the workplace. The location of the user was tracked using an active badge system that triggered the audio delivery. The design of these audio cues combined speech, music and sound effects to provide peripheral information such as calendar reminders, email status and information of activities of other colleagues. This information was relevant not only in the general context of the receiver, but also semantically connected to the physical

space. Unfortunately, no formal evaluation of these early prototypes was carried out to determine their effectiveness.

Since Bederson and Mynatt et al.'s early prototypes, exploratory indoor AAR applications have grown in complexity and spatial audio techniques are often used to provide access to information about different locations in the environment. In order to spatialise virtual sound sources, the user's position and head orientation is tracked in real time and 3D positional audio algorithms such as stereo panning[3] or HRTFs (Head Related Transfer Functions) [Begault 1994], for more accurate rendering of spatial audio, are used to deliver spatialised audio to the user through headphones. Audvert [Betsworth et al. 2013] is a recent example of a mobile AAR system that facilitates serendipitous discovery (i.e. wandering) and navigation through spatial audio. This system uses spatial auditory feedback to engage the user and guide it to a particular place in the environment. However, such an implementation results in the user not being co-located with the location being augmented when the information is being provided. This user co-location is critical when investigating situated interactions.

In order to deliver information that is co-located in the physical environment, mobile AAR applications usually situate user interaction within a proximity and an activation zone [Stahl 2007]. 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 [Bederson 1995] to multiple [Wakkary and Hatala 2007; Eckel 2001] audio streams. As the amount of information presented increases, more complex auditory displays within the activation zones are required.

The CORONA system [Heller and Borchers 2011] used virtual characters to present information to visitors about a coronation feast from the 16th century at particular physical locations in the Coronation Hall in Aachen, Germany. The virtual audio space was rendered using stereo panning on an Apple iPhone, which was presented over a pair of headphones. The audio space included ten source areas where information was presented to the visitor. Each of these source areas was surrounded by a circular proximity and activation zone (the voice of the virtual characters increased in volume within the proximity zone as the distance to the activation zone decreased) which were triggered by the visitors as they explored the space. Other systems like the one developed by the LISTEN project [Eckel 2001] proposed a tailored audio-augmented user experience in a museum environment in which not only a particular physical location was being augmented with virtual audio information but also a physical item, e.g. pictures or statues. In this system, Goßmann and Specht [2002] used an activation zone connected to smaller parts of the physical object containing more detailed audio comments. This detailed information was only audible if the visitor was located at a specific angle and distance away from the physical object and the visitor was facing the object. Similarly, the ec(h)o system [Wakkary and Hatala 2007] also allowed users to interact with multiple information about exhibits at the Canadian Museum of Nature in Ottawa. Using a tangible object in the form of a cube when in front of a display of artifacts, the visitor was able to interact with paired short audio sequences in the form of audio prefaces that acted as multiple-choice options for the audio objects, which contained a greater depth of information. The spatial arrangement of the auditory display was mapped to the tangible interface for selection. The design of the auditory display itself was simple using the left channel audio for the left, right channel for the right and both channels for the centre, presenting the prefaces to the user from left to right in a sequential order.

[3]This is a technique that reliably positions sound to the left or right of a listener, while variations in intensity can indicate distance.

Unfortunately, the lack of systematic user evaluation and the wide range of different spatial audio techniques used in these previous AAR systems make it difficult to compare the efficiency and usability of the auditory displays used across the different applications. A more detailed and controlled investigation into the effects of mobile spatial auditory display design on situated interaction may shed light on their usability and their impact on the user experience. In this article, we use a controlled mobile AAR environment to test both quantitatively and qualitatively a number of different choices for designing spatial auditory displays that support multiple information.

2.2. Spatial auditory displays for eyes-free mobile interaction

Nomadic radio [Sawhney and Schmandt 2000] was an early attempt to break away from the traditional desktop computing paradigm. This was not just an application but an audio-only wearable device that supported interaction with personal messages in a mobile environment. The output of the Nomadic Radio was spatialised and reproduced via shoulder-mounted loudspeakers, whereas the input was entered using spoken commands via a speech recognition interface. Messages were presented in the spatial position corresponding to the time of arrival, i.e. 12:00 in front of the user's nose, 3:00 and 9:00 to the right and left of the user, 6:00 behind the user. The spatialised nature of this system enabled the presentation of multiple and simultaneous auditory streams that users could distinguish and separate from each other. Navigation allowed users to actively browse these messages via a synchronised combination of non-speech audio, synthetic speech and spatial audio techniques. Unfortunately, although an informal evaluation of this application was performed, no formal usability evaluation was ever carried out. Thus, the usability and impact on user interaction of these 3D auditory interface design is still unknown for an eyes-free mobile environment.

Other spatial auditory interface designs have used spatial separation to convey menu structure. Foogue [Dicke et al. 2010] is an eyes-free spatial auditory interface purposely designed for state-of-the-art smartphones. Foogue allows the user to navigate, select or manipulate spatialised audio items from a hierarchical menu. All items are arranged in a 120 degree arc in front of the user and displayed in sequence. Unfortunately, no evaluation of this system was carried out to assess its effectiveness. Diary in the Sky by Walker et al. [2001] used a 3D audio radial pie menu, with the user's head in the middle of the pie, to encode the times of diary appointments. Using a desktop simulation, the diary entries were consecutively presented for selection according to their time of appointment, as in the Nomadic Radio system described earlier. Although spatial audio significantly improved user performance in this system, its usability in a mobile environment is not known and in addition, it is unclear to what extent the presentation of audio information sequentially might have affected user interaction. Similarly, the earPod application [Zhao et al. 2007] was used to evaluate the usability of a spatialised radial menu in which audio items were displayed sequentially. The efficiency of this audio menu was compared to that of an equivalent visual menu display. User interaction was performed using a circular touchpad that reinforced the user's cognitive mapping between menu items and spatial locations on the touchpad. It was found that earPod was efficient to use, relatively easy to learn and comparable in both speed and accuracy with a visual menu selection technique. Unfortunately, only an informal evaluation of this system was carried out in a mobile environment and simultaneous presentation was not explored.

Brewster et al. [2003] conducted a study to compare sequential and simultaneous sound presentation in a mobile radial audio pie menu interface. Three conditions were tested in this study in which sounds were presented sequentially in an egocentric or exocentric display and simultaneously in an exocentric display. A head gesture was used

for selection. The results from this study showed that the egocentric display design was more effective and simultaneous presentation led to faster performance, however the exocentric designs evaluated in this study were only partially exocentric as they depended on head orientation but not on user position or user orientation. Marentakis and Brewster [2006] also investigated the usability of egocentric and exocentric auditory displays. Users were asked to select a target sound amongst a number of distracters using a physical pointing gesture while standing, with the help of a loudness cue, a timbre cue and an orientation update cue and combinations of these cues. The results showed that in the egocentric display participants were faster but less accurate, whereas in the exocentric display they were slower but more accurate. However, the exocentric display design used in this study was again, as in [Brewster et al. 2003], only partially exocentric. Furthermore, the sounds in these exocentric displays did not relate to any targets physically located in the space.

In summary, previous work has used spatial audio to present multiple information sequentially in an egocentric [Dicke et al. 2010] or exocentric [Terrenghi and Zimmermann 2004] display; or simultaneously in an egocentric [Sawhney and Schmandt 2000] or exocentric display [Brewster et al. 2003]. However, the usability of these designs has not been compared against each other as part of an interactive audio-only environment. An egocentric display can be particularly useful for mobile users as changes in orientation when moving are inevitable. On the other hand, an exocentric display can aid user navigation as display elements appear to be fixed to the world 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 [Marentakis and Brewster 2006] and that simultaneous presentation allows for faster user interactions [Brewster et al. 2003]. However, to our knowledge, no previous research has compared the use of combined ego- and exocentric designs within the same spatial hierarchical auditory display. In this article, we build on this previous work by presenting a systematic evaluation of a complex multilevel spatial auditory display including egocentric and exocentric designs and sequential and simultaneous presentation, tested in a mobile AAR environment.

3. EXPERIMENTAL STUDY

In this section, we present a detailed description of the experimental setup and motivation for the study design.

3.1. Audio-augmented Art Exhibition

A conceptual art exhibition was used as the setting for this study. A variety of different mobile auditory interfaces designed to provide access to multiple location-based information were implemented and tested in this exhibition space, always aiming at a full eyes-free interaction between the user and the mobile device running the auditory interfaces so the user's visual attention could be focused on the interaction with the object being audio augmented.

The virtual audio environment superimposed on the art exhibition was run on a Nokia N95 8GB and the built-in Head Related Transfer Functions (HRTFs) and the JAVA JSR-234 Advanced Multimedia Supplements API (AMMS) [2007] were used to position the auditory sources. User position was determined using an Infrared (IR) camera tracking an IR tag powered by a 9V battery (see Figure 1a) and mounted on top of a pair of headphones. Coordinate information was fed to the mobile phone over a WiFi connection and was used to activate the zones associated with the art pieces in the exhibition space. User orientation (compass heading) was determined using a head-mounted JAKE Sensor Pack [2010] (see Figure 1b) connected to the mobile phone via Bluetooth. No visual aids were provided on the screen of the mobile device and,

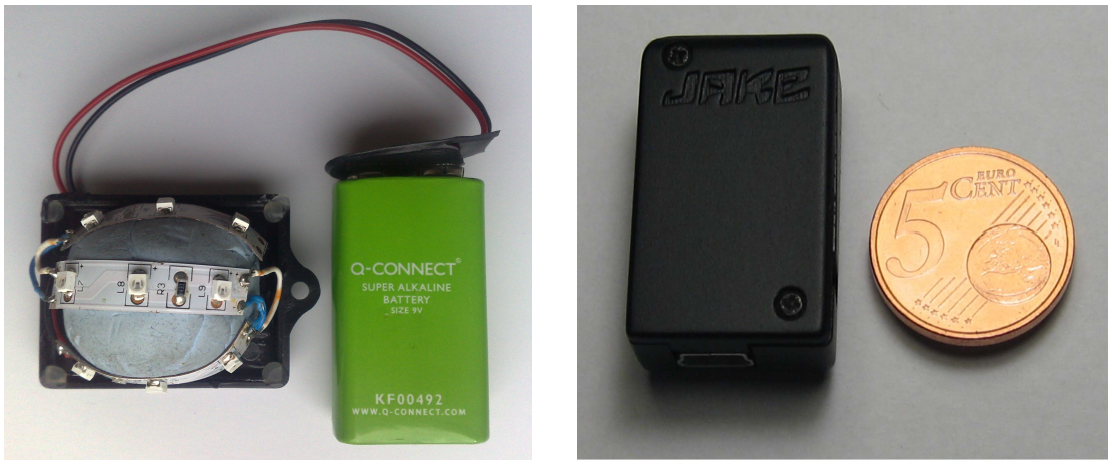

Fig. 1. (a) IR tag with 9V battery attached, (b) JAKE sensor pack shown with a five cent euro piece. Image provided courtesy of Stephen Hughes.

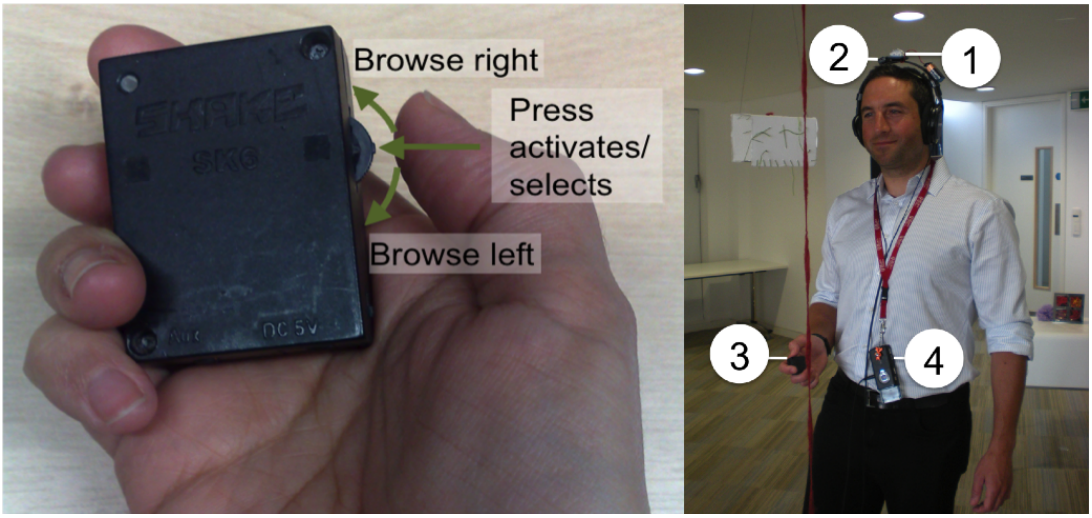

Fig. 2. Experimental setup (right): 1) IR tag and 2) JAKE sensor (as shown in Figure 1, both mounted on headphones), 3) SHAKE SK6 sensor pack and 4) mobile device; and interaction technique using the navigation switch on the SHAKE sensor pack (left).

to ensure a full eyes-free experience, the phone was placed on a lanyard around the user's neck (see Figure 2 (right)). The navigation switch on a SHAKE SK6 sensor pack [2010], also connected via Bluetooth, was used to feed user input into the system while users were holding it in their hands. This navigation switch allowed users to activate and deactivate audio content by pressing the switch and also to browse the content by pushing the switch left or right (see Figure 2 (left)). The audio was played over a pair of DT431 Beyerdynamic open-back 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 and both mounted using Velcro tape. Figure 2 (right) shows the final system setup and Figure 3 shows the system architecture.

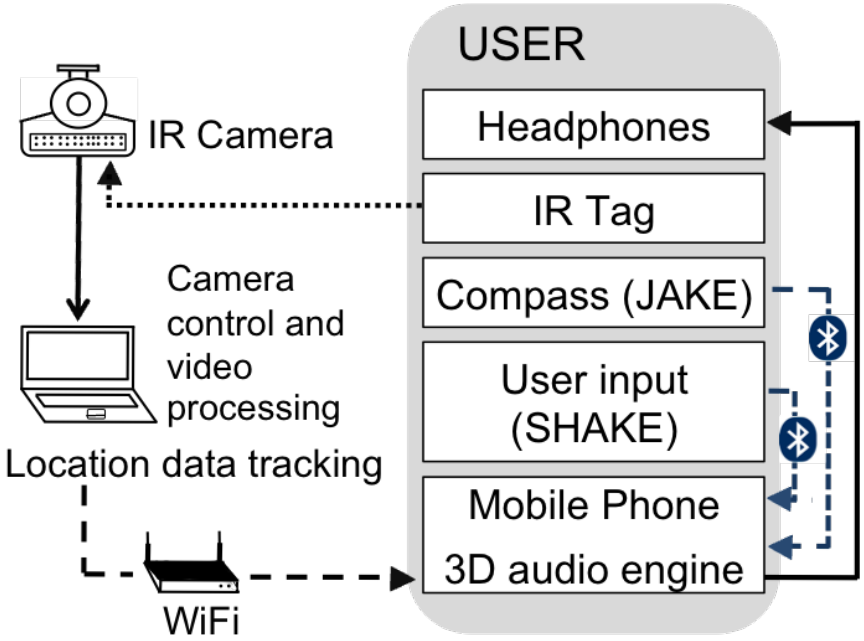

Fig. 3. Schematic representation of the system architecture.

3.1.1. Conceptual Art Exhibition Space. A Conceptual Art exhibition was chosen because as LeWitt writes "the artist has no control over the way a viewer will perceive the work. Different people will understand the same thing in a different way." [LeWitt 1967]. This offers an interesting use case where a single written description would be inappropriate whereas a multiple set of comments from different visitors and the artist could help inform and enrich a visitor's experience, requiring the use of multiple items of information for different exhibits and motivating the multi-level displays we wished to explore.

The exhibition consisted of six different art pieces from the *Weaving the City* project (www.weavingthecity.eu) kindly donated by Rocio von Jungenfeld, from the Edinburgh College of Art. Four art pieces were made of woolen threads and paper and exhibited in a space that measured 3m wide(x) x 3.85m long(y). They were complemented by another two media pieces placed outside the exhibition space. One media piece captured the participants' image via a webcam as they walked past and, after being processed using a Max/MSP patch running on a Mac mini, projected on the wall. A second media piece was a movie about the Weaving the City project playing in a loop on an iBook G4. The media pieces were not audio-augmented, i.e. no audio information was offered about these pieces, and their purpose was to make the exhibition space more playful and immersive with the help of the projected images and sounds. Two of the art pieces were suspended from the ceiling hanging at eye level and the rest, including the media pieces, were placed on small tables (see Figure 4 for an illustration of the setup).

3.1.2. User Location Tracking. In this study, the indoor tracking system consisted of a PlayStation® Eye camera modified to work as an IR camera and an IR tag used to track the user location in the exhibition space. The IR camera was attached to the

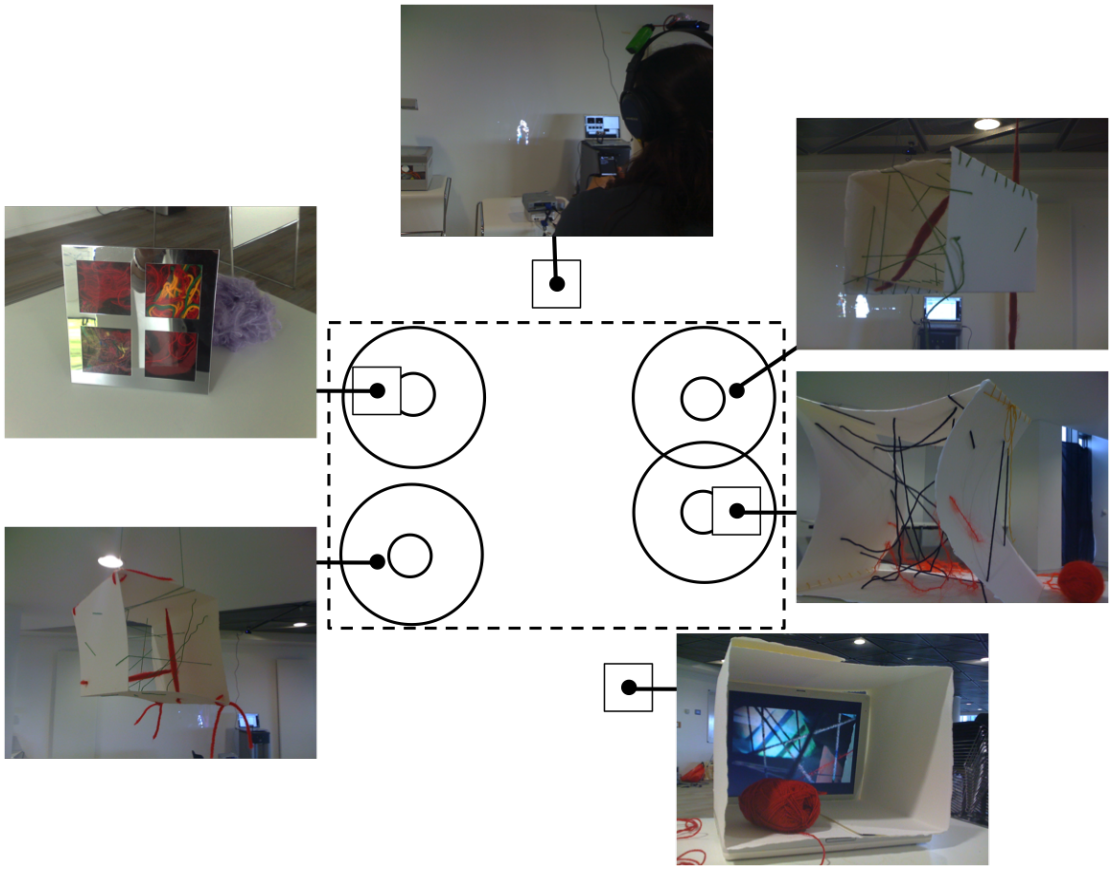

Fig. 4. Illustration of the exhibition area layout and the top-level sonification layer showing the location of the proximity and activation zones surrounding each art piece. The dashed-line area identifies the audio-augmented exhibition space measuring 3m(x) x 3.85m(y). The small squares with a dot at its centre identify the art pieces placed on tables and a dot alone the ones that hung from the ceiling.

ceiling using velcro and connected via USB to a MacBook computer running Community Core Vision - CCV (http://ccv.nuigroup.com) and Processing (http://processing.org) open source software. CCV takes a video input stream and outputs tracking data as coordinates and it is frequently used in building multi-touch applications [Correia et al. 2010; Fu et al. 2010; Roth et al. 2010; Leftheriotis and Chorianopoulos 2011; Zhang et al. 2012]. In this study CCV tracked the position of the IR tag mounted on the participant's headphones. Then, the Processing application used the TUIO (Table-Top User Interfaces Objects) API to decode the TUIO messages sent out from CCV and output coordinate information at 2Hz to a multicast network socket. The coordinate information could then be accessed by the mobile phone running the audio-augmented environment. The Processing application also plotted user location in real-time on the computer screen so the experimenter could confirm the tracking was active.

This indoor tracking system was first calibrated before measuring the tracking error (i.e., the average Euclidean distance deviation in centimetres (cm) between the actual position of a number of target points and the estimated user position as detected by the

IR camera at the same target points). Three volunteers of different heights (1.58cm, 1.73cm and 1.92cm) took part in the initial IR camera calibration. The IR camera was first calibrated with the middle height (1.73cm) using a total of nine reference points across the 3m x 3.85m exhibition space and these camera settings were used for all three participants. In this way, the tracking error for shorter or taller heights could be calculated. Then, a total of 19 fixed target points were identified across the tracked exhibition space. An application was devised to present the fixed target points one at a time on the screen of the mobile device so the participant could then walk to the location indicated. Once the location was reached, pressing a button on the device logged the user location. Results showed that the tracking error was 55.69cm, 28.45cm and 76.33cm for the 1.58cm, 1.73cm and 1.92cm high participants respectively. Given that the tracking error varied considerably depending on height and it was lower for the middle height used to calibrate the IR camera in the first place, it was decided that the IR camera would be calibrated individually for each of the participants taking part in the evaluation study.

The main constraint on processing and latency was the ability of the Nokia N95 8GB to update the sound source locations during the exocentric interaction and deal with the multi-threaded application required for streaming audio, updating head position from the external jake device, and updating user position from the network socket opened to the location tracker computer. Mariette [2010], showed a decrease in user performance occurring between latencies of 400ms - 800ms. An update frequency of 500 ms was adopted which balanced the stability of the application running on the Nokia mobile phone while producing audio location updates at a speed which worked well with subjects walking and exploring the relatively small exhibition space. Latency for both location tracking and head tracking was negligible in terms of this update rate ($<$ 100 ms). This was reflected on the participants' high ratings of the audio experience (over 4.5 on a 5 point likert scale - see Figure 8). As Mariette [2010] showed, latency, accuracy and other detailed implementation aspects can have an important effect on user experience and should be carefully assessed against the activities they need to support. For instance, a competitive game encouraging rapid movement would require a different support to the casual exploration featured in our gallery setup.

3.1.3. Multilevel Auditory Display Design and Stimuli. There were two levels in the multilevel auditory display: 1) A top-level sonification layer and 2) a secondary interactive layer.

The design of the top level display was chosen based on results from previous work [Vazquez-Alvarez et al. 2012] with the Madeira sound garden. In this study, spatial auditory interfaces of gradually increasing complexity were explored. These varied in terms of the use of Earcons, proximity zones, and spatial audio cues. The fully spatialised Earcon-based display was preferred. Furthermore, an exocentric configuration using spatial audio is a design that has been used in previous studies (e.g. [Eckel 2001; Heller and Borchers 2011; Stahl 2007]). Thus, of the very large possible configurations for a top level auditory interface, this was therefore considered the most most appropriate for the top layer auditory display.

Non-verbal audio messages in the form of Earcons [Blattner et al. 1989] were used to advertise the content of each exhibit. Earcons provide an abstract and symbolic relationship between the sounds and the information they are representing. Vazquez-Alvarez et al. [2012] showed that choosing Earcons that fit the environment and have an ambient quality can contribute to a sense of immersion. For this study the following Earcons were chosen:

— Top-level sonification layer: A chattering voices Earcon was used to advertise content about each art piece. This Earcon was chosen because it fitted the gallery environment by representing an item of public interest which would encourage discussion.
— Secondary interactive layer: For each art piece, different Earcons were used in an audio menu to identify comments left by the artist and those left by non-expert reviewers. In order to provide a uniform listening environment in this layer, Earcons representing the elements (i.e., water, fire and wind) were chosen to identify the different audio menu items as follows: "water waves": for the artist's comments to represent a deeper understanding of the work, "open crackling fire": representing warmth and excitement for positive non-expert reviews, and "stormy wind": for negative (cold) non-expert reviews. When a menu item was selected, an approx. 25 secs long audio clip was played containing the comment or review.

The top-level sonification layer attracted visitors towards the artwork and advertised the existence of information at that location. A circular proximity zone (radius 1.25m) advertised content and a smaller activation zone (radius 0.75m) enabled user access to it. The chattering voices Earcon used in this layer was mono, 16-bit and sampled at 16kHz. It was presented within the proximity zone surrounding each art piece using an exocentric design (sound positions were updated in real-time according to the user orientation and the loudness of the sound increased as the distance to the art piece decreased) to provide the user with orientation and distance information, while the activation zone was user-activated. The proximity zones overlapped for two of the art pieces while the other two were isolated (see Figure 4).

The secondary interactive layer was only accessible when in the activation zone of the top-level sonification layer. It consisted of a variable number of audio menu items from a minimum of one to a maximum of three. User interaction with the audio menu items varied for the different experimental conditions, as will be described in the next section. Both the Earcons and their related information were mono, 16-bit and sampled at 16kHz.

The chattering voices Earcon in the top-level sonification layer and the audio menu items in the secondary interactive layer were adjusted to conversational volume (approx. 60-70dB).

3.2. Study Design

3.2.1. Participants. 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 were paid £6 for participation, which lasted just over an hour. 12.5% (n=4) of the participants reported that they rarely went to museums or art galleries, 12.5% (n=4) reported they went once a year at most, 53.1% (n=17) two to three times a year, 18.8% (n=6) no more than once a month and 3.1% (n=1) at least once a week. Only 15.6% (n=5) of the participants had never used an interactive museum system in the past.

Participants were split equally into two groups: sequential and simultaneous presentation, in a between-subjects design. In the sequential audio group the Earcons in the interactive auditory display were presented sequentially one at a time, whereas in the simultaneous presentation group Earcons were presented simultaneously all at the same time.

3.2.2. Conditions. Each group (*sequential* and *simultaneous presentation* group) was tested in a Baseline condition and two other conditions, in which the secondary interactive layer varied in complexity, resulting in a total of 5 different conditions:

(1) *Baseline*: Each Earcon was *always* played sequentially at each push of the navigation switch either right or left for both presentation groups. There was no spatialisation of the audio menu 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 5a and 6a.
(2) *Sequential* egocentric: 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 0°, -90° and +90° azimuth (see Figure 5b).
(3) *Sequential* exocentric: In the exocentric conditions, each Earcon was situated in the exhibition space exocentrically in front of the art piece oriented towards the centre of the exhibition space and at a minimum 45° separation of each other (see Figure 7). The menu items were perceived as if they were fixed to a location. Selection for the sequential presentation group was performed pushing the navigation switch either right or left (see Figure 5c).
(4) *Simultaneous* egocentric: Similar to the sequential egocentric condition except that all of the Earcons were played simultaneously. When a menu item was selected, the volume increased for the selected item to bring it into focus and decreased for the rest. Again, selection was performed pushing the navigation switch either right or left (see Figure 6b).
(5) *Simultaneous* exocentric: As in the sequential exocentric condition, Earcons were perceived as if they were fixed to a location. However, selection was performed by walking around an art piece and then standing at the location where a menu item was situated (see Figure 6c). A loudness cue identified the activation area where menu items could be selected. The proximity zone around each menu item was 3m to ensure all items would overlap and play simultaneously and the activation zone was 1m. Here, a consistent design across an exocentric top-level sonification layer and an exocentric secondary interactive layer was tested.

The Baseline condition was used as a control in both groups and the order of conditions was randomised per participant to control for ordering effects (see Table I for a summary of experimental conditions). Participants were tested in the mobile environment provided by the conceptual art exhibition space.

Table I. Summary of Experimental conditions. Condition order was randomised for each subject per presentation group to control for ordering/learning effects.

Presentation group	*Condition*		
Sequential	Baseline	Egocentric	Exocentric
Simultaneous	Baseline	Egocentric	Exocentric

3.2.3. Layout. The number of exhibits we could use was limited by the physical space available, the area over which we could carry out effective indoor tracking, and the art pieces available to us. Six pieces were chosen to form the exhibition of which four were audio augmented. We chose a conservative maximum number of audio items per exhibit (three audio items at azimuths -90, 0, +90 degrees) as in Wakkary and Hatala [2007]. To control for item density, the number of items was varied from 1 to 3 and the

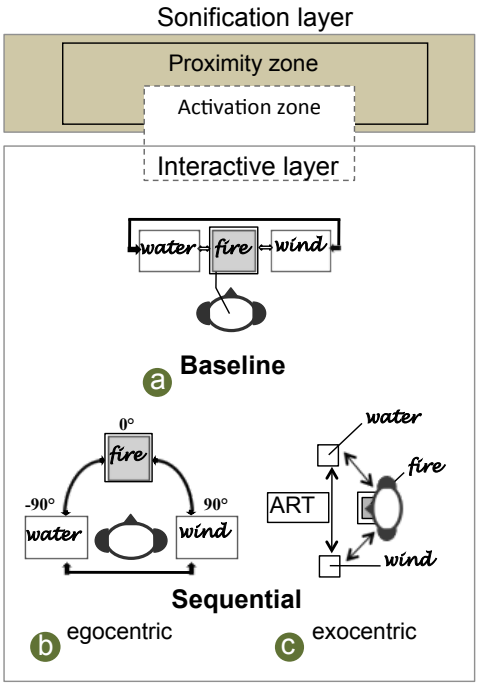

Fig. 5. Schematic illustration of the multilevel auditory displays surrounding each art piece for the sequential presentation group. Each multilevel auditory display consisted of a top-level sonification layer and a secondary interactive layer. In the sonification layer, there was a proximity zone and an activation zone. The interactive layer could only be activated when the user was situated in the activation zone. The interface design tested in the interactive layer varied in complexity for the different experimental conditions (The greyed-out areas indicate the number of sound sources playing at one time): **a)**Baseline: non-spatialised Earcons were played sequentially by pushing the navigation switch right or left. **b)** *Sequential* egocentric: each Earcon was played sequentially from a location around the user's head at each navigation button push. **c)** *Sequential* exocentric: Earcons were situated in the exhibition space and perceived as if they were fixed to a location in the physical space. Selection of an audio menu item was performed by pushing the navigation switch independently of were the user was situated around the art piece.

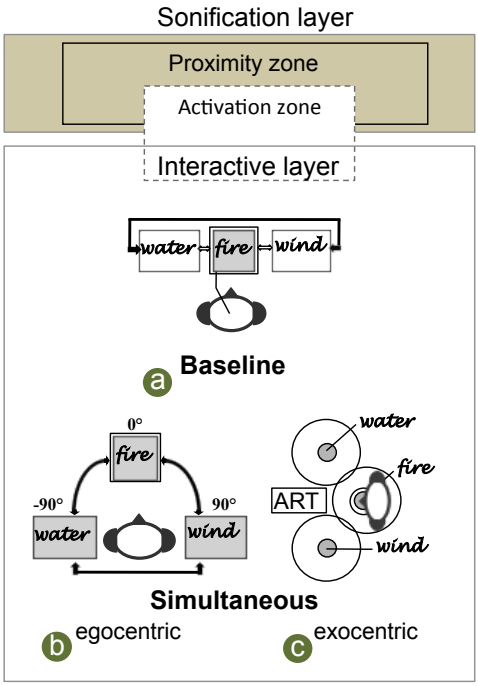

Fig. 6. Schematic illustration of the multilevel auditory displays surrounding each art piece for the simultaneous presentation group. Each multilevel auditory display was designed as described in Figure 5. The interface design tested in the interactive layer for the simultaneous presentation group also varied in complexity for the different experimental conditions (The greyed-out areas indicate the number of sound sources playing at one time): **a)**Baseline: non-spatialised Earcons were played sequentially by pushing the navigation switch right or left. **b)** *Simultaneous* egocentric: all Earcons were played simultaneously. Selecting one Earcon using the navigation switch would increase the volume of the selected Earcon and decrease the volume of the non-selected ones. **c)** *Simultaneous* exocentric: all Earcons were played simultaneously from a fixed location around the art piece. Circular proximity zones around each art piece enabled Earcons to play simultaneously. To select an Earcon, users walked around the art piece till the Earcon was perceived as louder than the rest. This indicated the Earcon was selected and the comment or review could be played.

overlap of the proximity zones between exhibits was also controlled with two overlapping exhibits and two isolated exhibits. The overall exhibition was larger than the area covered by indoor tracking and extended beyond the audio augmented exhibits allowing participants to walk out of the tracking area and return as desired. The objective was to design a playful and pleasant exhibition space to explore at leisure.

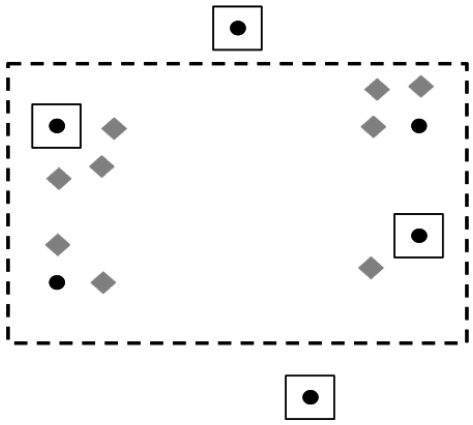

Fig. 7. Illustration of the exhibition space identifying the location of the audio menu items in the secondary interactive layer for each of the art pieces in the Exocentric condition. The dashed-line area identifies the audio-augmented exhibition space measuring 3m(x) x 3.85m(y). The small squares with a dot at its centre identify the art pieces placed on tables and a dot alone the ones that hung from the ceiling. Grey diamonds identify the location of the audio menu items for each of the art pieces situated in the exhibition space. This schematic illustration of the secondary interactive layer for the Exocentric condition complements Figure 4, which identifies the top-level sonification layer configuration used in all conditions.

3.2.4. Procedure. The experiment included a calibration procedure and a training session before the test conditions. First, the indoor tracking system was calibrated for the height of each participant. This followed a training session using the starting test condition to familiarise the participant with the multilevel 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. The experimental instructions and brief introduction to the exhibition can be found in Appendix A.1 and Appendix A.2 respectively. Also, the auditory display description provided per test condition can be found in Appendix A.3. As participants walked closer to the audio-augmented art pieces, the proximity zone in the top-level sonification layer was triggered and the sound of chattering voices was played to indicate the presence of information at that location. As participants approached the art piece more closely, they were able to reach the activation zone in which they were able to activate the secondary interactive layer. Participants knew they had reached the interactive area when the chattering voices were louder and heard in both ears. To activate the interactive layer, the participant pressed down (long press $>$ 2 secs) the navigation switch on the SHAKE sensor pack. Once activated, the audio menu could be browsed by pushing the navigation switch to the right or left. To select one of the menu items the navigation switch was pressed down (short press $<$ 2 secs). Once a menu item was selected, its content was made available to the user. When the participant finished listening to the information available in the audio menu, pressing down (long press) the navigation switch exited the interactive layer and the sound of the chattering voices in the top-level sonification auditory display was played again. The participant could then walk away and continue to explore the space. In order to keep each experimental session within one hour duration, 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 and a satisfaction questionnaire (see Appendix A.4) and also provide some informal feedback on their experience interacting with the system being tested in that condition. Once all three test conditions were completed, participants were instructed to provide feedback on how the different auditory interfaces tested in the exhibition space compared to each other. Finally, participants were invited to add an entry to a visitor's book especially created for the exhibition in which they could write any thoughts the information contained in the art pieces had provoked in them, if any.

3.2.5. Experiment hypotheses and metrics. In this study, two hypotheses were formulated:

i A consistent design (the same *exocentric* auditory display design in both the top-level sonification layer and the secondary interactive layer) in the multilevel audio display would follow the design principle of consistency and reduce subjective workload and increase user satisfaction.
ii The use of spatial audio techniques in the secondary interactive layer of the multilevel auditory display will encourage an exploratory behaviour, which will result in significantly more time taken interacting with the system without a significant drop in user satisfaction or a significant increase in perceived workload.

In this evaluation, user satisfaction and workload metrics (user experience) together with performance indicators were combined to assess the effectiveness of the interactive displays. The independent variable (IV) was the type of condition (the *Baseline* condition, the *Egocentric* condition and the *Exocentric* condition) per sequential and simultaneous presentation group, and the dependent variables (DVs) were a combination of subjective (level of user satisfaction and perceived subjective workload) and objective measures (time taken while interacting with the secondary interactive layer). In addition to participants' comments and opinions, user location coordinates and head orientation data were also collected for an in-depth analysis of participant behaviour.

The satisfaction questionnaire used in this experiment was a modified version of the one used in Wakkary and Hatala [2007] to evaluate the overall reaction to the system, the user interface, learning how to use the system, perceptions of the system's performance, the experience of the content, and degree of navigation and control. The questionnaire used in this study was modified to reflect the differences in the design of the system, in particular the user interface (questions on the SHAKE sensor pack and open-back headphones instead of the original interaction cube and wireless headset) and the content management (questions on the audio menus instead of the original audio preface). In addition, a question on the level of immersion experienced by the user was added to the set of questions grouped under the 'overall reaction to the system' category (see Appendix A.4). The inclusion of this question was motivated by the user feedback reported in Vazquez-Alvarez et al. [2012], in which participants remarked on the level of immersion experienced when interacting with a fully spatialised system.

4. RESULTS

4.1. User Experience

User Satisfaction. The user satisfaction questionnaire taken from Wakkary and Hatala [2007] grouped questions into 8 categories (see Appendix A.4). Each question was rated on a continuum from "low" (1) to "high" (5) satisfaction. The eight question categories were: *overall reaction to the system*, *user input interface*, *comfort level of headphones or headset*, *learning how to use the system*, *perceptions of the system's per-*

formance, *quality of the content*, *audio experience*, and *degree of navigation and control*. Satisfaction mean scores were calculated from the participants' responses to the multiple questions contained in each category. These mean scores could then be analysed using parametric statistics [Boone Jr and Boone 2012]. Overall, and across all categories, the eyes-free audio-augmented experience tested in the exhibition space was judged by participants as very satisfactory (mean=3.97). The categories that stood out as most satisfactory were the interaction with the user input interface device, learning how to use the system and the comfort experienced while wearing the headphones required to interact with the audio interface. Multiple two-way mixed-design ANOVAs were carried out on the 8 different categories in the satisfaction questionnaire. In order to protect from a Type I error, results from the ANOVAs were adjusted with a Bonferroni correction for multiple tests. Only results for *overall reaction to the system* showed a significant interaction between condition type and presentation group ($F(2,60)=9.134$, $p<0.01$ with Bonferonni correction). No significant main effect was found for presentation group or condition type. *Post hoc* paired samples t-tests with Bonferroni correction for condition type showed that the satisfactory reaction was significantly higher for the Baseline condition (mean=4.10, SD=.49) ($t(15)=3.014$, $p<0.030$) and Egocentric condition (mean=3.99, SD=.56) ($t(15)=4.011$, $p<0.005$) than for the Exocentric condition (mean=3.56, SD=.73) in the simultaneous presentation group. No significant differences were found between the conditions for the sequential presentation group. See Figure 8 for more detailed results per category.

This result shows that users were less satisfied with the simultaneous exocentric design than the other designs (see Figure 9). Thus, the first hypothesis that a consistent design across layers in a multilevel auditory display would increase user satisfaction was rejected, according to these results. However, the second hypothesis that using spatial audio techniques in the secondary interactive layer would encourage an exploratory behaviour (i.e. significantly more time taken interacting with the system without a significant drop in user satisfaction or a significant increase in perceived workload) was partially confirmed, as the other spatial conditions did not show a significant drop in user satisfaction.

Overall Workload. Raw overall workload means were calculated from the NASA-TLX questionnaire completed after each condition (see Figure 10). A two-way mixed-design ANOVA on overall workload with condition type as a within-subjects factor and presentation group as a between-subjects factor showed a significant main effect for condition type ($F(2,60)=4.606$, $p<0.015$). There was also an interaction between condition type and presentation group ($F(2,60)= 4.672$, $p<0.015$). No significant difference was found between presentation groups. *Post hoc* Paired samples t-tests with Bonferroni correction per presentation group showed that overall workload was significantly higher for the Exocentric condition ($t(15)=-3.480$, $p<0.01$) (mean= 45.00, SD=24.55) than the baseline (mean= 27.00, SD=15.62) and the Egocentric condition ($t(15)=-3.406$, $p<0.015$) (mean=31.63, SD=19.05) for the simultaneous presentation group. No significant differences were found between the conditions for the sequential presentation group.

These results show that workload was higher for the simultaneous exocentric design than for the other designs and supports the rejection of the first hypothesis that a consistent design across layers in a multilevel auditory display would reduce subjective workload. However, the second hypothesis was partially confirmed, as the other spatial conditions did not show a significant increase in perceived workload.

4.2. User Performance

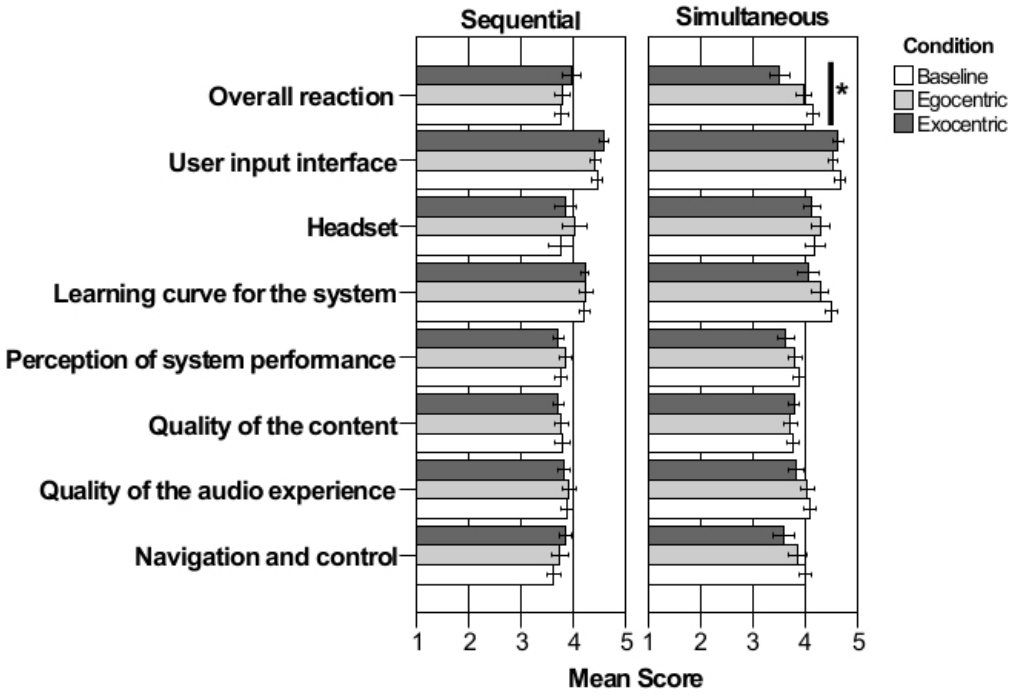

Fig. 8. Mean scores calculated from summated responses to the 62 questions on the satisfaction questionnaire (see Appendix A.4 for an example of the questionnaire, a modified version of the one used in Wakkary and Hatala [2007]) grouped into eight categories (shown on the left of the figure); per condition type: Baseline, Egocentric, and Exocentric; and presentation group: Sequential and Simultaneous. Answers were collected using a 5-point Likert scale from negative to positive, where 5 was 'best'. Means are well above 3 (all positive) for all eight categories with only the Overall reaction category showing significant differences between conditions for the Simultaneous presentation group,(*) Indicates a significant difference with corrected $p<0.008$ after Bonferroni adjustment (x8 factor given the eight categories in the satisfaction questionnaire). Error bars show Standard Error of Mean $\pm$ 1.0.

Time taken: interactive layer. The total time participants spent interacting with the secondary interactive layer was also computed. See Figure 11. A two-way mixed-design ANOVA on time taken with condition type as a within-subjects factor and presentation group as a between-subjects factor showed a significant main effect for condition type ($F(2,60)=5.971$, $p<0.005$). No significant main effect was found for presentation group or interactions with condition type. *Post hoc* pairwise comparisons with Bonferroni correction for condition type showed participants spent significantly less time interacting with the auditory display in the Baseline conditions (mean= 249 secs, SD= 50) when compared to the spatial conditions (Egocentric: mean= 307 secs, SD= 109, $p<0.025$; Exocentric: mean= 322 secs, SD= 102, $p<0.005$).

These results show that the multilevel auditory displays designed with a spatialised secondary interactive layer encouraged users to spend longer interacting with the artwork. This finding together with previous results from workload and user satisfaction not showing a significant increase in perceived workload or a significant drop in user satisfaction, confirm the second hypothesis formulated in this study (i.e. that spatial audio techniques would encourage an exploratory behaviour) for all spatial conditions except for the simultaneous exocentric.

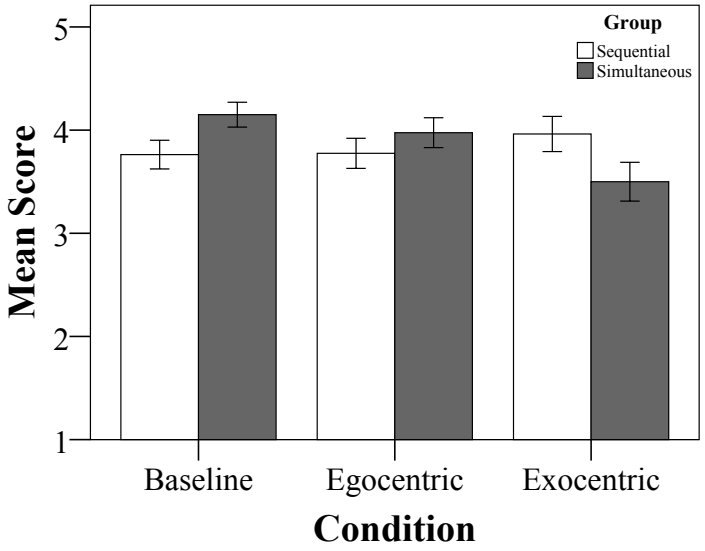

Fig. 9. Detail from figure 8 showing mean scores for the 'overall reaction to the system' category per condition and presentation group: Sequential (white) and Simultaneous (dark grey), see section 3.2.2 for more detail on the experimental conditions. Error bars show Standard Error of Mean $\pm$ 1.0.

4.3. User Feedback

Based on the user feedback collected after each condition was completed, twenty-nine participants out of the thirty-two that took part in this study found the experience enjoyable or interesting and they agreed that the provision of audio comments about the art pieces enhanced their experience. In addition, three participants reported that this experience had made them more likely to use an audio guide next time they visited a museum/gallery. In conditions where spatialised audio was used to present information sequentially, three participants described the interfaces as "thought provoking". The occasional spatial audio latency problem affected the user experience in the spatial conditions but overall all participants enjoyed the idea of being able to walk around the space freely. Informal user feedback is presented for each of the three multilevel auditory display conditions.

Baseline. In general, the interaction with this auditory display was described as "easy", "enjoyable" and "most of all playful and entertaining", with "informative" audio content. Two participants felt "more in control" when using this auditory display as the audio content was triggered by a simple press of a button in a sequential order. However, other participants reported that, although this display was "faster to use", it was simply "less immersive" and "less fun" than the spatialised ones. One participant remarked: "I felt the experience was slightly less immersive and interacting with the menus less enjoyable. I felt a little less certain that I had heard all the comments when they were not categorised hierarchically. However, I would say I put less mental effort into using the system, but it was closer to a simple accompaniment tape, which was much less interesting".

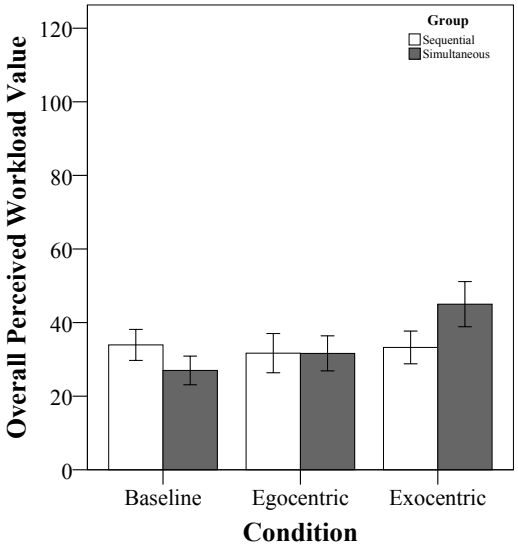

Fig. 10. Overall perceived workload per condition and presentation group: Sequential (white) and Simultaneous (dark grey), see section 3.2.2 for more detail on the experimental conditions. Error bars show Standard Error of Mean ± 1.0.

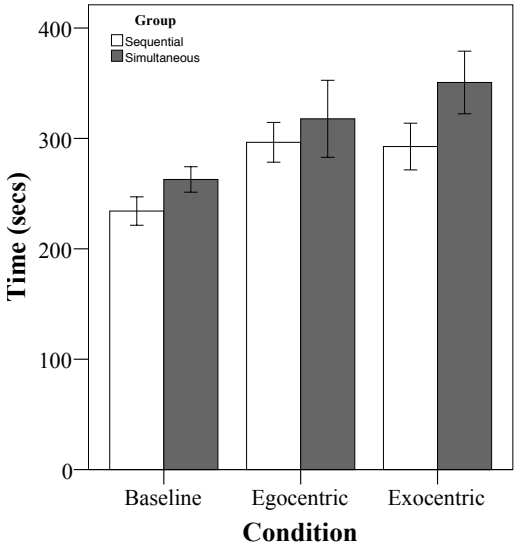

Fig. 11. Mean time taken interacting with the secondary interactive layer per condition and presentation group: Sequential (white) and Simultaneous (dark grey), see section 3.2.2 for more detail on the experimental conditions. Error bars show Standard Error of Mean ± 1.0.

Sequential egocentric. The participants' experience of the sequential egocentric condition, in which the Earcons were spatialised around the user's head and played one at a time when selected, was described as "novel", "fun", "enjoyable", "easy" and "informative". Although one participant reported that the spatialisation of Earcons did not affect the experience when using this auditory display, overall, the use of spatialisation had a positive impact in the user experience and interface usability. One participant felt that having the Earcons separated spatially "made them easier to differentiate" and another suggested that "it gave a real life feeling, as if someone was indeed talking to me".

Sequential exocentric. Participants in the sequential exocentric condition, in which Earcons were perceived as if they were fixed to a location in the exhibition space and selected sequentially by the press of a button, described their experience of the auditory display as "enjoyable", "stimulating", "fun" and "informative" with one participant finding the concept "innovative". Having the auditory sources fixed to a physical location was reported to have "made the experience even more immersive" as "having the audio cues distributed around the artifact encouraged you to examine it from all sides". This interface was found to be "easy to use" and "quick to learn". One participant highlighted how "it was unique in the way the menu was in your head and you had to use your hearing to navigate through the menu".

Simultaneous egocentric. Having all the Earcons play simultaneously in the simultaneous egocentric condition was found to "enhance the experience" and to be "less mechanic and artificial". One participant reported: "I liked the fact that the sounds played simultaneously, that I had much more control over which of them I played. It was also much easier to confirm that I had listened to all that were available, and

it was nice to be able to control the movement of the sound around your head." On the other hand, three participants remarked that playing the Earcons simultaneously made it a little bit more difficult to remember which option they had already listened to and consequently made the interaction more confusing. Two participants suggested that more training could offer a solution to enjoying the system more.

Simultaneous exocentric. Earcons in the simultaneous exocentric condition were not only presented simultaneously but participants were also required to walk around the art pieces in order to locate and select them. Some participants felt distracted by the need to move around more to locate the Earcons and perhaps for that reason they felt unsure of whether they had found all the menu items around the art piece. One participant remarked: "it required a lot more physical activity going back and forth, focusing more on the commentary and it made me focus less on the artwork". Although there were participants that enjoyed having to move more and felt that it added to the playfulness, entertainment and "our awareness of space", the small size of the exhibition that resulted in Earcons being closer together had a negative effect on the user experience. One participant reported: "Having the audio cues distributed around the artifact encouraged you to examine it from all sides, but finding the correct space to play certain cues was occasionally tricky". However, two participants remarked on the potential of this interface. One participant suggested that "if art pieces were much larger walking around to get menu items would make more sense" and another reckoned that "making you *search* for the audio information instead of it being available straight to you, I think this would be applied more to the public exhibitions in numerous very new experiences [*sic*]".

4.4. User Behaviour: Qualitative Analysis

A detailed analysis of the logged data (including user location coordinates and head orientation data) revealed three key features concerning the exploration behaviour of the participants.

Spatialised audio feedback lead to a deviation from systematic exploratory behaviour. Participants in the Baseline condition showed a tendency to cover less of the exhibition space, backtrack more infrequently and turn their heads less when compared to the Egocentric and Exocentric conditions. Compare Figure 12a - simultaneous baseline condition with Figure 12b,c,d - simultaneous egocentric, simultaneous exocentric, sequential exocentric conditions. The solid lines illustrate the direction of travel and the short splines illustrate the participant's head orientation logged at 2Hz.

The simultaneous exocentric secondary display lead to too much physical displacement. Participants in the simultaneous exocentric condition showed a tendency to cover much more of the exhibition space, backtrack often, and turn their heads more frequently when compared to participants in the sequential exocentric condition. The requirement of moving to a menu item location in order to activate it in the simultaneous exocentric condition succeeded in increasing physical displacement but at the expense of increasing workload and reducing satisfaction (compare Figure 12c - simultaneous exocentric condition with Figure 12d - sequential exocentric condition).

The exocentric secondary displays encouraged a 'head scanning' behaviour. In the Exocentric displays, in order to locate a menu item, participants would stop and turn their head, scanning for the location of the Earcon they were interested in selecting. (See the two gray circles highlighted on Figure 12c,d for examples of this behavior). Such head scanning behavior may be connected to a sense immersion but further work is required to investigate this. This behaviour was, however, not consistently observed in the Egocentric conditions (compare Figure 12c,d - simultaneous and sequential exocentric conditions with Figure 12b - sequential egocentric condition).

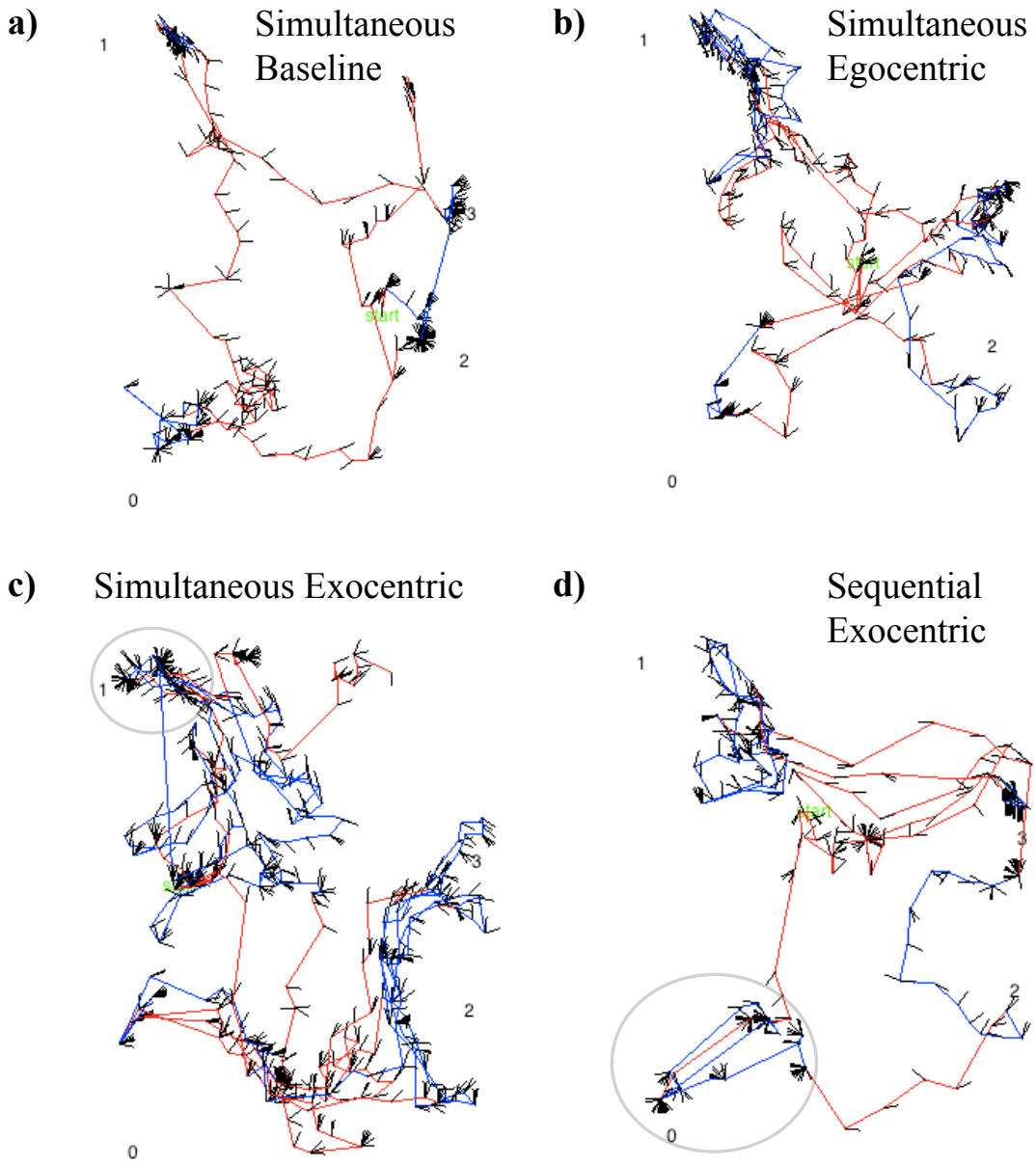

Fig. 12. Route taken around the art pieces (0-3) by two different participants: one from the *simultaneous* presentation group (**a**,**b** and **c**) and another from the *sequential* presentation group (**d**). Three key features concerning the exploration behaviour of the participants were identified: 1) *Spatialised audio feedback lead to a deviation from systematic exploratory behaviour*. Compare **a** with **b**,**c**,**d**. 2) *The simultaneous exocentric secondary display lead to too much physical displacement*. Compare **c** with **d**. 3) *The exocentric secondary displays encouraged a 'head scanning' behaviour*. Compare **c**,**d** with **b**. Solid red and blue lines illustrate the direction of exploration in the sonification layer and the interactive auditory display respectively. Short splines illustrate the participant's head direction logged at 2Hz. Route data within the gray circle identify examples of 'head scanning' behaviour.

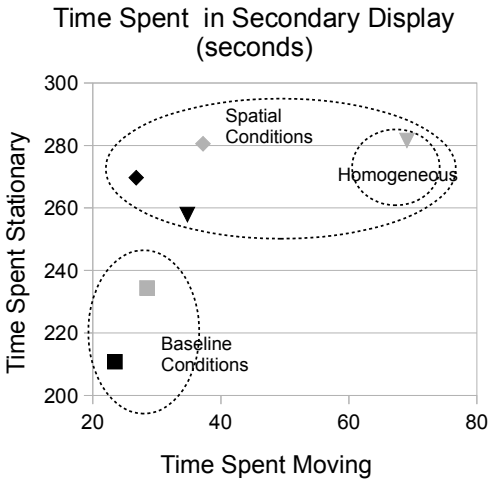

Fig. 13. In order to show the differences between conditions that required a lot of physical displacement and conditions that required a lot of stopping, time spent stationary (secs) is plotted against time spent moving (secs) for all conditions: Baseline (square), Egocentric (diamond) and Exocentric (triangle); and presentation group: Sequential (black) and Simultaneous (light grey). See section 3.2.2 for more detail on the experimental conditions. Black-dotted circumference identifies significantly different groups.

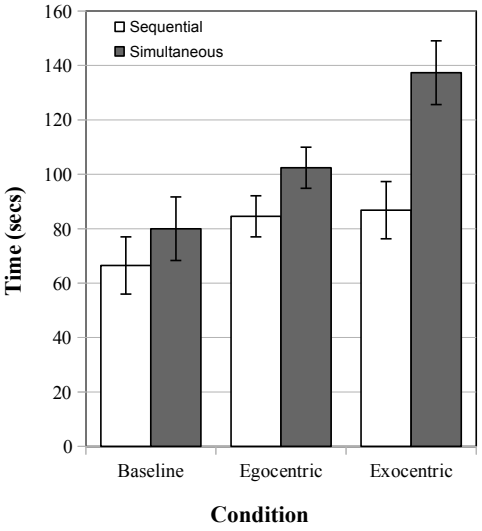

Fig. 14. Time spent stationary, when head orientation is changing by more than 5per second, per condition and presentation group: Sequential (white) and Simultaneous (dark grey), see section 3.2.2 for more detail on the experimental conditions. Error bars show Standard Error of Mean $\pm$ 1.0.

4.5. User Behaviour: Quantitative Analysis

Table II. Mean values for average speed, distance covered, time spent moving, time spent stationary and, while stationary, time spent head scanning (head orientation change $>$ than 5°/sec).

	Baseline		*Egocentric*		*Exocentric*	
	Seq.	Sim.	Seq.	Sim.	Seq.	Sim.
Speed (m/s)	0.029	0.027	0.027	0.032	0.037	0.062
Distance (m)	6.725	7.203	8.256	11.843	10.091	21.826
Time Spent Moving (secs)	23.465	28.463	26.791	37.213	34.771	69.071
Time Spent Stationary (secs)	210.763	234.366	269.698	280.568	257.877	281.624
Time Spent Stationary and Head Scanning (secs)	66.479	80.003	84.585	102.420	86.802	137.336

In this section, a further quantitative analysis of the logged data will extend the qualitative analysis in section 4.4 and provide a deeper insight into the differences between the experimental conditions.

The results from this additional analysis should be interpreted carefully given that the metrics are developed based on the previous qualitative evaluation, they are likely

to be correlated with our previous metrics (i.e. time spent), and are not independent of our initial analysis. Therefore they cannot be treated as evidence for supporting or rejecting hypotheses. However, these metrics provide a deeper insight into the user behavior observed in this study and offer potential for hypothesis testing and modelling in future work.

Two-way mixed-design ANOVAs were carried out on average speed (meters/second), distance covered in (metres), time spent stationary (seconds), and head scanning behaviour (seconds spent stationary and with head movement greater than 5°per second in the same direction), see Table II for mean values. ANOVA results were adjusted with a Bonferroni correction for multiple tests. We present the statistical analysis followed by our interpretation of the results.

Average speed. Significant effect by condition $F(2,60)=21.734$, $p<0.005$. No significant main effect was found for presentation group, but a significant interaction between condition and presentation group was found $F(2,60)=7.753$, $p<0.01$. *Post-hoc* paired samples t-tests with Bonferroni correction for condition by presentation group showed that average speed was significantly higher for the Exocentric condition ($t(15)=-8.433$, $p<0.01$) than the baseline and the Egocentric condition ($t(15)=-6.198$, $p<0.01$) for the simultaneous presentation group.

Distance covered. Significant effect by condition $F(2,60)=12.317$, $p<0.005$. No significant main effect was found for presentation group, but a significant interaction between condition and presentation group was found $F(2,60)=4.982$, $p<0.01$. *Post-hoc* paired samples t-tests with Bonferroni correction for condition by presentation group showed that distance covered was significantly higher for the Exocentric condition ($t(15)=-8.245$, $p<0.01$) than the baseline.

Time stationary. A trend was observed by condition $F(2,60)=4.447$, $p<0.064$. No significant main effect was found for presentation group. *Post-hoc* paired samples t-tests with Bonferroni correction per condition showed a tendency for participants to spend more time stationary in the Exocentric condition ($t(31)=-2.962$, $p<0.05$) than in the baseline and also in the Egocentric condition ($t(31)=-3.055$, $p<0.05$) when compared to the baseline across presentation group.

Head scanning. Significant effect by condition $F(2,60)=7.446$, $p<0.005$. A significant main effect was found for presentation group with the simultaneous presentation group showing significantly more time spent head scanning than the sequential group $F(1,30)=6.309$, $p<0.05$. *Post-hoc* paired samples t-tests with Bonferroni correction per condition showed that time spent head scanning was significantly higher for the Exocentric condition ($t(31)=-4.774$, $p<0.005$) than the baseline across presentation group.

Figure 12 shows greater physical movement in the simultaneous exocentric condition than in any other condition. In Table II, we can see that this movement resulted in a significantly increased average speed and average distance covered for participants in the simultaneous exocentric condition.

However, average speed ignores the stopping behaviour we observe in Figure 12. Examining the time spent stationary and the time spent in motion individually shows more clearly the difference between baseline and spatial conditions and between the homogeneous condition (simultaneous exocentric) and the other spatial conditions (See Figure 13). The time spent moving shows, as with average speed and distance, that the simultaneous exocentric condition resulted in much more movement. As we suggest in section 4.4, this extra requirement for movement increases workload and reduces user satisfaction for this condition. However, by looking at the time spent stationary, we can control for this extra movement effect. The ANOVA on time spent stationary showed a different trend between the Baseline conditions and all four spatial conditions, echoing our initial results for overall time spent in section 4.2, showing that participants spent

significantly less time interacting with the auditory display in the Baseline conditions when compared to the spatial conditions.

The head scanning behaviour illustrated in Figure 12 occurs when stationary so we can approximate this behaviour quantitatively by looking at time spent stationary, when head orientation is changing by more than 5°per second in the same direction. Examining this specific behaviour more closely (see means in Table II), we can observed a significant effect for presentation group with participants in the simultaneous group exhibiting more head scanning. This effect is also significant by condition with *post-hoc* tests showing a significant difference between Baseline and Exocentric head scanning behaviour (See Figure 14).

5. DISCUSSION

In this study, a number of different multilevel auditory display designs that varied in complexity were evaluated as part of a non-guided mobile audio-augmented reality environment. Previous work on eyes-free interaction design has focused on evaluations of different mobile spatial audio designs in semi-controlled task-based assessments [Marentakis and Brewster 2006; Brewster et al. 2003; Vazquez-Alvarez and Brewster 2011], whereas work on mobile audio-augmented reality has mainly focused on the design of a unique auditory display to deliver location-based information as part of the main system implementation [Wakkary and Hatala 2007; Eckel 2001; Heller and Borchers 2011]. This study combined both a systematic assessment of mobile spatial auditory displays and situated interactions within a mobile audio-augmented reality system to provide usability and user behaviour information. Here we discuss the findings relating to interface usability and the impact on user experience. We then outline implications for the design of spatial auditory displays, limitations and future work.

5.1. Usability and Impact on User Experience

A *consistent* multilevel layout, similar to the one used in ZUIs [Bederson 2011], was evaluated for presenting multiple information. We found that consistency across the multilevel auditory display had a profound negative impact on both usability and cognitive load. The simultaneous exocentric auditory display used an exocentric design that was consistent across levels but this *did not* improve usability. This was the only secondary display that *required* physical displacement, resulting in users experiencing higher workload and being less satisfied with this auditory display. In this study, it was the appropriateness of the interface to the task, rather than the consistency of the auditory display design which was the key requirement. The top-level sonification layer was used for searching for audio-augmented locations but using the same exocentric design for browsing content in the secondary interactive layer was not acceptable for users. The other three spatialised secondary interactive layers in the sequential exocentric, simultaneous egocentric, sequential egocentric auditory displays were found to encourage users to spend longer interacting with the artwork without a drop in user satisfaction or increase in workload reflecting greater exploration. Changing from an exocentric to an egocentric perspective did not appear to confuse the users who were able to move smoothly between the different layers without an increase in workload, or drop in user satisfaction.

As hypothesised, informal feedback suggests that the secondary spatialised interactive layer allowed for a more exploratory behaviour. Although, overall, the interface tested in the Baseline condition was reported as easy and faster to use, it was also found to be less immersive and less fun than the spatialised interfaces. Users liked the control over the interaction with the location-based information provided by the egocentric design and found that the exocentric design made the experience even more immersive. Performance results supported user feedback and helped characterise an

exploratory behaviour as one where interaction times will increase without an increase in workload and a decrease of user satisfaction. In this way, the Baseline and the Exocentric display with simultaneous presentation did not encourage an exploratory behaviour. However, the other three spatialised auditory displays, including the simultaneous egocentric display, did encourage an exploratory behaviour without a significant increase in workload. As one participant wrote in the visitor's book:

> The simultaneous presentation of *the menu items gave the whole experience a nice ambience and encouraged me to spend more time exploring the pieces.*

Overall, the mobile audio-augmented reality environment implemented for this study provided a successful user experience. One participant summed up well the sentiment of enjoyment that was echoed repeatedly throughout the visitor's book:

> *I enjoyed the idea of being able to move around a space and have the commentary adapt to me rather than the other way round. An altogether pleasant experience.*

In addition, this system was able to engage participants who mostly identified themselves as 'non-arty' by making the exhibit 'thought provoking'. A participant wrote on the visitor's book reflecting on the experience:

> *the audio comments helped provoke thoughts and appreciate the exhibition in a way I usually wouldn't. As a person who is not very 'arty' I spent more time looking at the pieces than I usually would.*

Although participants were required to wear a pair of headphones at all times in order to experience the audio-augmented environment, the use of a headset as part of this system was greatly supported by all participants. In the same way that graphical user interfaces can divert the attention from the exhibits, delivering content through headphones can have the negative effect of isolating users from their surroundings and their companions [Martin 2000; Stahl 2007]. In contrast to Wakkary and Hatala's [2007] study, in which user discomfort and low user satisfaction were reported, in our study headphones were always reported as comfortable and high levels of user satisfaction were observed (see Figure 8 for more detailed results on the *Headset* category per condition and presentation group). This suggests that open-back headphones successfully reduced the isolation from the physical environment usually experienced by users wearing closed-back headphones.

5.2. Implications for Design

The systematic evaluation of the multilevel spatial auditory displays presented in this article lead to the creation of a set of guidelines. These design guidelines aim at informing how to better support situated interaction with multiple location-based information in a mobile AAR environment.

— *Consistency is not paramount* When using a multilevel auditory display, using the same consistent configuration across levels is less important than using an appropriate configuration for the task at hand.
— *Spatial audio in an exocentric auditory display can be used to encourage exploratory behaviour.* Co-locating information in the physical space of an object generates a positive user involvement while navigating the space as it results in a more personal serendipitous exploration.
— *An exocentric auditory display operated by physical displacement increases user workload.* Take care when using such a display in an interactive multilevel design. De-

signing this type of display should take into account the complexities of navigating the environment and listening and interacting with the sounds within it.

5.3. Study Limitations and Future Work

The experimental task in this study was focused on exploring the gallery space and interacting with the location-based information as required. This is quite different from a traditional *treasure trail* task where the challenge of finding information is part of the experience, or an environment were a participant is accessing information in order to complete a set of predefined tasks. As previous work (e.g. [Vazquez-Alvarez and Brewster 2011]) has shown, the extent an audio interface divides attention has a strong impact on the user experience. In a different context, for instance where audio is used to provide instructions, participant behaviour may well be very different from that observed in our study. Furthermore, the top level exocentric display was designed specifically for an exploratory audio-augmented experience (as in [Vazquez-Alvarez et al. 2012]). This is an important use case but the results might not generalise to an interface for a different purpose.

A limitation relating to the auditory display design was the positioning of all the auditory sources only on the horizontal plane around the user's head. Considering elevation in the design of egocentric displays was outside the scope of this study. Including elevation in a spatial auditory display could offer greater flexibility and more complex design possibilities, however further baseline studies would be required to test its accuracy as part of an interactive auditory interface.

Another limitation is related to the use of non-individualised HRTFs to position virtual sound sources around the user. Individualised HRTFs provide better localisation results as they are custom generated for each individual user but they are difficult to employ as the setup and equipment to acquire them is complex and very expensive. Also, as shown by Mariette [2010], individualised HRTFs, although improving front-back localisation, can increase the perception of tracking deficits, where non-individualised HRTFs can blur angular distinction and produce a better overall user experience. However, for auditory displays to become more mainstream interfaces in mobile devices, the issues of user differences in 3D audio localisation ability need to be addressed. Non-individualised HRTFs provide worse localisation results but can be used by a much bigger number of users, which is the main reason why HRTF-based mobile phones use non-individualised HRTFs. Anatomical differences mean that systems using non-individualised HRTFs perform differently for different users. There is scope for systems to adapt to a user or be calibrated more easily by a user. For example, a set of headphones could adapt to the user's head width, or a system could analyse a photographic image of a user's pinnae, altering the HRTFs accordingly (see work by University of Sidney and University of York [2013] on mapping HRTFs to Ear Morphology). These approaches would not address non-anatomical individual variation in 3D audio perception, such as that caused by ear dominance [Noonan and Axelrod 1981]. In this case, either some type of smart user adaptation would be required or some sort of initial calibration. Some systems already offer users a choice between several HRTFs with a simple calibration exercise to help choose the most appropriate [Papa Sangre 2012]. Yet, the extent such a calibration phase could improve an auditory display is unclear. Should interfaces be designed to be resilient to individual variation or should they be designed to take user variation into account? Future work is required to deal with these issues, especially if auditory displays become more complex and more widespread.

Another area for improvement is to model the room acoustics more carefully. The current system produces the same audio experience whatever the room acoustics. Tailoring the augmented audio to the acoustics of the space they are in could improve

the sense of immersion and allow a more effective blend of virtual and environmental audio over a pair of open-back headphones.

Finally, in an exocentric display, user position together with real-time updating of the sound sources relative to the user orientation make users perceive the sound sources as being fixed to the physical space. Critical to the accuracy, immersiveness and believability of such interfaces is the precision and responsiveness of user location and head orientation tracking. Despite the successful implementation of this mobile audio-augmented reality prototype system, user location tracking still poses challenges. In this study, the system had to be calibrated for each user due to their height having an effect on the positional accuracy of the system. In addition, the fact that only one participant could be tracked at any time, and the multiple devices required made the system inappropriate for a less controlled environment. Recent work in computer vision [Eichner et al. 2012] looking at human pose estimation could be applied to address some of these issues. Potentially, such a system would be able to detect both user location and user head orientation for more than one user in a space equipped with multiple cameras. Such a system would also offer the advantage that no additional sensors would be required other than a user's mobile phone. These advantages would allow for deployment in a semi-public space, such as an art gallery or supermarket.

6. CONCLUSIONS

In this article, we investigated the efficiency and usability of complex spatial auditory displays designed to enable user interactions with concentrated areas of information. We compared the users' experience and performance when interacting with a number of multilevel spatial auditory displays in an exploratory mobile audio-augmented reality environment. Multilevel displays enable the presentation of simultaneous auditory streams and allow the structuring of information in concentrated areas in a location-based system. Both egocentric and exocentric designs were combined in the multilevel auditory display to test whether a consistent design across levels would be preferred over a mixed-design and whether these spatial audio techniques would encourage an exploratory behaviour. Our findings show that using a consistent exocentric design in the multilevel auditory display was not preferred. Also, by including a formal assessment of perceived workload and user satisfaction as part of the evaluation of user experience, it was possible to determine that a consistent exocentric design also failed to encourage an exploratory behaviour. However, the combination of a top-level exocentric configuration and an egocentric secondary configuration did encourage exploratory behaviour without overloading the user, even when auditory sources were presented simultaneously.

Our results suggest that spatial audio encourages both an immersive experience and an exploratory behaviour but it is important to avoid overloading the user. Results also show that users can switch between egocentric and exocentric display types readily, so using the same configuration is less important than using an appropriate configuration for the task at hand. Informal feedback suggests that an interface allowing for simultaneous presentation can also be more immersive but such an interface should be very carefully designed as simultaneous presentation can increase workload.

We hope that our findings will allow designers to make more informed decisions when designing eyes-free auditory interfaces for mobile audio-augmented reality environments.

ELECTRONIC APPENDIX

The electronic appendix for this article can be accessed in the ACM Digital Library.

ACKNOWLEDGMENTS

The authors would like to thank all the participants who took part in the study. We are also grateful to David McGookin for the support given during the user trials and to other colleagues (i.e., Graham Wilson and Martin Halvey) that helped with pilots and video documentation. This work was jointly funded by Nokia and EPSRC research grant EP/F023405.

REFERENCES

B. B. Bederson. 1995. Audio Augmented Reality: A Prototype Automated Tour Guide. In *Proceedings of CHI 1995*, Vol. 2. ACM Press, New York, NY, 210–211.

B. B. Bederson. 2011. The promise of zoomable user interfaces. *Behaviour & Information Technology* 30, 6 (2011), 853–866.

D. R. Begault. 1994. *3D Sound for Virtual Reality and Multimedia*. Academic Press, Boston, MA. USA.

L. Betsworth, N. Rajput, S. Srivastava, and M. Jones. 2013. Audvert: using spatial audio to gain a sense of place. In *Human-Computer Interaction – INTERACT 2013*. Springer, 455–462.

M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg. 1989. Earcons and Icons: Their Structure and Common Design Principles. *Human-Computer Interaction* 4, 1 (1989), 11–44.

H. N. Boone Jr and D. A. Boone. 2012. Analyzing Likert Data. *Journal of Extension* 50, 2 (2012).

S. A. Brewster, J. Lumsden, M. Bell, M. Hall, and S. Tasker. 2003. Multimodal 'Eyes-Free' Interaction Techniques for Wearable Devices. In *Proceedings of CHI 2003*. ACM Press, New York, NY, 463–480.

A. W. Bronkhorst. 2000. The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions. *Acustica* 86 (2000), 117–128.

M. Cohen, S. Aoki, and N. Koizumi. 1993. Augmented audio reality: Telepresence/VR hybrid acoustic environments. In *Proceedings of RO-MAN: 2nd IEEE International Workshop on Robot and Human Communication*. IEEE, Tokyo, Japan, 361–364.

N. Correia, T. Mota, R. Nóbrega, L. Silva, and A. Almeida. 2010. A multi-touch tabletop for robust multimedia interaction in museums. In *ACM International Conference on Interactive Tabletops and Surfaces (ITS '10)*. ACM Press, New York, NY, 117–120.

C. Dicke, K. Wolf, and Y. Tal. 2010. Foogue: eyes-free interaction for smartphones. In *Proceedings of MobileHCI 2010*. ACM Press, New York, NY, 455–458.

G. Eckel. 2001. Immersive audio-augmented environments: the LISTEN project. In *IV'01: Proceedings of the 5th International Conference on Information Visualisation*. IEEE Computer Society Press, London, England, UK, 571–573.

M. Eichner, M. J. Marn-Jimnez, A. Zisserman, and V. Ferrari. 2012. 2D Articulated Human Pose Estimation and Retrieval in (Almost) Unconstrained Still Images. *International Journal of Computer Vision* 99, 2 (2012), 190–214.

C. Fu, W. Goh, and J. A. Ng. 2010. Multi-touch techniques for exploring large-scale 3D astrophysical simulations. In *Proceedings of CHI 2010*. ACM Press, New York, NY, 2213–2222.

W. W. Gaver. 1997. Auditory interfaces. *Handbook of human-computer interaction* 1 (1997), 1003–1041.

J. Goßmann and M. Specht. 2002. Location Models for Augmented Environments. *Personal and Ubiquitous Computing* 6, 5-6 (2002), 334–340.

M. G. Helander, T. K. Landauer, and P. V. Prabhu. 1997. *Handbook of Human-Computer Interaction*. Elsevier Science, Amsterdam, The Netherlands.

F. Heller and J. Borchers. 2011. Corona: Audio Augmented Reality in Historic Sites. In *Proceedings of MobileHCI 2011*. ACM Press, New York, NY, 51–54.

JAKE Sensor Pack. 2010. http://code.google.com/p/jake-drivers/

JSR-234 Advanced Multimedia Supplements API (AMMS). 2007. http://developer.nokia.com/resources/library/Java/_zip/GUID-D3E35E6F-0C45-48ED-B09D-F716E14C1C02/overview-summary.html

H. Lam and T. Munzner. 2010. *A guide to visual multi-level interface design from synthesis of empirical study evidence*. Morgan & Claypool, San Rafael, Calif. (1537 Fourth Street, San Rafael, CA 94901 USA).

I. Leftheriotis and K. Chorianopoulos. 2011. User experience quality in multi-touch tasks. In *Proceedings of the 3rd ACM SIGCHI symposium on Engineering interactive computing systems (EICS '11)*. ACM Press, New York, NY, 277–282.

Sol LeWitt. 1967. Paragraphs on conceptual art. *Artforum* 5, 10 (1967), 79–83.

A. M. Lund. 1997. Expert Ratings of Usability Maxims. *Ergonomics in Design* 5, 3 (1997), 15–20.

G. N. Marentakis and S. A. Brewster. 2006. Effects of Feedback, Mobility and Index of Difficulty on Deictic Spatial Audio Target Acquisition in the Horizontal Plane. In *Proceedings of CHI 2006*. ACM Press, New York, NY, 359–368.

N. Mariette. 2010. Navigation Performance Effects of Render Method and Head-Turn Latency in Mobile Audio Augmented Reality. In *Auditory Display*. Lecture Notes in Computer Science, Vol. 5954. Springer Berlin Heidelberg, 239–265.

D. Martin. 2000. Audio Guides. *Museum Practice* 5, 1 (2000), 71–81.

J. McCarthy and P. Wright. 2004. *Technology as Experience*. MIT Press, Cambridge, MA, USA.

D. McGookin. 2004. *Understanding and improving the identification of concurrently presented earcons*. PhD thesis. School of Computing Science, Glasgow, UK.

D. McGookin and S. Brewster. 2012. PULSE: The Design and Evaluation of an Auditory Display to Provide a Social Vibe. In *Proceedings of CHI 2012*. ACM Press, New York, NY, 1263–1272.

Microsoft. 1995. *The Windows Interface Guidelines for Software Design*. Microsoft Press, Redmond, WA, USA.

A. J. Morrison, P. Mitchell, and M. Brereton. 2007. The lens of ludic engagement: evaluating participation in interactive art installations. In *Proceedings of the 15th international conference on Multimedia, ser. MULTIMEDIA 07*. ACM Press, New York, NY, 509–512.

E. Mynatt, M. Back, R. Want, M. Baer, and J. B. Ellis. 1998. Designing Audio Aura. In *Proceedings of CHI 1998*. ACM Press, New York, NY, 566–573.

J. Nielsen. 1994. *Heuristic evaluation*. John Wiley & Sons, New York, NY, USA.

M. Noonan and S. Axelrod. 1981. Earedness (Ear choice in monaural tasks): Its measurement and relationship to other lateral preferences. *Journal of Auditory Research* 21 (1981), 263–277.

Papa Sangre. 2012. http://www.papasangre.com

V. Roth, P. Schmidt, and B. Güldenring. 2010. The IR ring: authenticating users' touches on a multi-touch display. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology (UIST '10)*. ACM Press, New York, NY, 259–262.

N. Sawhney and C. Schmandt. 2000. Nomadic Radio: Speech and Audio Interaction for Contextual Messaging in Nomadic Environments. *ACM Transactions on Computer-Human Interaction* 7, 3 (2000), 353–383.

SHAKE SK6 sensor pack. 2010. http://code.google.com/p/shake-drivers/

B. Shneiderman. 1998. *Designing the User Interface*. Addison-Wesley, Reading, Massachusetts, USA.

C. Stahl. 2007. The Roaring Navigator: A Group Guide for the Zoo with a Shared Auditory Landmark Display. In *Proceedings of MobileHCI 2007*. ACM Press, New

York, NY, 282–386.
L. J. Stifelman. 1994. *The cocktail party effect in auditory interfaces: a study of simultaneous presentation*. Technical Report. MIT Media Laboratory Technical Report.
L. Terrenghi and A. Zimmermann. 2004. Tailored audio augmented environments for museums. In *IUI 04: Proceedings of the 9th international conference on Intelligent user interfaces*. ACM Press, New York, NY, 334–336.
University of Sidney and University of York 2013. http://sydney.edu.au/engineering/electrical/carlab/hrtfmorph.htm. (2013).
Y. Vazquez-Alvarez and S. A. Brewster. 2011. Eyes-Free Multitasking: The Effect of Cognitive Load on Mobile Spatial Audio Interfaces. In *Proceedings of CHI 2011*. ACM Press, New York, NY, 2173–2176.
Y. Vazquez-Alvarez, I. Oakley, and S. A. Brewster. 2012. Auditory display design for exploration in mobile audio-augmented reality. *Personal and Ubiquitous Computing* 16, 8 (2012), 987–999.
R. Wakkary and M. Hatala. 2007. Situated Play in a Tangible Interface and Adaptive Audio Museum Guide. *Journal of Personal and Ubiquitous Computing* 11, 3 (2007), 171–191.
A. Walker, S. Brewster, D. McGookin, and A. Ng. 2001. Diary in the Sky: A Spatial Audio Display for a Mobile Calendar. In *People and Computers XV-Interaction without Frontiers*, Ann Blandford, Jean Vanderdonckt, and Phil Gray (Eds.). Springer-Verlag, London, 531–539.
H. Zhang, X. Yang, B. Ens, H. Liang, P. Boulanger, and P. Irani. 2012. See me, see you: a lightweight method for discriminating user touches on tabletop displays. In *Proceedings of CHI 2012*. ACM Press, New York, NY, 2327–2336.
S. Zhao, P. Dragicevic, M. Chignell, R. Balakrishnan, and P. Baudisch. 2007. Earpod: eyes-free menu selection using touch input and reactive audio feedback. In *Proceedings of CHI 2007*. ACM Press, New York, NY, 1395–1404.