

2016-03-01

MobiLAudio – a Multimodal Content Delivery Platform for Geo-Services

James Carswell

Technological University Dublin, james.carswell@tudublin.ie

Keith Gardiner

Technological University Dublin, keith.gardiner@tudublin.ie

Charlie Cullen

Technological University Dublin, charlie.cullen@tudublin.ie

Follow this and additional works at: <https://arrow.tudublin.ie/dmcart>

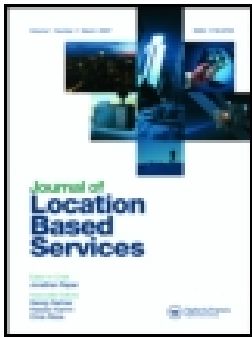


Part of the [Databases and Information Systems Commons](#)

Recommended Citation

Gardiner, K., Cullen, C. & Carswell, J. D., (2016) MobiLAudio – a multimodal content delivery platform for geo-services, *Journal of Location Based Services*, doi:10.1080/17489725.2015.1119319

This Article is brought to you for free and open access by the Digital Media Centre at ARROW@TU Dublin. It has been accepted for inclusion in Articles by an authorized administrator of ARROW@TU Dublin. For more information, please contact arrow.admin@tudublin.ie, aisling.coyne@tudublin.ie, vera.kilshaw@tudublin.ie.



MobiLAudio – a multimodal content delivery platform for geo-services

Keith Gardiner, Charlie Cullen & James D. Carswell

To cite this article: Keith Gardiner, Charlie Cullen & James D. Carswell (2016): MobiLAudio – a multimodal content delivery platform for geo-services, Journal of Location Based Services, DOI: [10.1080/17489725.2015.1119319](https://doi.org/10.1080/17489725.2015.1119319)

To link to this article: <http://dx.doi.org/10.1080/17489725.2015.1119319>



Published online: 01 Mar 2016.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

MobiLAudio – a multimodal content delivery platform for geo-services

Keith Gardiner, Charlie Cullen and James D. Carswell

Digital Media Centre, Dublin Institute of Technology, Dublin, Ireland

ABSTRACT

Delivering high-quality context-relevant information in a timely manner is a priority for location-based services (LBS) where applications require an immediate response based on spatial interaction. Previous work in this area typically focused on ever more accurately determining this interaction and informing the user in the customary graphical way using the visual modality. This paper describes the research area of multimodal LBS and focuses on audio as the key delivery mechanism. This new research extends familiar graphical information delivery by introducing a geoservices platform for delivering multimodal content and navigation services. It incorporates a novel auditory user interface (AUI) that enables delivery of natural language directions and rich media content using audio. This unifying concept provides a hands-free modality for navigating a mapped space while simultaneously enjoying rich media content that is dynamically constructed using such mechanisms as algorithmic music and phrase synthesis to generate task-relevant content based on the path taken. This paper outlines the innovative ideas employed in the design of the AUI and details the geoservices platform developed for facilitating the authoring and delivery of multimodal LBS applications. The paper concludes with a discussion on the results of a live user trial. The results are analysed and presented to validate the original hypothesis for this approach, address the research questions outlined and to inform further research directions. The results show that the proposed solution significantly progresses the state-of-the-art in terms of mobile tour production. The results also show that an AUI is an effective modality for the delivery of audio content and natural directions when used in combination with a graphical user interface, producing significantly reduced overheads in terms of content size and network usage. The results also indicate that the AUI provides a good overall user experience, performing well in the user trial.

ARTICLE HISTORY

Received 23 December 2014
Accepted 4 November 2015

KEYWORDS

Multimodal content delivery; geoservices; immersive audio; phrase synthesis; natural directions; algorithmic music

1. Introduction

This paper describes research and development of a novel geoservices platform for multimodal delivery of high-quality and task-relevant content to constrained mobile devices

(e.g. spatially enabled smartphones). Importantly, it investigates delivering location-based content to the user by means of the phone's aural modality as the primary delivery mechanism. The visual modality is also presented using a graphical user interface (GUI) that incorporates and builds on many of the ideas developed in previous work on mobile spatial interaction for location-based services (LBS) (Carswell, Gardiner, and Neumann 2003; Gardiner and Carswell 2003; Gardiner 2004; Gardiner et al. 2005; Rizzini et al. 2006; Kilfeather et al. 2007; Rooney, Gardiner, and Carswell 2007; Gardiner, Yin, and Carswell 2009; Carswell, Gardiner, and Yin 2010a,b).

The aural modality incorporates an auditory user interface (AUI), which is the main innovation in terms of delivering context-sensitive information to the user using focus independent (eyes-free) means. The AUI is intended to be a non-visual interface that can be used in combination with a GUI, or not. The idea is that all information in the application (including media content, directions, advertisements, etc.) can be delivered without the user ever needing to touch the device.

There are two main motivations for this work. Firstly, the concept of multimodality is investigated in terms of both multimodal input and output. Historically, mobile applications used just one form of multimodal input or output. More recently, the idea of combining multiple modalities at the input and output stages has gained significant traction in the mobile spatial interaction research community. Combining touch/gesture interaction with spatial navigation can provide a unique method of information discovery to enhance the overall user experience. Similarly, combining modalities at output level by using the visual and aural modalities together can provide a synergistic method of content delivery. The second motivation for this work is the development of an AUI for geoservice content delivery. The idea of combining mobile spatial interaction data with high-quality media content in the form of audio assets is explored where the goal is to enhance the user experience using multiple modalities while achieving a significant reduction in processing/data overheads. By using a phrase synthesizer to dynamically construct all content delivered by the system, physical storage requirements for the content is significantly reduced, resulting in the possibility to store all content data in-app. This further reduces network overhead as the requirement for external data calls is minimized.

Considering the above motivations, there are two main research questions related to this work. The first of these (Q1) is concerned with issues surrounding mobile tour production. To date, producing mobile tours necessitates a significantly arduous manual process, where geodata is collected and combined/linked with business data (content) to produce the point-of-interest (POI) objects used by the application. This manual process is repeated for each tour and requires significant resources. One important question is whether this process can be improved by reducing this manual overhead. Our work proposes a solution by developing a mobile tour production platform for audio tours, where each stage feeds data to the next using a structured workflow. In traditional GUI-based applications, this has been achieved already; however in our case, the intended output is a multimodal application that also delivers content using an AUI. A second important research question (Q2) is whether the use of an AUI is even effective in the delivery of content for geoservices. The hands-free and eyes-free nature of audio as a delivery mode has significant advantages over the focus dependant nature of GUIs; however, GUIs do provide more control. One of our aims therefore is to measure the overall user experience while using the AUI as the primary modality for content delivery.

The proposed approach to addressing the research questions outlined above is the development of a platform for multimodal geoservice delivery that will incorporate an AUI for multimodal content exploration. There are four main parts to this research.

- Survey Stage – Development of a mobile app for collecting region-specific geo-information.
- Author Stage – Implementation of an authoring tool that focuses on the structure of the content and how the overall process can be optimized.
- Delivery Stage – Development of a mobile app for delivery of content to users in a clear and natural way using an AUI and/or GUI.
- User Trial Stage – Testing the effectiveness of the multimodal tour guide in terms of functionality, usability and user experience.

The fourth and arguably most significant part of this research is the live user trial that was performed to test the functionality, usability, and overall user experience of the app. The user trial was carried out in Dublin City where a James Joyce Tour was created using the three-stage process outlined above and released as part of James Joyce week. The trial included 31 participants who used the app to navigate through the City Centre, visiting 12 plaques on route relating to James Joyce's book *Ulysses*. The app provided information on each plaque, directions between plaques and information on local amenities and commercial offers.

In Section 2, related work is presented with a focus on four of the most relevant aspects of previous research in this area, LBS, mobile spatial interaction and AUIs and multimodal content delivery. In Section 3, the geoservices platform is introduced which describes the three-stage process based on the idea of a survey, author and delivery methodology for app development, and deployment. Section 4 gives a comprehensive account of the live user trial carried out on the AUI including a detailed analysis of the results, outlining relevant observations, outcomes and recommendations. In Section 5, conclusions are drawn and further work is proposed.

2. Background

Location-based information retrieval has been a very active and pertinent area of research for many years (Bellotti et al. 2002; Persson et al. 2003; Wilson and Shafer 2003; Shi et al. 2010). Since the first methods of digitally creating maps, POI locations and geotags, the idea of providing context-sensitive services based on a user's location supported by an underlying geographic map have become increasingly important. These developments form the basis for previous research in this area and in particular work in the areas of location-based information retrieval, mobile spatial interaction, AUIs and multimodal content delivery.

2.1. Location-based information retrieval

With advances in mobile device technology, the idea of LBS has become a reality in many ways. However, even before such devices were commercially available, initial research presented by (Carswell et al. 2002; Gardiner 2004), describes work that simulated the concept of LBS by querying an underlying map on a desktop PC. Using a user's position and orientation in the map, a content delivery service was provided in the form of simple text and images. A 3D virtual map of Dublin City was developed to demonstrate and monitor the user's

physical interaction within the environment. The map was stored in a spatial database and registered to real-world coordinates together with a number of POIs representing notable locations relevant to Ireland's 'Easter Rising' in 1916. Using these three pieces of structured data, an early example of the potential of LBS was demonstrated.

Subsequent research focused on the accurate detection of POIs in a 2D space (Gardiner and Carswell 2003) based on the senses, i.e. how we identify and consume information presented to us. The visual modality was studied as the primary means of consuming information and more importantly the restrictions that line-of-sight (LOS) presented (Gardiner 2004). Defining actual visibility presented varying degrees of complexity because the location and definition of the interaction have to be accurate (Carswell et al. 2002; Rooney, Gardiner, and Carswell 2007), to effectively deliver location-based information in the correct context.

2.2. Mobile spatial interaction

Mobile spatial interaction is the study of the interaction between a mobile user and a world of spatial information. In (Gardiner and Carswell 2003; Gardiner 2004; Chan et al. 2005; Fröhlich et al. 2006; Simon and Fröhlich 2007a, 2007b; Simon, Fröhlich, and Grechenig 2008; Gardiner, Yin, and Carswell 2009; Carswell, Gardiner and Yin 2010a), the idea of an egocentric view of the world is investigated in a mobile context. In (Gardiner and Carswell 2003), the idea of using a triangular 2D shape to represent a user's field-of-view (FOV) to query POIs stored in a spatial database was introduced. By using a user's actual FOV as the 'query window' in the database, the query results are limited to what is actually visible to the user in a particular direction. This research proved successful in reducing the associated problem of information overload.

To increase query relevance and reduce information overload further, in (Strachan and Murray-Smith 2009; Gardiner, Carswell and Yin 2009), the idea of generating a more accurate FOV was demonstrated. This research focused on determining the exact shape of the user's viewport. Extending this an Isovist algorithm (Benedikt 1979) was used to determine 360° LOS. This resulted in a more representative viewport shape, minus the geometry of any buildings or objects that interrupt a user's LOS.

In (Strachan, Williamson, and Murray-Smith 2007; Fröhlich, Simon, and Baillie 2009; Gardiner, Yin, and Carswell 2009; Carswell, Gardiner and Yin 2010b), the focus shifted to the area of 3D directional queries, where the idea of 'point-to-query' was investigated. Before this, the idea of performing this specific type of spatial search was not feasible for the following reasons. Firstly, although storing 3D data was possible in some commercial database management systems at the time, performing 3D queries at database level was not. This was due to a lack of built-in 3D spatial query operators. Secondly, the availability of mobile devices that contained the sensor hardware necessary to measure both the horizontal and vertical orientations of the device was still limited to high-end or customized research devices. However, with the introduction of smartphones and improvements in spatial database technology, these barriers diminished and new research in the area of mobile 3D queries was now possible (Robinson, Eslambolchilar, and Jones 2008; Gardiner, Yin, and Carswell 2009).

A practical example of mobile egocentric-visibility querying was successfully demonstrated in a simulated *Internet of Things* environment by the 3DQ query processor developed in (Carswell, Gardiner and Yin 2010a; Yin and Carswell 2013). 3DQ enabled a user to point a smartphone anywhere at a building to obtain information about any visible feature (e.g. window, door, floor, environmental sensors, etc.).

2.3. Auditory user interfaces

An AUI is defined as ‘the use of sound to communicate information about the state of an application or computing device’ (McGookin 2004; Shneiderman 2005). AUIs have traditionally been used for accessibility by the visually impaired in desktop applications with the main objective of transforming a graphical representation to an auditory one (Edwards 1989; Gaver 1989; Mynatt 1994; Donker, Klante, and Gorny 2002; Jagdish et al. 2008; Shinohara and Tenenbergh 2009). Auditory non-speech audio has also been used to provide additional information to users about navigating a GUI (Kramer 1993; Absar and Guastavino 2008). This idea has gained popularity in the LBS and augmented reality (AR) space to provide navigational information to users of mobile applications (Walker and Lindsay 2006; Baus et al. 2007; Stahl 2007).

In (Holland, Morse, and Gedenryd 2002), the use of tones and audio panning is used in an AudioGPS application to communicate direction and distance to direct a user through a space. User trials show that the use of simple tones and the placement of them on the stereo stage are effective for indicating direction and that a simple oscillating tone was effective for communicating distance. Related work by (Strachan et al. 2005; Strachan, Williamson, and Murray-Smith 2007) investigated the use of music for the purposes of navigation where a listener’s music is modulated according to the changing predictions of a user’s position to guide them through a space.

In the context of multimodal geoservice delivery, the use of AUIs as an output modality has shown significant potential in the area of multimodal content delivery. In addition to providing instructional and navigational content, AUIs can deliver audio content in the form of real or synthesized speech to inform users about the surrounding environment (Abowd et al. 1997). This approach was previously used to augment the environment by providing context-sensitive information with applications in the field of mobile tour guides (Bederson 1995).

2.4. Multimodal content delivery

The concept of mobile multimodal content delivery therefore has been the topic of considerable research in recent times with multiple output modalities now available on today’s off-the-shelf devices (Chittaro 2010). Combining output modalities to provide information in a synergistic manner by augmenting a space has also gained attention (Lalanne et al. 2009; Turk 2014). The use of audio for content delivery is described in (Bartie and Mackaness 2006) where a combination of visibility querying and audio AR is used to create a CityGuide application. The application proved to be an effective, non-invasive approach to augmenting the user’s reality as it offered hands-free and eyes-free usage, enabling the user to keep focused on the environment.

The use of spatialized soundscapes to convey contextual information has been used in many cases in an attempt to deliver an immersive user experience (Wenzel 1994; Misra, Perry Cook, and Wang 2006; Stahl 2007; Vazquez-Alvarez, Oakley, and Brewster 2012). The LISTEN project (Eckel 2001) provides users with intuitive access to personalized and situated audio information to enhance the sensual impact of applications ranging from art installations to entertainment events. This is achieved by augmenting the physical environment using immersive, spatial audio and advanced user modelling methods.

An interesting approach was taken by (Heller et al. 2009) with the development of an interactive virtual tour where users can remain independent or can interact with other users on the tour, essentially 'checking in' with them. The ec(h)o system described in (Wakkary et al. 2004; Droumeva 2005) uses an audio AR interface that is used to deliver exhibition data based on a semantic structure and utilizes spatialized soundscapes to provide an immersive experience. Work carried out in (McGee and Cullen 2009a, 2009b) suggested the usefulness of audio as a delivery modality for LBS. In particular, the advantages identified were both its hands-free and eyes-free nature, plus fast neural processing rate. An experimental case study in (Vazquez-Alvarez, Oakley, and Brewster 2012) tests the use of earcons, proximity and spatial audio queues for effective information discovery using audio. A combination of 3D spatial audio techniques together with earcons proved the most effective audio display. This work also suggests that the location and orientation sensing technologies available in off-the-shelf devices are effective when used to create rich and compelling outdoor augmented-audio environments. All these works point to the benefits of the hands-and-eyes-free nature of non-visual modalities and to the potential the aural modality has for the provision of navigational and context-relevant information in a geoservices context.

3. MobiLAudio

The *MobiLAudio* system described in this paper extends the above related work by proposing a mobile audio tour guide production platform that incorporates a structured workflow for the collection of data, the production of media assets and the effective multimodal delivery of region specific information all in one application.

There are two significant innovations that differentiate the *MobiLAudio* platform from previous work, the first is the workflow itself which uses a unique survey–author–deliver metaphor for tour production, and secondly is the method by which the content is synthesized and delivered using an AUI at the delivery stage. The *MobiLAudio* platform builds on the work carried out in the development of the *mobiSurround* engine (Gardiner, Cullen, and Carswell 2015) with an emphasis on dynamic tour production and multimodal content delivery. In earlier research (Carswell et al. 2002, 2008; Carswell, Gardiner, and Neumann 2003; Gardiner and Carswell 2003; Rizzini et al. 2006, 2010), numerous issues with the amount and size of the media content required were encountered. In many cases, the amount of redundant data in the system is very high, where multiple files contained much the same content with small variations. In *MobiLAudio*, because a sizable amount of the content is now audio, the opportunity to reduce data redundancy using phrase synthesis is exploited.

The concept of phrase synthesis is based on taking a dictionary of recorded words and phrases, and combing them into sentences. This technique can deliver the required content via the AUI using a set of pre-recorded words and phrases that are synthesized at runtime using a set of production rules, which significantly reduces the required content in terms of size, thus reducing the overall size of the app. Taking advantage of this reduction in overhead, further research was carried out in the area of multichannel content delivery via the AUI. The multichannel AUI can provide up to 32 channels of audio information on a mobile device (Apple Developer Library 2014). In addition to the multiple channels of speech audio, multiple channels of background audio are used in the form of algorithmic music and 3D soundscaping to provide ambient audio and other context-relevant sound effects.

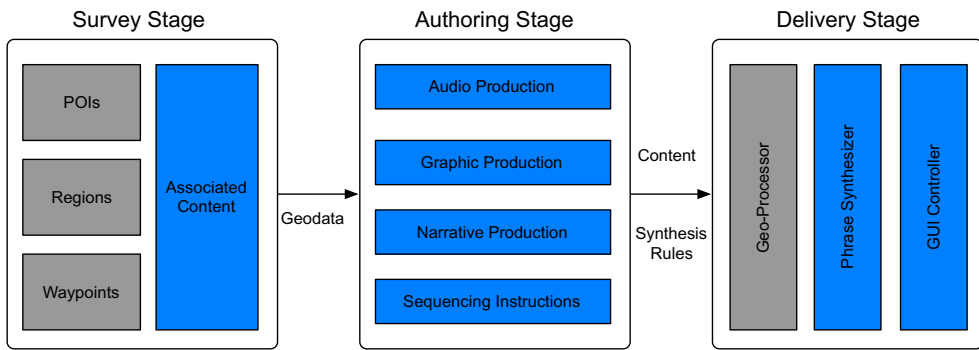


Figure 1. MobiLAudio system architecture diagram.

Note: This diagram describes the overall system architecture, which includes the survey, author and delivery stages of the tour production process.

Combining these, the AUI can deliver an immersive experience using real speech assets, background music generated on-the-fly, transient sound effects and natural directions. *MobiLAudio* developed a number of these innovative features, all of which improve and augment the viability of the AUI for real-world applications. Our focus is on the delivery of high-quality authored media content that is organized in such a manner that location events trigger timely media transitions. The *MobiLAudio* platform provides a toolset where professional HD tours (using high fidelity content) can be planned, authored and delivered. The production of a tour involves a three-stage process and begins with the use of a mobile survey tool to collect geodata. These data along with associated field-based content are passed to the content authoring tool, which models the geodata/content for the tour for delivery via multiple mobile platforms. The architecture of the *MobiLAudio* platform is illustrated in Figure 1.

The following three sections outline the three-stage process describing the novel aspects of tour production and multimodal content delivery that differentiates it from other approaches. The main focus is on the authoring and delivery stages of the process as this is where the main advances beyond the state-of-the-art are made (research question Q1).

3.1. Survey stage

The core concept of the *MobiLAudio* platform is the synergy of location data and professionally produced media content. To achieve this, a mobile survey tool was developed to provide a full LBS map of a given space and to geotag content for effective tour production. The tool allows a tour creator to annotate specific locations within a space by attaching rich media in the form of text, images, audio and video (Figure 2). The geodata collected by the survey tool are used to define the three core elements of a tour space, and are referred to as nodes, waypoints and regions.

3.1.1. Nodes

Nodes are considered a basic unit of data in the *Delivery Engine* and are used to represent a basic POI, service or facility (Figure 2). Each node contains a number of pieces of data that firstly identifies the node, the four neighbouring nodes, the closest waypoint and the region

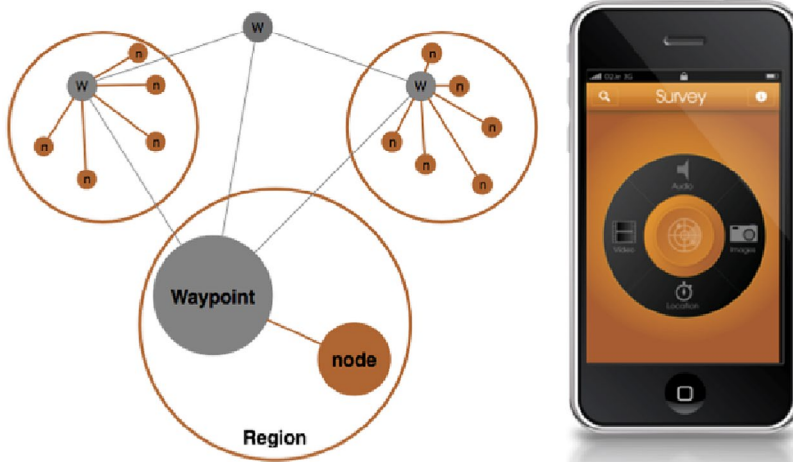


Figure 2. MobiLAudio surveying tool.

Note: The mobile app used to survey a space before the content production stage; it provides the geodata for nodes, waypoints and regions.

the node belongs to. Using the node ID, the engine determines what content to load on arrival, whether to change region and what node is next based on the direction the user has arrived from.

Each node has a number of associated media assets; these include images, text and audio. The audio is split into parts, each containing separate facts about a POI which are dynamically loaded at runtime by the phrase synthesizer. The phrase synthesizer is used in all node content delivery to allow a relatively small set of commentary phrases (delivered by the narrator) to be reused for a given node based on a set of linguistic rules. In this manner, node content is assembled at runtime to give the user the impression of listening to a documentary-style interview between narrator and expert.

3.1.2. Waypoints

Waypoints are the main navigational unit in the *Delivery Engine*. They can be considered as reference or control points for all other nodes in a space, where each node essentially belongs to or is governed by a waypoint such that navigation around a space is achieved by specifying directions between waypoints (Figure 2). Once navigation between waypoints has been achieved, directions from the destination waypoint and all of its associated nodes are accessible. Directions to the destination node are then provided. More significantly, waypoints are defined as natural landmarks that exist in an area making them perfectly suitable for delivering natural language directions. For example, ‘walk down as far as the Big Tree and take a left; the shop is on your right hand side’. In this instance, ‘the Big Tree’ is the destination waypoint and ‘the shop’ is the destination node.

3.1.3. Regions

Regions are defined as functional areas that divide a space. Each region has a set of physical boundaries that give it a unique geospatial identity. Regions can encompass multiple

waypoints and nodes, essentially helping to define the context in which they exist (Figure 2). Navigating between regions triggers a region change, which in effect signifies a change in context. The result is a managed fade-out of all contextual audio and sound effects for that region, followed by an introduction to the new region and a gradual fade-in of all contextual audio for this region. This subtle effect is a key feature of the adaptive virtualized 3D soundscaping described more in Section 3.3.

With the geodata mapping completed, the data are then used to design a brief for content authoring, indicating where trigger points lie, where waypoints should be and the boundary lines for each region. This helps to define exactly what needs to be rendered by the AUI and when (i.e. content modelling). This is done at the *Authoring Stage* where a set of production guidelines for speech audio, soundscaping, SFX and adaptive music are produced.

3.2. Authoring stage

With the geodata collected for nodes, regions and waypoints, the audio soundscape, adaptive music and narrative content for nodes and regions within a given tour must be composed and structured to trigger accurately as the user navigates the space. This is done using the *Authoring Tool*. The Authoring Tool is used to load the survey information for a given tour, and inform the design and production of soundscape, music, and narrative assets for all points and regions in the tour (Figure 3).

The data collected by the survey tool are raw geodata that needs to be structured and augmented so that it can be used. A data delivery matrix is used to structure the data where a node is defined as a logical POI, service or facility. Waypoints are defined as navigational reference points in the space and regions are logical boundaries that define a space. These basic elements are the spatial foundation of all functionality in the system. The translation of the required functionality into a content model for authoring is undertaken for each of the three elements in the matrix. These elements can be broadly defined in terms of:

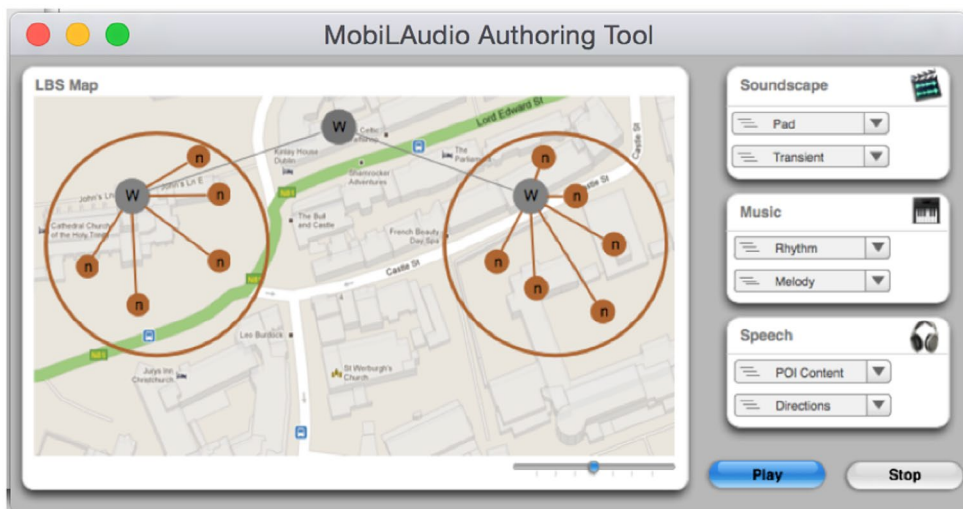


Figure 3. MobiAudio authoring tool.

Note: Desktop application used to organize audio assets for nodes, waypoints and regions by linking the assets for each node to geodata collected by the survey tool.

- Content Delivery – information relating to a specific node in the space.
- Navigation – information needed to traverse between points in the space using waypoints.
- Narrative Context – changes between defined regions allow the narrative to adapt to the new context of the user.

In order to collect and model tour data for effective delivery of content, navigation instructions, and narrative context, a content modelling strategy is used. This process (outlined below) helps ensure that the AUI is provided with accurate sequencing instructions at runtime.

3.2.1. Content modelling and selection

The central component of the *Authoring Tool* is based on a three-stage data modelling process that informs the various parts of the application that feed both the visual and aural presentation modalities (Figure 4).

3.2.1.1. Geodata modelling. The geodata modelling (GM) task is a process by which all geographic data that defines POIs, navigational waypoints and regions is modelled. These data inform the rest of system and is the basis of the content modelling strategy.

3.2.1.2. User modelling. The user modelling (UM) task is where user preferences are modelled, essentially defining the type of user and their context. These data have a scaling effect on the content modelling strategy where varying amounts of content is presented based on a user's profile (Figure 4).

3.2.1.3. Content modelling. The content modelling task is where the structure of the media content is defined for each node based on the geodata (GM) and application preferences data (UM). It is this strictly controlled process of data modelling from multiple perspectives that ensures data in the system is structured correctly, enabling the dynamic loading of media content to give the perception of an adaptive coherent narrative with real speech content and natural directions, and algorithmic music and 3D soundscaping (Figure 4).

The following section outlines the innovative approach to providing this experience and describes how they integrate to deliver the functionality that forms the basis of the Delivery Engine. In particular, aspects relating to phrase synthesis, adaptive 3D soundscapes, algorithmic music sequencing and natural directions are outlined.

3.3. Delivery stage

Following considerable investigations into the area of contemporary LBS coupled with an increasing focus on the affect that quality media assets have on providing a good user experience; the need for improved, innovative methods of delivering this content became apparent. This led to the development of a content *Delivery Engine* capable of delivering an immersive experience using multiple modalities while minimizing data redundancy using a reusable content approach. The research focused on several distinct areas of functionality necessary to deliver the required content, navigational instructions, music and soundscaping. This was achieved through several advances on the state-of-the-art, notably in the areas of

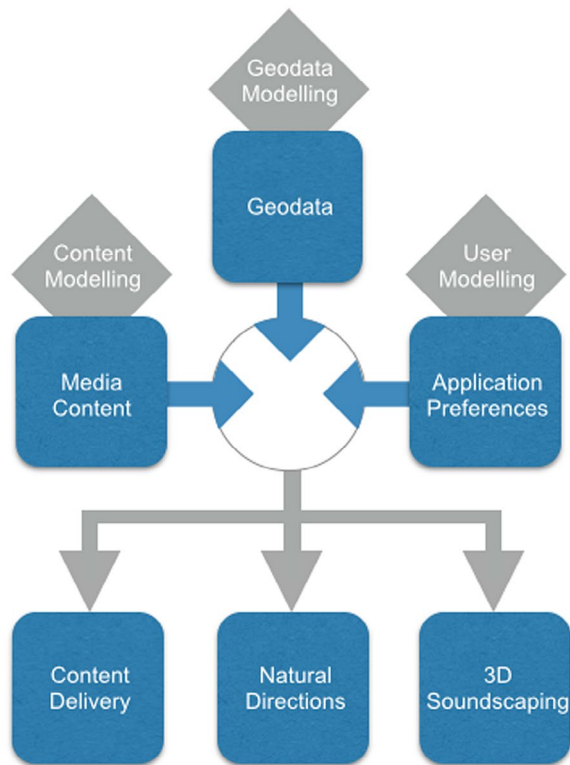


Figure 4. Three-stage data modelling process.

Note: This diagram describes the structure of the data required to build a mobile audio tour based on survey data, contextual information and media content assets.

multimodal content delivery and AUIs. The architecture of the *MobiLAudio Delivery Engine* is illustrated in Figure 5.

The concept of using an AUI for content delivery in a mobile context has been investigated before (Wang et al. 2008; Magnusson et al. 2009); however, in most cases, pre-recorded audio is used without consideration for a user's limited ability to perceive multiple audio sources at once (Potamianos and Perakakis 2008). For example, audibility problems can arise when speech and background music or sound effects are playing together. A model where multiple channels of audio are used to deliver speech, music and soundscaping separately based on priority is a more effective approach and is the main motivation for the development of the *MobiLAudio AUI*. The main problem here (relating back to research question Q2) is whether the use of an AUI based on a prioritized listening model can provide an immersive user experience, while effectively delivering speech content and navigational instructions to the user.

Our approach uses a multichannel audio mixer and phrase synthesizer to deliver the required music, speech and soundscaping based on a background to foreground listening model. This approach advances the state-of-the-art in the area of mobile multimodal interfaces with the development of an AUI that models content and narrative context to effectively deliver speech, navigation instructions, algorithmic music and soundscaping in a LBS context. More specifically the engine comprises:

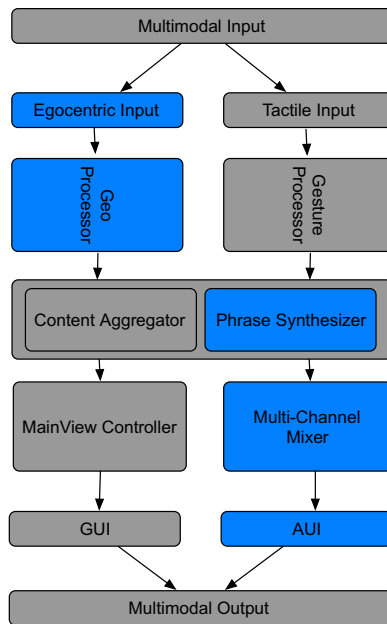


Figure 5. Delivery engine architecture.

Note: The diagram illustrates the architecture of the mobile delivery app, which includes the geoprocessor, phrase synthesizer and GUI controller.

- A background to foreground listening model.
- An intelligent content phrase synthesizer.
- A natural directions model for real-time navigation.
- A location manager with adaptive resolution monitoring.
- Adaptive virtualized 3D soundscapes.
- An algorithmic music sequencer.
- A multichannel auditory mixer.

The delivery platform chosen is an iOS application that uses a GUI in conjunction with an AUI for content delivery using earphones. The GUI for the iOS application is shown in Figure 6.

The most fundamental part of the AUI design is the underlying listening model. The following section describes the listening model and how it forms the foundation of each component in the system. The subsequent five subsections describe the main research contributions made to their respective areas during the development of the AUI. The final subsection describes the multi-channel audio mixer, which details how all components of the AUI integrate to produce an immersive user experience.

3.3.1. Listening model

The AUI delivery model is based on a background to foreground listening paradigm (Truax 2001). Research has shown that using this approach helps users to delineate foreground speech from background audio resulting in improved auditory perception (Beauchamp 2007; Darwin 2008). Aligning the audio features of the AUI with this listening paradigm prioritizes speech over background audio and music. This ensures that speech content and directional instructions are always in the foreground ensuring they are heard and coherent.



Figure 6. Delivery engine graphical user interface (GUI).

Note: This interface is used in manual mode to view data on the map and trigger POI playback.

The listening model illustrated in Figure 7 shows how content within the AUI is delivered to the user. A spatialized 3D soundscape is overlaid with algorithmic music that the user can detect but not prioritize directly. Speech content (either node information or directions) will always be foreground content. The prioritization of node content as the primary function of the AUI is defined as part of the speech model, with waypoint navigation as a specific subset of natural speech directions within this model. Region content is required to give change of context, and so a combination of adaptive virtualized 3D soundscaping and specialized algorithmic music is implemented to convey broader conceptual variations within the narrative of the space.

3.3.2. Intelligent content phrase synthesizer

Concatenative synthesis has been the subject of considerable research in the area of speech and music sound for many years (Schwarz 2005; Neustein and Markowitz 2013). Traditionally the technique synthesizes sounds by concatenating short sounds of up to 1 s, into sounds or words. In speech synthesis, concatenative synthesis methods are most widely used and result in more natural-sounding and intelligible speech (Schwarz 2004). The *MobiLAudio AUI* uses the more simplified approach of unit selection synthesis, where pre-recorded words and phrases are assimilated at runtime by the AUI using a process called phrase synthesis.

The intelligent phrase synthesizer developed delivers speech and natural direction content for mobile tours. The primary motivation for this approach was the reduction in mobile application size, which is an important consideration in mobile app development, and the significant reduction in data redundancy that this method achieves. By using this approach, all content is stored in-app, reducing the cost/reliance on mobile network data connectivity, avoiding issues experienced in previous projects with unreliable data connectivity. The phrase synthesizer is used in all node content delivery to allow a relatively small set of commentary phrases (delivered by the narrator) to be reused for a given node based on a set of linguistic rules. In this manner content for each node is assembled at runtime to give the impression of listening to a documentary-style interview between a narrator and a domain expert.

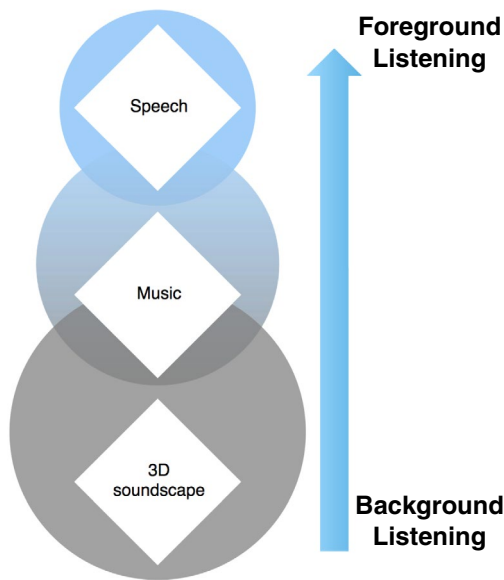


Figure 7. Delivery engine listening model.

Note: This model is based on a background to foreground listening paradigm, prioritizing speech over background and music.

Figure 8 illustrates how the phrase synthesizer assembles node, region, help and direction content sentences in real-time. Each channel (Intro, Content, altContent and Outro) is loaded with audio using a handle and a content phrase. The handle is randomly selected from a predefined library and the content relates to the nodeID of the current active node. An example of a synthesized sentence is as follows:

[You are now arriving at] + [the Snow Leopard exhibit] + [Lets ask john to give us some information] + [Snow Leopards are large cats that live way up high in the mountains ranges of central Asia] + [Wow that's interesting] + [FACT2] + [Cool can you give us another fact] + [FACT3] + [Interesting, what else can you tell us about it] + [FACT4] + [Well now you have heard about it lets move on to] + [the red river hog exhibit] + [DIRECTIONS]

Each node has two narrators associated with it, the main narrator and a domain expert. Facts 1 and 2 are initially loaded into the Content and altContent channels, and then swapped out for Facts 3 and 4. The phrases are constructed in real-time, loaded into the audio graph (MixerBus) and played. All node content and directions are constructed in this manner, significantly reducing the overhead of recording and playing large static files.

3.3.3. Natural directions model for real-time navigation

The area of personal way finding has gained significant attention in recent years due to the proliferation of GPS enabled mobile devices and personal navigation systems (Klippel et al. 2005; Kainulainen et al. 2007; Modsching et al. 2007; Stark, Riebeck, and Kawalek 2007). The augmentation of traditional map-based approaches with other forms of navigation such as the use of directional arrows and photographs of landmarks has been widespread (Burnett 2000; Chittaro and Burigat 2005; Millonig and Schechtner 2007; Hile et al. 2008).

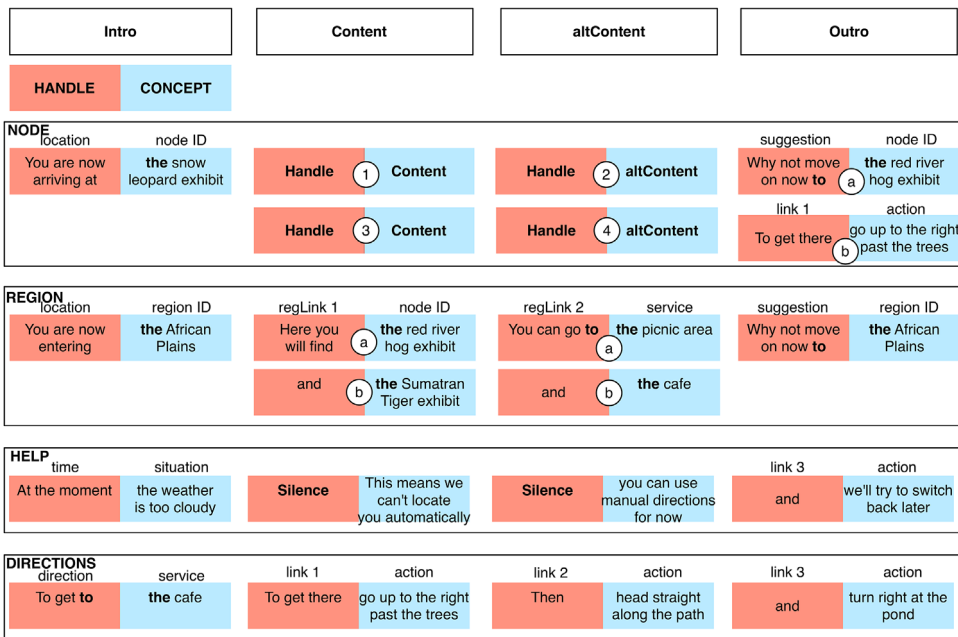


Figure 8. Phrase synthesis structure for node, region, help and directions content.

Note: The phrase synthesizer requires 4 MixerBus channels to facilitate the assembly and playback mechanism employed by the AUI.

This has also resulted in the use of audio as an aide to personal way-finding with voice guidance gaining particular interest. In (Rehrl et al. 2014), the use of voice, map and AR are used in the context of pedestrian navigation. Test results show that the use of audio-only guidance leads to significantly better navigation and user experience amongst users. A study in (Rehrl, Häusler, and Leitinger 2010) also indicates that voice-only navigation is preferable over other modalities and that users preferred landmark-based navigation instruction over metric based distances, mainly due to a preference for the more natural eyes-free and hand-free approach of egocentric audio directions (Chittaro and Burigat 2005). However, this approach also presents some difficulties mainly in the area of audio production and landmark surveying, where both aspects require significant manual effort.

These difficulties are the primary motivation to our approach taken in *MobiLAudio*, where a phrase synthesizer is used to reduce the amount of audio required to provide directions, and the initial survey stage of the three-stage tour production model is used to collect information on landmarks. The main research question here is whether natural directions can be constructed and delivered effectively using the survey–author–deliver methodology employed by the *MobiLAudio* platform.

In *MobiLAudio*, natural directions are defined as the use of audio to deliver directions based on visible points or landmarks in a space rather than providing metric distances and cardinal points. The advantages of natural directions are speed (humans do not need to translate data) and adaptability (there is no need to hold a compass in a certain way to obtain a heading) that allow them to be delivered hands free. The disadvantage of natural directions is the size of the content model required, where each direction is a unique narrative in itself.

To address this, the *Delivery Engine* natural directions model uses a combination of node and waypoint information to control the phrase synthesizer, providing a set of natural directions (which can also be displayed on a map if required) to any given point in the space. The model uses the concept of waypoint linkage to create a lookup table of directions between waypoints that is then completed by directions from the last waypoint to the node required. In this manner, natural points of focus within the space are first defined as waypoints using the survey tool and then the authoring tool is used to script and create the directions needed to link them together. When a specific node is queried, the only information required is its nearest waypoint, which allows a lookup table to complete the navigation path to it and subsequently configure the phrase synthesizer (Figure 9).

The example above demonstrates how waypoint linkages are used to provide directions to any node in the space. For the three destination nodes (e.g. zoo exhibits) in Figure 9, a common set of waypoint linkage directions is used to get to waypoint 3, where different waypoint combinations then lead to the specific nodes. Each node is associated with one waypoint only. When a direction query is performed, three pieces of information are needed.

- (1) Directions from the current node to its waypoint.
- (2) Directions from the current waypoint to the destination waypoint.
- (3) Directions from the destination waypoint to the destination node.

By building a set of waypoint linkages in both directions, (getting from W2 to W1 is not the same as going from W1 to W2), each node is linked to its nearest waypoint, which is then linked by a lookup table to every other waypoint. The phrase synthesizer is then loaded with the entries from the lookup table and provides the directions to the user as required (Table 1).

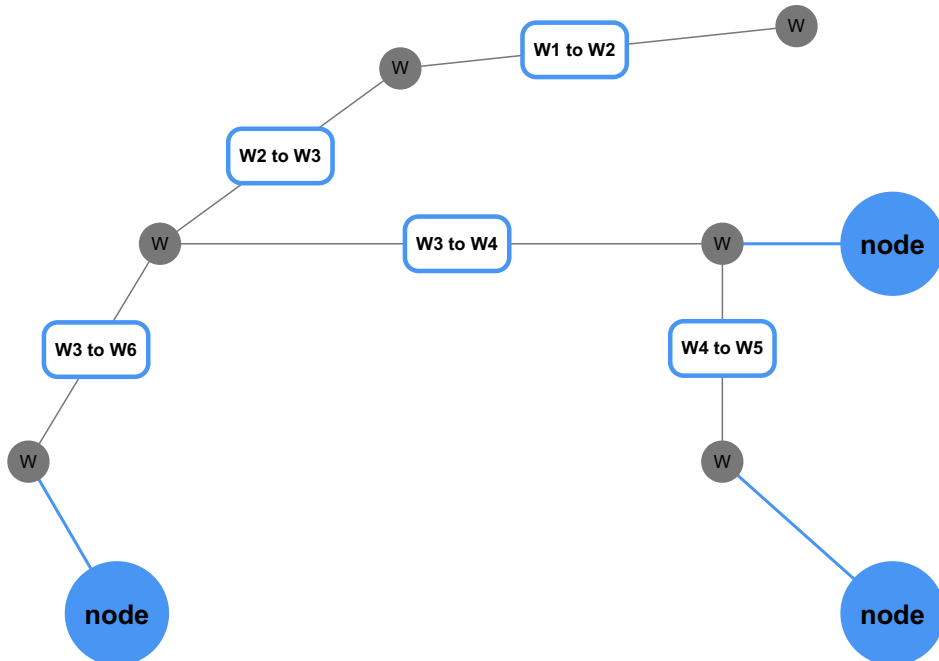


Figure 9. Waypoint linkage example.

Note: Each node is owned by a waypoint; directions are given between waypoints and each waypoint contains direction to each of its nodes.

Table 1. Waypoint lookup table.

| Waypoint | 1 | 2 | 3 | 4 | 5 | 6 |
|----------|--|--|--------------------------|--|--|--|
| 1 | | WP1 to WP2 | WP1 to WP2 WP2 to WP3 | WP1 to WP2 WP2 to WP3 WP3 to WP4 | WP1 to WP2 WP2 to WP3 WP3 to WP4 WP4 to WP5 | WP1 to WP2 WP2 to WP3 WP3 to WP6 |
| 2 | WP2-WP1 | | WP2 to WP3 | WP2 to WP3 WP3 to WP4 | WP2 to WP3 WP3 to WP4 WP4 to WP5 | WP2 to WP3 WP3 to WP6 |
| 3 | WP3 to WP2 WP2 to WP1 | WP3 to WP2 | | WP3 to WP4 | WP3 to WP4 WP4 to WP5 | WP3 to WP6 |
| 4 | WP4 to WP3 WP3 to WP2 WP2 to WP1 | WP4 to WP3 WP3 to WP2 | WP4 to WP3 | | WP4 to WP5 | |
| 5 | WP5 to WP4 WP4 to WP3 WP3 to WP2 WP2 to WP1 | WP5 to WP4 WP4 to WP3 WP3 to WP2 | WP5 to WP4 WP4 to WP3 | WP5 to WP4 | | WP5 to WP6 |
| 6 | WP6 to WP3 WP3 to WP2 WP2 to WP1 | WP6 to WP3 WP3 to WP2 | WP6 to WP3 | WP6 to WP3 WP3 to WP4 | WP6 to WP3 WP3 to WP4 WP4 to WP5 | |

Note: This lookup table is used to determine the files required by the phrase synthesizer to assemble and play natural directions between destination nodes.

Table 1 illustrates the structure of the lookup table used to determine directions between waypoints. The variable lists of linkages (part 2) are used in combination with parts 1 and 3 outlined above, to deliver effective directions using the phrase synthesizer. For example, if the user wants to navigate from node 6 to node 5, they first need directions from node 6 to waypoint 6, then directions from waypoint 6 to waypoint 5 and finally directions from waypoint 5 to node 5. Directions between nodes and their waypoints are named accordingly and are loaded automatically. The format of the directions between waypoints determined using the waypoint lookup table is shown below:

[To get to] + [Node 6] + [N6 to WP6] + [Then] + [WP6 to WP3] + [Then] + [WP3 to WP4] + [Then] + [WP4 to WP5] + [and] + [WP5 to N5] + [I hope this is clear, if not request directions again]

The example illustrates how the phrase synthesizer dynamically constructs directions using randomly selected handles combined with a series of directions segments. The result is rendered by the AUI, and animated visually on the map (Figure 10).

3.3.4. Location manager with adaptive resolution monitoring

The *Delivery Engine* is a spatialized AUI, but practical LBS require accurate positioning in order to deliver a satisfactorily immersive and hands-free user experience. The engine is by design fully automated in its response to changes in location, so a *Location Manager* was developed to manage this process (Figure 10).

The *Location Manager* detects both current region and nearest node in real time, with a location resolution monitor used to determine when the engine should deliver content for a given node. In addition, the manager can distinguish between different types of node (e.g. POI, facilities, events), as demonstrated in the Dublin Zoo tour (Figure 10).

In any situation where positioning accuracy is too low to deliver content effectively, an intelligent audio housekeeping manager updates the user in real time by informing them of their positioning conditions and instructs them to switch to manual mode, where all nodes



Figure 10. The delivery engine location manager interface.

Note: The location manager manages the visual representations of natural directions (demonstration using Dublin Zoo tour).

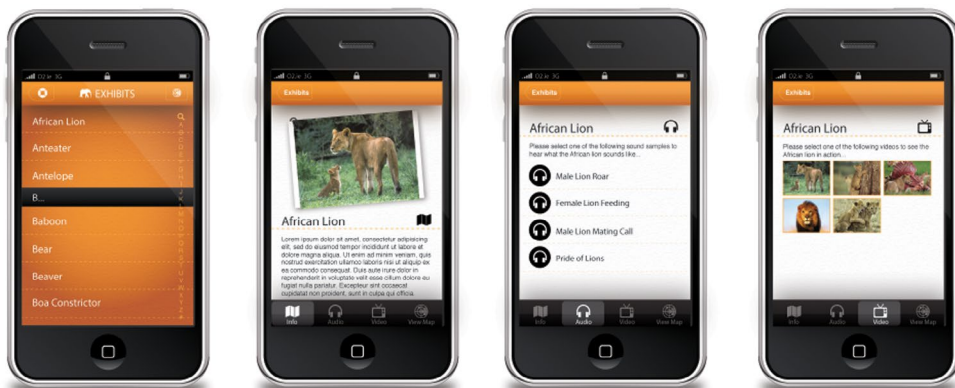


Figure 11. Delivery engine graphical user interface for manual content delivery.

Note: This mode is used when GPS is unavailable or the user requires a visual aide.

can be triggered by manual selection (Figure 11). In this manner, the engine ideally never delivers inaccurate content to the user providing them with a higher degree of usability. As an extension to this, the location manager incorporates live weather updates into its positioning accuracy assessment (accuracy is often linked to cloud cover) that is also reported to the user by the housekeeping manager.

3.3.5. Adaptive virtualized 3D soundscapes

The *Delivery Engine* is capable of delivering immersive soundscaping based on the region context of the user. Transitions between regions are dealt with automatically as fades between atmospheric 3D background sounds, as the user moves between different regions in a space. Fades are timed to function at walking pace so the user perceives a natural spatial transition between one context and another.

Each region has a specifically authored soundscape comprised background synth pad sounds and foreground transient sound effects (SFX), where a transient sounds (such as a bird singing high in a tree behind your right shoulder) are delivered using real-time transient positioning to give the listener a sense of a live environment. This is achieved using the DTS Audio Virtualizer to give the impression of movement within the soundscape, allowing sounds relating to non-stationary objects (e.g. flying birds) to be delivered algorithmically as part of the narrative. The engine automatically adapts to changes in location and context from a top down perspective, so transient SFX never clash with more important foreground sounds such as node content or directions.

The Virtualizer contains a low bitrate version of the codec for potential implementation on constrained devices. The virtualization algorithm takes a head related impulse response for a given azimuth (measured using a binaural dummy head) and convolutes this with the incident signal to deliver an output signal which is perceptually similar to one from that azimuth. In the case of the DTS Virtualizer, azimuths are provided for a common 5.1 multichannel audio setup that defines rear left (rL) and rear right (rR) positions. This Virtualizer was performance tested using the iOS audioGraph render model to determine whether a real-time fast Fourier transform convolution could be performed using the iOS render callback model.

The results (Figure 12) define a cardioid response pattern for rear content when delivered using the Virtualizer. This leads to a practical constraint of positioning where idealized radius r (translating to distance to source) is replaced by r_1 in the rear condition and r_2 for frontal cues in the virtualized case. Taking the desired effect into account, these results proved sufficient to produce a satisfactory result.

3.3.6. Algorithmic music sequencer

The *Delivery Engine* contains a full multipart algorithmic music sequencer, where regions are given a specific musical score containing several parts that vary automatically over time. The sequencer swaps content based on compositional rules provided in a similar manner to algorithmic music found in video games to ensure that the user never becomes bored with a continuous loop of music. The *Delivery Engine* sequencer also responds to contextual changes, swapping all music for each region to compliment the underlying soundscape using the same fading strategy. In this manner, the user may move wherever they wish, with SFX and music seamlessly adapting to their current context. Music and background pads are juxtaposed by placing them at the front and rear of the virtualized soundscape, respectively.

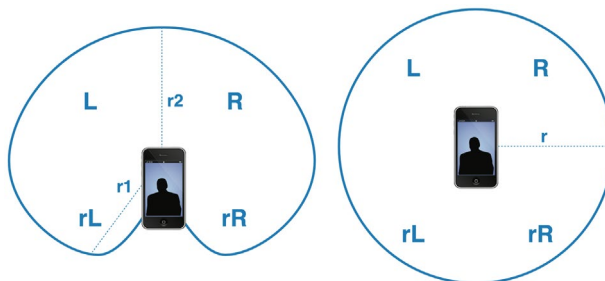


Figure 12. Virtualizer perception test results.

Note: This figure shows the comparison between the actual and ideal cardioid response pattern for delivering rear content using the DTS Virtualizer.

This provides the maximum spatial distance possible between content elements, providing the user with a much wider auditory vista within the interface.

To provide all functionality described in the previous five subsections, a very flexible approach to content delivery is required, where each element of the content can be controlled and managed separately. A multichannel audio mixer (*MixerHost*) was used to achieve this.

3.3.7. Multi-channel auditory mixer

The audio mixer component of the iOS Core Audio Framework was selected as it provides up to 32 independently controllable channels of audio. For our purposes, we used 16 channels of audio for phrase synthesis (content, help and directions), algorithmic music sequencing and 3D soundscaping. This requires a rigid approach to file management based on the linguistic rules and sequence instructions modelled at the authoring stage. This fundamental aspect of the entire process is really only evident at this stage in the process. The breakdown of the assigned channels for each aspect of audio content is illustrated in Figure 13.

The diagram above illustrates the structure of the *MixerBus* and the allocation of the channels for SFX (3D soundscaping), Music (algorithmic music) and Speech (content, directions, region changes and help).

The first four channels are reserved for SFX, which are a combination of ambient synth pads and virtualized transient sound effects. These are loaded at runtime from the *swapSFXStructArray*, which is a buffer array of randomly selected files for the current region. When the user enters a new region, the SFX for the current region are gradually faded-out on the *MixerBus*. The *swapSFXStructArray* is loaded with the new region's synth pads and transient sounds, ready to be loaded in the *MixerBus*. When the fade-out is complete, the new files are loaded into the *MixerBus* and faded back in. The next eight channels are reserved for the

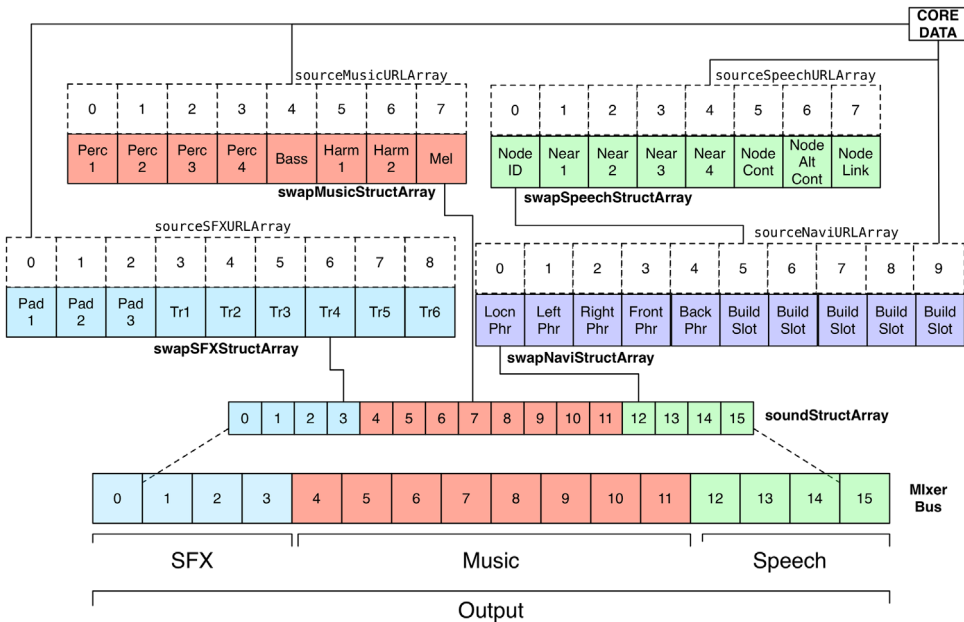


Figure 13. MixerBus logic diagram. The MixerBus provides 32 independently controllable channels of audio, which are used to configure speech, music and SFX.

construction of algorithmic music composed of melodies, rhythms and harmonics randomly selected and organized at runtime (Schoenberg 1967). Similar to the SFX, the *swapMusicStructArray* is a buffer array that feeds the *MixerBus* with audio files. The eight channels are carefully loaded when a region change occurs (fading in and out) and are algorithmically modified over time so that the music does not seem repetitive.

The last four channels are reserved for phrase synthesis where node, region, help or natural directions content is constructed at runtime. These four channels are the most active part in the overall process and require careful planning. Both the *swapSpeechStructArray* and the *swapNaviStructArray* use these channels for the production of phrases that will inform the user about their current node or region, give them directions to the next node or provide help. These buffers are managed using a prioritization matrix that only plays the audio with the highest priority first and queues the rest. The values that control this are assigned at the authoring stage of the process. One of the most significant features of the *MixerBus* is the ability to manage each channel or set of channels independently, which supports the dynamic loading of audio separately for speech content, music and background SFX.

This section described the main features of the *MobiLAudio* platform, how they function in isolation and how they integrate to deliver an immersive user experience using multimodal interfaces. The main component of the platform is the AUI, which can only operate with the support of the preceding survey and authoring process stages. Considerable contributions to the state-of-the-art in the area of mobile multimodal content delivery have been presented, next they are validated in the form of a user trial.

The following section details the user trial carried out to answer the research questions presented in Section 3, and to assess the effectiveness of the approach taken. Details of the trial are given, followed by the results of functionality and usability tests and concluded by a discussion of the results.

4. User trial

A central goal of this research was to test the usability and user experience of the *MobiLAudio* prototype. This was achieved in the form of a detailed user trial to determine specific user responses to the functionality, usability and content provided by the iOS audio tour developed. The trial focused extensively on user feedback to determine which elements of the proposed functionality are most important to the end user. Specific levels of detail were of interest (e.g. no music or SFX, voice only, etc.) alongside additional functions utilizing a variety of media to try and accurately model the user experience.

The user trial was carried out in central Dublin based around the theme of James Joyce's *Ulysses* and its central character Leopold Bloom. The tour tested the functionality of the survey and authoring tools to determine how effectively multiple tours could be produced. A total of 31 participants took part in the trial, and all participants completed the trial fully (Figure 14).

The trial focused on the execution of several location-based tasks, which were evaluated as part of functionality testing. Quantitative questions relating to the user experience of the application were given alongside post-test NASA Task Load Index (TLX) tests to assess how much of a cognitive load the application places on the user. The testing schedule defined three separate test areas; functionality, user experience and post-test TLX evaluation. The

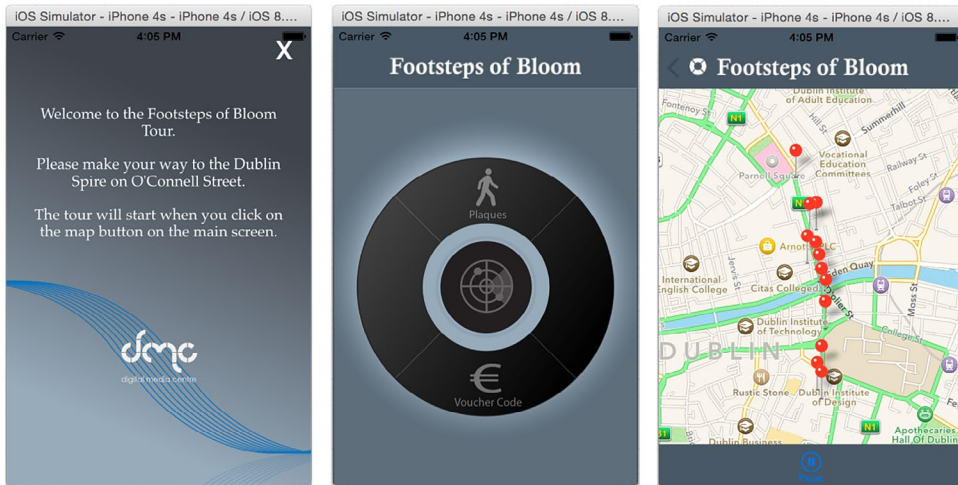


Figure 14. Bloom tour graphical user interface.

Note: This interface was used for the user trial in Dublin City centre for use in the event that manual control was necessary.

results provided both quantitative (functionality and TLX) and qualitative (user experience) data for analysis. This combination provided a comprehensive picture of the performance of the application, within the defined testing space of the *Footsteps of Bloom* audio tour.

Functionality Testing: After using the application, participants were asked a series of questions relating to the performance and functionality of the application. These included questions on the music, SFX, speech and directions provided, alongside the general operation of the application on the device.

User Experience Testing: After using the application, each participant was asked to evaluate the experience by answering a short set of qualitative questions. Questions included ease of use, evaluation of audio tours, and whether the app is worth paying for.

NASA TLX Testing: After the user experience testing was completed, a brief TLX test was carried out. The smallest TLX test contains a series of five scales that can be marked by either the tester or participant relating to metrics such as mental, physical and temporal demand. These metrics allow the participant to express how much cognitive demand the app placed upon them, which is crucial to the analysis of both functionality and user experience results as an increased cognitive load can negatively affect user experience.

4.1. *Footsteps of bloom* tour

The *Footsteps of Bloom* tour was constructed specifically for the purposes of the user trial and commercial evaluation and validation. The tour comprises 10 points of interest on a linear narrative beginning at the O'Connell Street Spire and ending at Davy Byrne's Pub on Grafton Street. The tour points are based on various sidewalk plaques found in central Dublin that punctuate Leopold Bloom's journey in *Ulysses*, and provides information based on the novel at each point. The tour takes approximately 35–40 min to complete and utilizes all the major functional elements of the AUI.

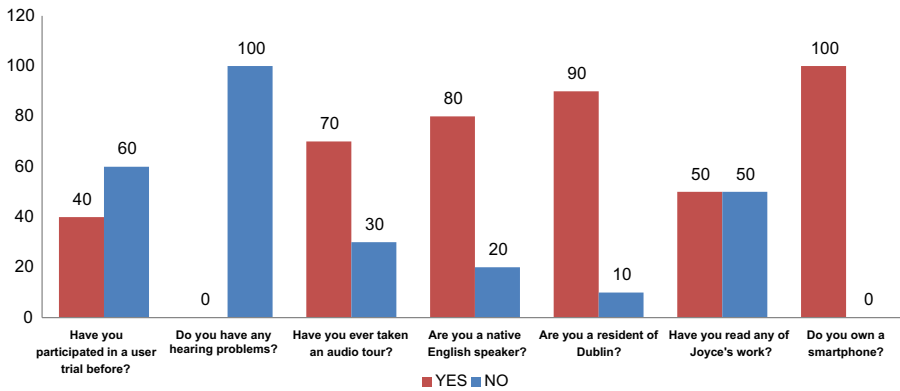


Figure 15. MobilAudio user trial pre-test question results.

Note: The table shows answers to yes/no questions relating to the user and their prior experience.

4.1.1. Pre-test questions

Each participant was asked several pre-test questions to determine whether they had any hearing problems (which would invalidate their test) and to obtain general information about their language skills, familiarity with the subject matter and their familiarity with the technologies involved. The results of each question are shown in Figure 15.

Of the participants questioned, none had any hearing difficulties that would preclude them from taking part in the trial. A total of 70% had previously taken an audio tour, which was considered useful in providing a benchmark of familiarity for the application. Most participants were native speakers (80%) living in Dublin (90%) though only 50% were familiar with the work of James Joyce. All participants own a smartphone, and as the test was designed to be largely hands free this was considered to be a sufficient level of prior exposure to the technologies involved in the tour.

4.1.2. Functionality questions

Each participant was asked a series of questions relating to the functionality of the application. These questions related to the SFX, music, speech and directions modelling provided as part of the *Footsteps of Bloom* tour. Questions were also included to determine the general operation of the application, notably its performance as a hands-free audio tour (Figure 16).

All participants could distinguish between the music, SFX and speech within the application and considered the balance between the three to be correct. Similarly, they considered the application to have functioned automatically and as required.

4.1.3. Directions modelling

Only 65% of participants felt the application had directed them correctly, though in many cases this was due to physical occlusion of several Leopold Bloom sidewalk plaques (e.g. due to outside tables at Davy Byrne's Pub, or telecom repair work on Aston Quay for 2 days of the trial). Of those who were misdirected in another way, two participants, after given the first instruction to proceed down O'Connell Street towards the bridge (from the Dublin Spire), walked instead in the opposite direction towards Parnell Street. In another three instances, participants struggled to find a plaque on Westmoreland Street because the directions to

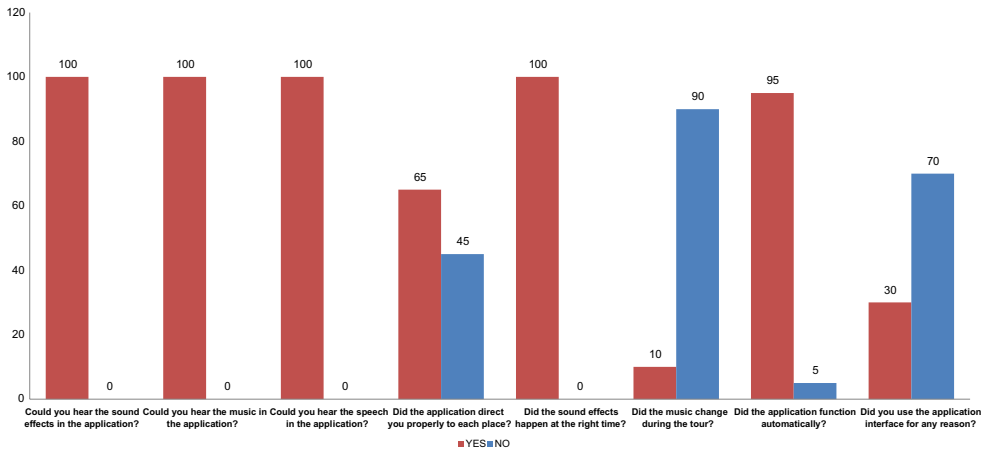


Figure 16. MobiLAudio user trial functionality question results.

Note: The table shows answers to yes/no questions relating to the operation of the application and its component functionality.

Conways Pub were not strictly accurate (as the sidewalk sign is actually outside Charlie's Chinese Takeaway).

The most significant directional issues occurred at the second POI on Middle Abbey Street, as there were several components making up the navigation directions from Joyce's Statue on Earl Street. Many participants reported having difficulty keeping up with the tour whilst also trying to remember the series of steps needed to get to the Easons building. Some participant comments on the application (taken post-test) included:

... directions were a great idea, but difficult in places ...

... (I) had only one chance to listen to the directions ...

Staged directions would have been a lot more useful ...

... Street names are no good for tourists ... landmarks are better ...

... a wrong turning sound would have been helpful around Easons ... otherwise it was grand

This feedback suggests that the directions component, though considered very useful by participants, did not function to the level of accuracy and detail that they would have liked. Future work will have to consider directions modelling not just in terms of node/hub networks (as performed in *Delivery Engine*) but also in terms of the complexity of the route. In addition, the use of street names in tandem with visual landmarks proved not to provide any useful additional information, and indeed was considered slightly confusing in some cases.

4.1.4. AUI performance

All participants indicated that they could hear sound effects (SFX), music and speech correctly, and in particular the SFX were considered to be a really good addition to the experience. All users considered the timing of the SFX to be correct (100%) and again this was backed up by qualitative comments indicating that SFX were a significant positive element of the tour.

The determination of all three audio elements (SFX, music, speech) validates the design of the listening mode model employed by the application, but there was an issue indicated

with repetition of music during the tour. Only 10% of participants felt the music changed or varied during the tour, with many commenting that the music was either too much or too distracting:

... voice could be a bit louder relative to the music ...

I didn't like the music, it didn't fit in with Joyce ...

This indicates that the use of immersive background music may not be as important or necessary as was initially thought, with listeners indicating a preference for a music 'mute' button in the application. Indeed, the use of the application interface (participants were told the application was automatic) by 70% of participants was mainly to change the volume between POIs, with only two participants stating they had used the application's map out of personal interest in the tour. Interestingly, the expectation of traffic noise as a distraction to users did not arise. This may have been due to the level of the background music or the overall immersion experienced by users.

4.1.5. User experience questions

Each participant was also asked a series of qualitative questions relating to the user experience of the *Footsteps of Bloom* tour. These questions focused on the engagement of the application, with questions on user tours and social media integration being included alongside the opportunity to make more general comments (Figure 17).

All participants found the application easy to use and most found it engaging (95%) and would be interested in other similar tours (95%). Comments on the general concept and content of the tour were positive, with many participants providing suggestions for other tours:

... easy to use, liked the directions between points ...

... liked the voice and the sounds, the barge going past sounded real!

... storytelling was great, would have liked more content though ...

... could you have running tours ... or cycling tours ...

An Easter Rising tour would be brilliant with the bombs going off around you!

One of the more significant comments made here aside from the high level of engagement experienced was that of the changes to the sound stage. Users reported on a noticeable change in the spatial image due to the use of sound effects (barges passing by and trams in the distance), with some reporting that they sounded 'very real'.

4.1.6. Tour pricing

Ninety per cent of participants said they would pay to use a mobile tour application, and gave an average price value of €2.95 for the amount they would be willing to pay. This figure is interesting, as 90% of the participants were Dublin residents (with 80% being native English speakers) who would have had some previous exposure to elements of the tour, its concept and content. Additionally, though only 50% said they had read Joyce's work, this does not seem to preclude interest in audio tours in general with 75% indicating they enjoyed the *Footsteps of Bloom* tour.

4.1.7. Social networking integration

One of the most interesting aspects of the results was the high values for social networking integration with tours (85%), alongside the possibility of user tour creation (90%). In more

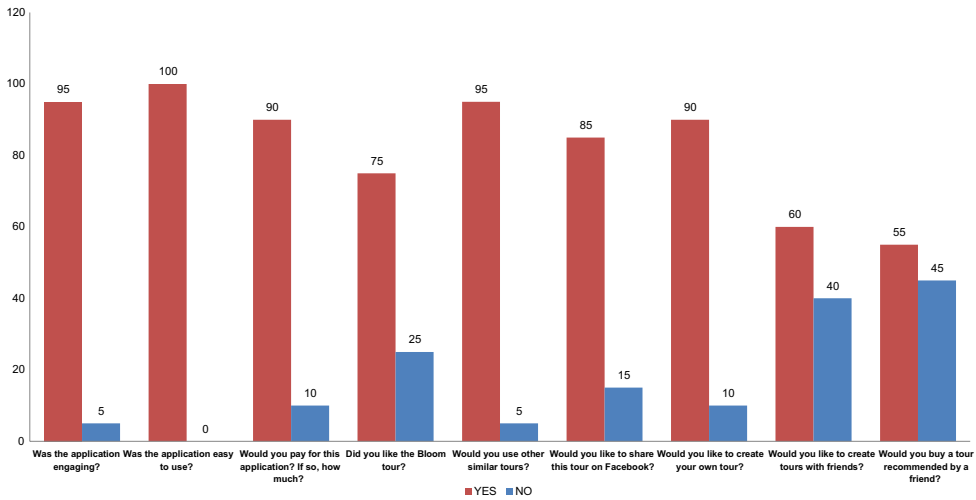


Figure 17. MobiLAudio user trial user experience question results.

Note: The table shows answers to yes/no questions relating to the engagement of the application, its potential for self-generated tours and social media integration opportunities.

detail, 60% considered the possibility of creating tours in a group to be interesting, though the social marketing potential of tour recommendations did not provide a definitive answer (55%). Comments about user tour creation included:

... I'd love to build a tour for visiting family, it would give them something to do while I'm at work during the day ...

... I'd build a tour for my boyfriend, linking all the places we've been since we met ...

... when friends come to visit it would be great to give them a pub tour ...

Is there a web interface for building the tours? I'd love to put one together ...

... my own voice would be nice for family, they would like the personal touch ...

4.1.8. Post test TLX results

After completing the tour and answering all test questions, a set of post-test NASA TLX tests were carried out to determine the load on the participant. NASA TLX tests were originally developed to evaluate astronaut performance and behaviour under demanding or stressful conditions. Under normal circumstances, the tests are useful in assessing the mental and physical load that a certain operation or task places on a participant and are a useful means of comparing how participants actually performed relative to how well they *thought* they performed (Figure 18).

The overall results were low (as desired), with physical demand and frustration having values of 22 and 20.5%, respectively, suggesting that the user experience results for engagement and ease of use (100%) were a fair reflection of the performance of the application. The important value of time pressure was also very low at 16.5%, suggesting that the application did not impinge upon the tour experience in its function or delivery. Mental demand was higher at 41% and this is potentially explained by the requirement for participants to listen to directions as they navigate the tour (no participant had any prior idea of the direction the

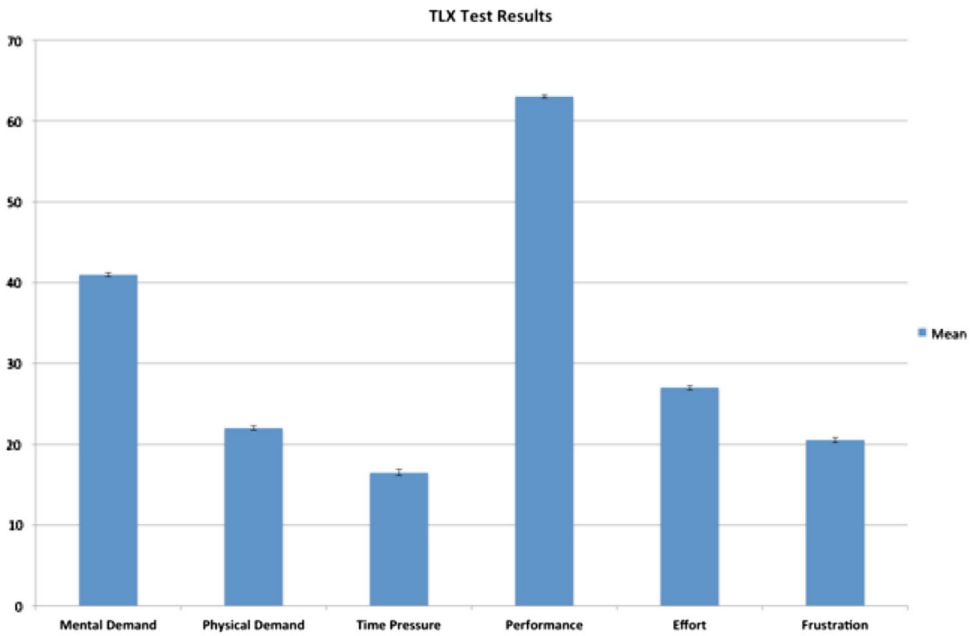


Figure 18. MobiLAudio user trial post-test TLX results.

Notes: The table shows the mean and standard deviation (in per cent) for each of the six TLX scales. TLX scales cover mental and physical demand, time pressure, performance, effort and frustration.

tour would take). The effort score of 27%, though still very low, can perhaps be attributed to the combination of walking and listening to a narrative at the same time. It is also interesting to consider the performance value of 63%, which may indicate in some respects difficulties with the directions modelling provided. In several instances, participants indicated that the directions were too long or confusing and expressed concern that they had not 'done well enough' even though they themselves were not directly being evaluated. It is also significant that percentile deviations were under 0.4% for all scales, indicating that the TLX values obtained were representative of the overall participant population.

4.1.9. Discussion

The overall reaction to the *Footsteps of Bloom* tour was very positive, and the technical elements of the application were all evaluated to function properly (though directional modelling did cause problems in its current delivery format). In particular, SFX was found to be a very positive addition to the experience of the application, though background music proved to be distracting for some users. The automated nature of the application was also considered a significant feature, with most people only using the graphical interface out of interest rather than necessity. The narrative and conversational flow of the tour was very well received, and the vast majority of participants would be interested in other audio tours. The concept of user-generated tours was strongly endorsed by participants, with integration into social networking being another positive element. Overall, participants reacted very positively to the tour and validated both the design and technical and directions taken in *MobiLAudio*.

The results of the Post Test TLX survey were mainly in line with expectations with significantly low scores for frustration, time pressure, physical demand and effort. The most interesting part of these results is the high value of 63% for performance. This could be attributed to the use of a combination of map-based visual directions and egocentric-based audio directions, which have been shown to cause confusion when users switch between frames of reference. Another reason could be that the audio directions were too long as some participants did express concern about the length and complexity of the directions and that they had not 'done well enough'.

5. Conclusions and future work

The *MobiLAudio* platform demonstrates significant advances in the area of multimodal content delivery. In addition to a traditional GUI, an AUI is used to deliver rich, interactive content using audio as the primary modality. The AUI comprises novel innovations in the areas of natural directions (providing directions using landmarks as opposed to cardinal directions) and content delivery using an intelligent content phrase synthesizer, controlled by a users location, orientation and previous visits to the space. In addition, an immersive audio experience is delivered using region detection and an algorithmic music sequencer. The result is an audio tour that provides timely information about an environment, 3D soundscaping that delivers immersive background audio and transient sounds based on location and natural directions between nodes on the tour. All of the content in the app is assembled on the fly with music, sound effects and speech audio being constructed dynamically using the audio sequencer and phrase synthesizer. This approach has proven to significantly reduce the amount of data redundancy in the system, enabling a far greater degree of efficiency overall.

Taking the results of the user trial into account, future work will focus on refining the delivery of directions by considering directions modelling not just in terms of node/hub networks but also in terms of the complexity of the route. In addition, the use of street names in tandem with visual landmarks proved not to provide any useful additional information, and was considered slightly confusing in some cases. Indeed, staged natural directions based on a user's current location are a more obvious choice in an urban environment.

Further testing will be carried out in order to fine-tune the variances in the listening model to determine a more optimum balance of audio in the AUI. In addition, an adaptive volume control based on external noise levels will be considered to alleviate the issues identified in the user trial around volume levels and algorithmic music.

Although the usability testing identified some interesting results regarding user engagement and level of immersion, this was somewhat inconclusive. Future usability testing will use the standardized system usability scale to determine the usability score for applications.

Extending the use of the authoring tool to include user-created tours will also be investigated with a view to testing the idea of social media integration enabling users to create, share and interact with users of a tour.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Abowd, G. D., C. G., Atkeson, J. Hong, S. Long, R. Kooper, and M. Pinkerton. 1997. "Cyberguide: A Mobile Context-aware Tour Guide." *Wireless Networks* 3 (5): 421–433.
- Absar, R., and C. Guastavino. 2008. "Usability of Non-speech Sounds in User Interfaces." *Proceedings of the 14th International Conference on Auditory Display (ICAD2008)*, June 24–27, 2008, Paris: International Community for Auditory Display.
- Apple Developer Library. 2014. "iOS – Core Audio Essentials." Accessed November 16. <https://developer.apple.com/library/ios/documentation/MusicAudio/Conceptual/CoreAudioOverview/CoreAudioEssentials/CoreAudioEssentials.html>
- Bartie, Phil J., and William A. Mackaness. 2006. "Development of a Speech-based Augmented Reality System to Support Exploration of Cityscape." *Transactions in GIS* 10 (1): 63–86. doi:10.1111/j.1467-9671.2006.00244.x.
- Baus, Jörg, Rainer, Wasinger, Ilhan, Aslan, Antonio, Krüger, Andreas, Maier and Tim Schwartz. 2007. "Auditory Perceptible Landmarks in Mobile Navigation." In *Proceedings of the 12th International Conference on Intelligent User Interfaces (IUI '07)*, 302–304. New York: ACM. doi:10.1145/1216295.1216352
- Beauchamp, J. W., ed. 2007. *Analysis, Synthesis, and Perception of Musical Sounds*. New York: Springer-Verlag .
- Bederson, B.B. 1995. "Audio augmented reality: a prototype automated tour guide." In *Conference Companion on Human Factors in Computing Systems (CHI '95)*, edited by Katz, I., R. Mack, and L., Marks, 210–211. New York: ACM. 10.1145/223355.223526
- Bellotti, F., C. Berta, A. de Gloria, M. Margarone. 2002. "User Testing a Hypermedia Tour Guide." *IEEE Pervasive Computing* 1 (2): 33–41.
- Benedikt, M. L. 1979. "To Take Hold of Space: Isovists and Isovist Fields." *Environment and Planning B: Planning and Design* 6 (1): 47–65.
- Burnett, G. E. 2000. "'Turn Right at the Traffic Lights': The Requirement for Landmarks in Vehicle Navigation Systems." *Journal of Navigation* 53 (3): 499–510.
- Carswell, J. D., A. Eustace, K. Gardiner, E. Kilfeather, and M. Neumann. 2002. "An Environment for Mobile Context-based Hypermedia Retrieval." *13th International Conference on Database and Expert Systems Applications (DEXA)*, IEEE CS Press, Aix en Provence, 532–536.
- Carswell, J., K. Gardiner, and M. Neumann. 2003. "Wireless Spatio-Semantic Transactions on Multimedia Datasets." 19th Annual ACM Symposium on Applied Computing, Nicosia, Cyprus.
- Carswell, J. D., K. Gardiner, A. Rizzini, M. Bertolotto, N. Mandrak. 2008. "A Web-based and Mobile Environmental Management System." *Journal of Environmental Informatics* 12 (1): 9–20.
- Carswell, J. D., K. Gardiner, and J. Yin. 2010a. "Mobile Visibility Querying for LBS." *Transactions in GIS* 14 (6): 791–809.
- Carswell, J. D., K. Gardiner, and J. Yin. 2010b. "3DQ: Threat Dome Visibility Querying on Mobile Devices." *GIM International* 24 (8): 24–26.
- Carswell, J. D., K. Gardiner, A. Rizzini, M. Bertolotto, and N. Mandrak. 2010. "A Web and Mobile System for Environmental Decision Support." In *Decision Support Systems in Agriculture, Food and the Environment: Trends, Applications and Advances*, edited by Basil Manos, Konstantinos Paparrizos, Nikolaos Matsatsinis, and Jason Papatthasiou, 317–338. Hershey, PA: IGI Global.
- Chan, L.-W., Y.-Y. Hsu, Y.-P. Hung, and J. Y.-J. Hsu. 2005. "Orientation-Aware Handhelds for Panorama-Based Museum Guiding System." *UbiComp 2005 Workshop: Smart Environments and their Applications to Cultural Heritage*, September 11–14, Tokyo, Japan.
- Chittaro, L. 2010. "Distinctive Aspects of Mobile Interaction and Their Implications For The Design of Multimodal Interfaces." *Journal on Multimodal User Interfaces* 3 (3): 157–165.
- Chittaro, L., and S. Burigat. 2005. "Augmenting Audio Messages with Visual Directions in Mobile Guides: An Evaluation of Three Approaches." In *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services*, 107–114. New York: ACM.
- Darwin, C. J. 2008. "Listening to Speech in the Presence of Other Sounds." *Philosophical Transactions of the Royal Society B: Biological Sciences* 363 (1493): 1011–1021. doi:10.1098/rstb.2007.2156.
- Donker, Hilko, Palle Klante, and Peter Gorny. 2002. "The Design of Auditory User Interfaces for Blind Users." In *Proceedings of the Second Nordic Conference on Human-Computer Interaction (NordCHI '02)*, 149–156. New York, NY: ACM.

- Droumeva, Milena. 2005. "Understanding Immersive Audio: A Historical and Socio-Cultural Exploration of Auditory Displays." Eleventh Meeting of the International Conference on Auditory Display (ICAD'05), July 6-9, Limerick, Ireland.
- Eckel, G. 2001. "Immersive Audio-Augmented Environments: The LISTEN Project." In *Proceedings Fifth International Conference on Information Visualisation* 571–573. IEEE. doi:10.1109/IV.2001.942112.
- Edwards, Alistair. 1989. "Soundtrack: An Auditory Interface for Blind Users." *Human-Computer Interaction* 4: 45–66. doi:10.1207/s15327051hci0401_2.
- Fröhlich, P., R. Simon, L. Baillie and H. Anegg. 2006. "Comparing Conceptual Designs for Mobile Access to Geo-Spatial Information." 8th International Conference on Human- Computer Interaction with Mobile Devices and Services (MobileHCI 06), Helsinki, Finland.
- Fröhlich, P., R. Simon, and L. Baillie. 2009. "Mobile Spatial Interaction." *Personal Ubiquitous Computing* 13 (4, May 2009): 251–253. doi:10.1007/s00779-008-0208-1.
- Gardiner, K. 2004. "A Framework for Processing Viewer-based Directional Queries in a Simulated Mobile Environment." Masters diss., Dublin Institute of Technology.
- Gardiner, K., and J. D. Carswell. 2003. "Viewer-Based Directional Querying for Mobile Applications." In *Proceedings of the Fourth International Conference on Web Information Systems Engineering Workshops*, edited by Giuseppe Santucci, Wolfgang Klas, Michela Bertolotto, Coral Calero and Luciano Baresi, 83–91. Washington, DC: IEEE Computer Society.
- Gardiner, K., A. Rizzini, J. D. Carswell, and M. Bertolotto. 2005. "MEMS – Mobile Environmental Management System" 13th Annual Conference on GIS Research UK, University of Glasgow, Scotland.
- Gardiner, K., J. Yin, and J. D. Carswell. 2009. "EgoViz –A Mobile Based Spatial Interaction System." In *9th International Symposium on Web and Wireless Geographical Information Systems (W2GIS)*, LNCS 5886, edited by James D. Carswell, A. Stewart Fotheringham, and Gavin McArdle, 135–152. Berlin Heidelberg: Springer. doi:10.1007/978-3-642-10601-9_10
- Gardiner, K., C. Cullen and J. D. Carswell. 2015. "MobiSurround: An Auditory User Interface for Geo Service Delivery." In *26th International Conference on Database and Expert Systems Applications (DEXA)*, Part II, LNCS 9262, edited by Q. Chen, A. Hameurlain, F. Toumani, R. Wagner, and H. Decker, 57–72. Cham: Springer International Publishing.
- Gaver, William. 1989. "The SonicFinder: An Interface That Uses Auditory Icons." *Human-Computer Interaction* 4: 67–94. doi:10.1207/s15327051hci0401_3.
- Heller, Florian, Thomas Knott, Malte Weiss, and Jan Borchers. 2009. "Multi-user Interaction in Virtual Audio Spaces." In *Extended Abstracts on Human Factors in Computing Systems (CHI EA '09)*, 4489–4494. New York, NY: ACM.
- Hile, H., R. Vedantham, G. Cuellar, A. Liu, N. Gelfand, R. Grzeszczuk, and G. Borriello. 2008. "Landmark-based Pedestrian Navigation from Collections of Geotagged Photos." In *Proceedings of the 7th International Conference on Mobile and Ubiquitous Multimedia (MUM '08)*, 145–152. New York, NY: ACM.
- Holland, Simon, David R. Morse, and Henrik Gedenryd. 2002. "AudioGPS: Spatial Audio Navigation with a Minimal Attention Interface." *Personal and Ubiquitous Computing* 6 (4): 253–259. doi: 10.1007/s007790200025.
- Jagdish, Deepak, Rahul Sawhney, Mohit Gupta, and Shreyas Nangia. 2008. "Sonic Grid: An Auditory Interface for the Visually Impaired to Navigate GUI-based Environments." In *Proceedings of the 13th International Conference on Intelligent User Interfaces (IUI '08)*, 337–340. New York, NY: ACM.
- Kainulainen, A., M. Turunen, J. Hakulinen, and A. Melto. 2007. "Soundmarks in Spoken Route Guidance." Proceedings of the 13th International Conference of Auditory Display (ICAD 2007), June 26-29, Montreal, Canada.
- Kilfeather, E., J. Carswell, K. Gardiner, and S. Rooney. 2007. "Urban Location Based Services Using Mobile Clients: The ICiNG Approach." GISRUK, Maynooth, Ireland.
- Klippel, A., H. Tappe, L. Kulik, and P. U. Lee. 2005. "Wayfinding Choremes—A Language for Modeling Conceptual Route Knowledge." *Journal of Visual Languages & Computing* 16 (4): 311–329.
- Kramer, Gregory. 1993. *Auditory Display: Sonification, Audification, and Auditory Interfaces*. New York: Perseus Publishing.
- Lalanne, Denis, Laurence Nigay, Philippe Palanque, Peter Robinson, Jean Vanderdonck, and Jean-Francois Ladry. 2009. "Fusion Engines for Multimodal Input: A Survey." In *Proceedings of the 2009 International Conference on Multimodal Interfaces (ICMI-MLMI '09)*, 153–160. New York, NY: ACM.

- Magnusson, Charlotte, Konrad Tollmar, Stephen Brewster, Tapani Sarjakoski, Tiina Sarjakoski, and Samuel Roselier. 2009. "Exploring Future Challenges for Haptic, Audio and Visual Interfaces for Mobile Maps and Location Based Services." In *Proceedings of the 2nd International Workshop on Location and the Web (LOCWEB '09)*, Article 8. New York, NY: ACM. doi: [10.1145/1507136.1507144](https://doi.org/10.1145/1507136.1507144).
- McGee, J., and C. Cullen. 2009a. "Vocate: Auditory Interfaces for the LOK8 Project." Proceedings of the 9th IT & T Conference, October 22-23, Dublin Institute of Technology, Dublin.
- McGee, J., and C. Cullen. 2009b. "Vocate: Auditory Interfaces for Location-Based Services." Physicality Workshop at 23rd Conference on Computer Human Interaction (HCI 2009), September 23, Cambridge, UK.
- McGookin, D. 2004. "Understanding and Improving the Identification of Concurrently Presented Earcons." PhD diss., University of Glasgow, 155–159.
- Millonig, A., and K. Schechtner. 2005. "Developing Landmark-based Pedestrian Navigation Systems." *IEEE Transactions on Intelligent Transportation Systems* 8(1): 43–49. doi:[10.1109/TITS.2006.889439](https://doi.org/10.1109/TITS.2006.889439)
- Misra, Ananya, R. Perry Cook, and Ge Wang. 2006. "A New Paradigm for Sound Design." Proceedings of the International Conference on Digital Audio Effects (DAFx-06), September 18-20, Montreal, Quebec, Canada.
- Modsching, M., R. Kramer, M. Riebeck, A. Stark, K. ten Hagen, and J. Kawalek. 2007. "Field Trial on the Efficiency and User Experience of GPS Based State of the Art Navigational Systems for Pedestrians." In *4th Workshop on Positioning, Navigation and Communication (WPNC '07)*, 129–134. IEEE. doi: [10.1109/WPNC.2007.353623](https://doi.org/10.1109/WPNC.2007.353623)
- Mynatt, Elizabeth D. 1990. "Auditory Presentation of Graphical User Interfaces." *Auditory Display: Sonification, Audification and Auditory Interfaces*, edited by G. Kramer, 533–556. Reading, MA: Addison-Wesley.
- Neustein, A., and J. Markowitz, eds. 2013. *Mobile Speech and Advanced Natural Language Solutions*. New York, NY: Springer. [10.1007/978-1-4614-6018-3](https://doi.org/10.1007/978-1-4614-6018-3)
- Persson, P., F. Espinoza, P. Fagerberg, A. Sandin, and R. Cöster. 2002. "GeoNotes: A Location-based Information System for Public Spaces." In *Designing Information Spaces: The Social Navigation Approach, Part I*, edited by Kristina Höök, David Benyon, and Alan J. Munro, 151–173. London: Springer.
- Potamianos, A., and M. Perakakis. 2008. "Multimodal Processing and Interaction." *Multimedia Systems and Applications Series* 33: 1–18. doi:[10.1007/978-0-387-76316-3](https://doi.org/10.1007/978-0-387-76316-3).
- Rehrl, K., E. Häusler, and S. Leitinger. 2010. "Comparing the Effectiveness of GPS-Enhanced Voice Guidance for Pedestrians with Metric-and Landmark-Based Instruction Sets." In *Geographic Information Science*, 189–203. Berlin Heidelberg: Springer.
- Rehrl, K., E. Häusler, S. Leitinger, D. Bell. 2014. "Pedestrian Navigation with Augmented Reality, Voice and Digital Map: Final Results from an *in Situ* Field Study Assessing Performance and User Experience." *Journal of Location Based Services* 8 (2): 75–96.
- Rizzini, A., K. Gardiner, M. Bertolotto, and J. D. Carswell. 2006. "MEMS Mobile GIS: A Spatially Enabled Fish Habitat Management System." 6th International Symposium on Web and Wireless Geographical Information Systems (W2GIS); Springer LNCS vol.4295; Hong Kong, China.
- Robinson, S., P. Eslambolchilar, and M. Jones. 2008. "Point-to-GeoBlog: Gestures and Sensors to Support User Generated Content Creation." Proceedings of the 10th Conference on Human-Computer Interaction with Mobile Devices and Services, Mobile HCI 2008, September 2-5, Amsterdam, The Netherlands. doi: [10.1145/1409240.1409262](https://doi.org/10.1145/1409240.1409262)
- Rooney, S., K. Gardiner, and J. D. Carswell. 2007. "An Open Source Approach to Wireless Positioning Techniques." In Proceedings of the 5th International Symposium on Mobile Mapping Technology (MMT), Padua, Italy.
- Schoenberg, A. 1967. *Fundamentals of Music Composition*. Boston, MA: Faber and Faber.
- Schwarz, D. 2004. "Data-driven Concatenative Sound Synthesis." PhD Thesis, Ircam Centre Pompidou, Paris, France.
- Schwarz, D. 2005. "Current Research in Concatenative Sound Synthesis." Proceedings of the International Computer Music Conference (ICMC), September 5-9, Barcelona, Spain.

- Shi, X., T. Sun, Y. Shen, K. Li, and W. Qu. 2010. "Tour-guide: Providing Location-Based Tourist Information on Mobile Phones." In *Computer and Information Technology (CIT), 2010 IEEE 10th International Conference on*, 2397–2401. IEEE.
- Shinohara, Kristen, and Josh Tenenber. 2009. "A Blind Person's Interactions with Technology." *Communications of the ACM* 52(8): 58–66. doi:10.1145/1536616.1536636
- Shneiderman, S. B., and C. Plaisant. 2005. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. 4th ed. Boston: Pearson Addison Wesley.
- Simon, R., and P. Fröhlich. 2007. "A Mobile Application Framework for the Geospatial Web." Proceedings of the 16th international conference on World Wide Web, Banff, Alberta, Canada.
- Simon, R., and P. Fröhlich. 2007. "The Point to Discover GeoWand." 9th International Conference on Ubiquitous Computing (UbiComp 07), Innsbruck, Austria.
- Simon, R., P. Fröhlich, and T. Grechenig. 2008. "GeoPointing: Evaluating the Performance of Orientation-Aware Location-Based Interaction under Real-World Conditions." *Journal of Location Based Services* 2 (1): 24–40.
- Stahl, Christoph. 2007. "The Roaring Navigator: A Group Guide for the Zoo with Shared Auditory Landmark Display." In *Proceedings of the 9th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '07)*, 383–386. New York, NY: ACM.
- Stark, A., M. Riebeck, and J. Kawalek. 2007. "How to Design an Advanced Pedestrian Navigation System: Field Trial Results." In *4th IEEE Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS 2007)*, 690–694. IEEE. doi:10.1109/IDAACS.2007.4488511
- Strachan, S., R. Murray-Smith. 2009. "Bearing-based Selection in Mobile Spatial Interaction." *Personal and Ubiquitous Computing* 13 (4): 265–280.
- Strachan, Steven, Parisa Eslambolchilar, Roderick Murray-Smith, Stephen Hughes, and Sile O'Modhrain. 2005. "GpsTunes: Controlling Navigation via Audio Feedback." Proceedings of the 7th international conference on Human computer interaction with mobile devices & services, Salzburg, Austria.
- Strachan, S., J. Williamson, and R. Murray-Smith. 2007. "Show Me the Way to Monte Carlo: Density-based Trajectory Navigation." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*, 1245–1248. New York, NY: ACM. doi:10.1145/1240624.1240812
- Truax, B. 2001. *Acoustic Communication*. 2nd ed. Westport, CT: Ablex Publishing.
- Turk, Matthew. 2014. "Multimodal Interaction: A Review." *Pattern Recognition Letters* 36: 189–195. doi: 10.1016/j.patrec.2013.07.003/.
- Vazquez-Alvarez, Yolanda, Ian Oakley, Stephen A. Brewster. 2012. "Auditory Display Design for Exploration in Mobile Audio-Augmented Reality." *Personal and Ubiquitous Computing* 16: 987–999. doi: 10.1007/s00779-011-0459-0.
- Wakkary, R, K Newby, M Hatala, D Evernden, and M Droumeva. 2004. "Interactive Audio Content : An Approach to Audio Content for a Dynamic Museum Experience through Augmented Audio Reality and Adaptive Information Retrieval." In *Museums and the Web 2004 Selected Papers*, edited by D. Bearman and J. Trant, 55–60. Arlington, VA: Archives & Museum Informatics.
- Walker, Bruce N., and Jeffrey Lindsay. 2006. "Navigation Performance with a Virtual Auditory Display: Effects of Beacon Sound, Capture Radius, and Practice." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 48: 265–278. doi:10.1518/001872006777724507.
- Wang, Lei, Paul Roe, Binh Pham, and Dian Tjondronegoro. 2008. "An Audio Wiki Supporting Mobile Collaboration." In *Proceedings of the 2008 ACM Symposium on Applied Computing (SAC '08)*, 1889–1896. New York, NY: ACM. doi:10.1145/1363686.1364145. <http://doi.acm.org/10.1145/1363686.1364145>
- Wenzel, Elizabeth M. 1994. "Spatial Sound and Sonification." In *Auditory Display: Sonification, Audification, and Auditory Interfaces*, edited by G. Kramer, 127–150. Reading, MA: Addison Wesley.
- Williamson, J., R. Murray-Smith, and S. Hughes. 2007. "Shoogle: Excitatory Multimodal Interaction on Mobile Devices." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 121–124. New York, NY: ACM.
- Wilson, A., and S. Shafer. 2003. "XWand: UI for Intelligent Spaces." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '03.*, 545–552. New York, NY: ACM.
- Yin, Junjun, and J. D. Carswell. 2013. "Spatial Search Techniques for Mobile 3D Queries in Sensor Web Environments." *ISPRS International Journal of Geo-Information* 2 (1): 135–154.