![Queen Mary University of London]

Wayfinding without Visual Cues: Evaluation of an Interactive Audio Map System

STOCKMAN, AG

Wayfinding without Visual Cues: Evaluation of an Interactive Audio Map System

ESTHER LOELIGER[1] AND TONY STOCKMAN[2]

[1]Algorithms Group
School of Computer Science
University of Hertfordshire
Hatfield, AL10 9AB, Herts, UK
[2]Cognitive Science Research Group and Centre for Digital Music
School of Electronic Engineering and Computer Science
Queen Mary, University of London
Mile End Road, E1 4NS, London, UK

We present the evaluation of an interactive audio map system that enables blind and partially sighted users to explore and navigate city maps from the safety of their home using 3D audio and synthetic speech alone. We begin with a review of existing literature in the areas of spatial knowledge and wayfinding, auditory displays, and auditory map systems, before describing how this research builds on and differentiates itself from this body of work. One key requirement was the ability to quantify the effectiveness of the audio map, so we describe the design and implementation of the evaluation, which took the form of a game downloaded by participants to their own computers. The results demonstrate that participants (blind, partially sighted and sighted) have acquired detailed spatial knowledge and also that the availability of 3D audio cues significantly improves wayfinding performance.

1. INTRODUCTION

This project began with the idea of an audio map that allows blind and partially sighted users to explore city maps from the safety of their home or office. Having explored oneŠs urban environment using a virtual map, further exploration and wayfinding should be cognitively less demanding, enabling the individual to focus on locomotion and potential hazards. Getting real users to engage with such a virtual audio map would also facilitate research into how audio maps can be made more effective. It would reveal and capture what works and what does not. In short, it would pave the way for more usable implementations of audio maps.

In terms of preparing the evaluation, it quickly became clear that data gathering had to focus on a narrowly defined question to yield quantitative, usable test results. In order to create a test environment that is fully accessible to both blind and sighted users, the evaluation compares one kind of audio map with another. It weighs up the relative effectiveness of the components that make up these maps: 3D sound, synthetic speech and sounds representing steps, turns, collisions, and so on.

The evaluation presented participants with a series of wayfinding challenges, which were chosen for two reasons. First, trying to reach a destination from a known location is a core application of maps. One can judge the effectiveness of a map by deciding on a route and following it, noting any unforeseen delays or wrong turns. Second, wayfinding challenges permit detailed quantitative as well as qualitative analysis. Every action was recorded and

timed, allowing each journey to be reconstructed and analysed. The evaluation was open for a period of just over a month and yielded extensive data logs, doodles, route descriptions and quantitative feedback. The results demonstrate very clearly that the audio map helped participants acquire spatial knowledge and also that 3D audio based on generic samples of traffic, birdsong, trains, shops, and so on significantly improved the ability of participants to complete the challenges. The software is released under an open source license and available from SourceForge.net at `http://team.sourceforge.net` (v. 4.0 released February 2013).

2. PREVIOUS WORK AND ITS INFLUENCE ON OUR DESIGN RATIONALE

2.1. Spatial knowledge and wayfinding

Much of the terminology used in this research including the term 'wayfinding' is rooted in Lynch (1960). Downs and Stea (1977) subdivide the process of wayfinding into four component parts: orientation, choice of route, keeping on track and the realisation that the destination has been reached. The evaluation described here focuses on the third, especially on the ways in which mapping tools can help the wayfinder keep on track. The first two, however, are ever present as participants pause to regain their bearings, especially after taking a wrong turn. In her study of gender differences in wayfinding strategies, Lawton (1994, 771) emphasises the split between 'orientation' and 'route' strategies, adding an additional scale she called the 'spatial anxiety scale'. (However, no attempt has been made to analyse the relative performance of male and female participants in the present evaluation.)

Montello (2005, 258–259) defines 'navigation' as a combination of 'locomotion' and 'wayfinding'. The former focuses on physical movement through the environment; the latter is goal-centred and thus requires a destination. As Montello observes, they have to do with different kinds of challenges. One requires us to use our senses and motor systems to avoid running into buildings or hazards. It requires attention to our 'proximal' environment. The other requires cognitive problem solving in 'distal' space as we choose suitably direct paths to our destination and calculate the journey time (Montello, 2005, 259). Wiener et al. (2009) have developed Montello's definition further by breaking up the wayfinding component into unaided and aided wayfinding. Unaided wayfinding may still be an undirected activity (the examples given are free exploration and pleasure walks), but path planning is only possible if (a) there is a specific destination, (b) one knows the route, and (c) one possesses 'survey knowledge' of the area (here defined as 'knowledge about the spatial relation of at least two points') (Wiener et al., 2009, 156).

According to Canada and Vinson (1999, 278), 'landmarks' are 'distinctive environmental features functioning as reference points', objects that stand out from a verbal route description. The kind of urban map on which this research is based fits Vinson's description of 'large-scale environments [. . .] that are so large that the navigator's viewpoint does not encompass the environment in its totality' (Canada and Vinson, 1999, 278). As Montello notes, these 'environmental spaces' are large 'relative to a person, more precisely, to a person's body and action (e.g., looking, walking)' (Montello, 1993, 315).

The challenge, then, is to build an audio map that enables users to become wayfinders in Wiener et al.'s sense of planning one's path through this environment, taking advantage of the most striking environmental features of the terrain.

The search for effective route directions is motivated in part by the fact that many – including most directions offered in good faith to passers-by – are not. Rinck (2005) observes that 'both the good descriptions and the skeletal ones yielded fewer direction errors, fewer hesitations, and fewer requests for assistance than the bad descriptions' (Rinck, 2005, 371). Rinck cites an article by Allen which reports that 'wayfinding was more successful if route directions were given in the correct temporal-spatial order, if they concentrated on information to choice points, and if they employed delimiters well known to most recipients' (Rinck, 2005, 371). The first point recalls Denis's observation that route directions contain temporal as well as spatial aspects: information about the environment of the route is spatial, whereas information about the actions that need to be taken is temporal (cited in Tschander et al. (2003, 91)).

In grouping 'good descriptions' with 'skeletal ones', Rinck also indicates that there is a lot to be gained from managing the cognitive load of instructions, that good route directions tend to be short or as short as possible. As Tversky points out, '[a] navigator doesn't need to remember where the highway exit or subway entrance is as the environment will mark them'; condensed instructions are likely to help recipients reach their destinations (Tversky, 2003, 6). Denis et al. agree that too much detail in route descriptions can be detrimental. They too recommend trimming them to a minimum (cited in Klippel et al. (2003, 17)). Where directions are too complex to be cut down to size, Klippel et al. propose 'chunking' directions according to (a) landmarks, (b) decision points and (c) a ranking of decision points based on 'change points' (decision points requiring a change of direction), 'uniqueness of a spatial structure' and 'processability' – a consideration memorably overlooked in the instruction to 'turn right at the 25th intersection' (Klippel et al., 2003, 21–22).

The process of selection that goes into the making of an effective route description is mirrored by an imperfect process of reorganisation on the part of the recipient. 'Essential as they are,' as Tversky puts it, 'perceptual organizing principles are guaranteed to produce error' (Tversky, 2003, 4). The situation is more complicated still after attempting to follow the instructions, as different views need to be integrated across modalities, mixing the original instructions with contextual information picked up along the way. 'Schematization', the attempt to achieve this integration 'through common elements and frames of reference' suits the limited capacity of working memory and helps individuals form mental representations of the route travelled (Tversky, 2003, 5).

Kuipers's work on knowledge of large-scale space in humans and robots led to what he called the 'spatial semantic hierarchy', an ontological hierarchy each level of which has different strengths and weaknesses, is based on different mathematical rules, and permits a distinct set of inference rules (Kuipers, 1994, 93). The spatial semantic hierarchy allows humans to cope with incomplete spatial knowledge and uncertainty in a robust way (Kuipers, 2000, 227). At the same time, it 'provides targets for opportunistic (hence inexpensive) assimilation of new knowledge and inference from existing knowledge' (Kuipers, 2000, 229). The theory draws on psychological studies that show, first, that 'a *topological* description of the environment is central to the cognitive map, and is logically prior to the metrical description' and, second, that 'spatial representation is grounded in the sensorimotor interaction between the agent and the environment' (Kuipers, 1994, 93). Accordingly, it is clear that the audio map should give precedence to topological relationships, not metrical measurements, and that it should be immersive.

Dede (2009) defines immersion as 'the subjective impression that one is participating in a comprehensive, realistic experience', encouraging the user's 'suspension of disbelief that she or he is "inside" a digitally enhanced setting' (Dede, 2009, 66). Dede's central argument is that digital immersion can enhance learning and 'transfer' of knowledge 'through simulation of the real world' (Dede, 2009, 67). In the immersive virtual setting of 'River City', for example, 'low-performing students do as well as their high-performing peers.' Dede contends that 'digital immersion allows these students to build confidence in their academic abilities by stepping out of their real-world identity of poor performer' (Dede, 2009, 67). Like River City, virtual audio maps seek to convey knowledge about the real world, creating a level playing field for sighted and blind participants.

There is a range of audio games that utilise the listeners' ability to orient themselves in 3D space using sound alone. The field has been examined in some detail in McCrindle and Symons (2000), Sánchez et al. (2001), Archambault et al. (2007), Archambault et al. (2008) and Oren et al. (2008). All games in this area have benefited hugely from advances in 3D audio quality developed for general-purpose games, but none of them makes the task of navigating a map the core challenge.

Parush et al. (2007) stress the importance of actively engaging the user in the wayfinding task. They observe that the use of GPS-based systems can lead to significant 'degradation in spatial knowledge acquisition' (Parush et al., 2007, 238) because it pushes humans 'out of the control loop' (Parush et al., 2007, 240). In particular, GPS-based devices can 'replace human perception by eliminating the need to gather information from the environment [...] replace human cognition by eliminating the need to integrate, comprehend the information, and process it [...] eliminate the need for wayfinding decision making and problem solving' (Parush et al., 2007, 240). The key problem is that although using a GPS device may lower the cognitive effort required, it also reduces the individual's ability to accomplish these tasks independently. For the purposes of this research it is clear that the audio map should force users to 'monitor and track their location in the environment' (Parush et al., 2007, 251) at all times.

2.2.   Auditory displays

Past work on auditory displays has focused on four distinctive techniques: auditory icons, earcons, actual (synthetic or pre-recorded) speech and spearcons (see Peres et al. (2008, 149–50) for an overview). This research makes extensive use of the first three but not spearcons – verbal cues played back at very high speeds – as they would have added learning effort for aspects that are well served by synthetic speech.

Gaver coined the term 'auditory icons' in a 1986 article of the same title: 'auditory icons are caricatures of naturally occurring sounds such as bumps, scrapes, or even files hitting mailboxes' (Gaver, 1986, 169). Gaver explored this emphasis on 'everyday listening' further in a 1989 paper on SonicFinder, an auditory interface which he developed for Apple (Gaver, 1989, 4). However, it has to be stressed that Gaver's immediate context (early graphical user interfaces) is far removed from auditory icons treated as elements in an auditory map. The sound of rustling paper reinforces the interaction of throwing a document in the trash, but not in the way the sound of a rumbling train alerts the user of a auditory map to the proximity of a railway crossing.

Blattner et al. (Blattner et al., 1989) developed the concept of 'abstract earcons': musical 'motives' that can be grouped to form 'families, where earcons with similar meanings have similar sounds'. The authors defined

earcon motives as 'a rhythmicized sequence of pitches' (Blattner et al., 1989, 23). The best length, they argued, is two to four notes as longer diatonic patterns would tend to form tunes and be likely to annoy users.

In 1992, Brewster et al. subjected the proposals of Blattner et al. to a detailed experimental investigation (Brewster et al. (1992); see also Brewster (1991)). The finding was that earcons do work, but also that Blattner et al. had been overly optimistic regarding the degree of differentiation and hierarchical organisation that is achievable by changing motives very slightly: participants were unable to tell one member of a given earcon family from another (Brewster et al., 1992, 496).

As Peres et al. (2008) have observed, many auditory interfaces call for a hybrid approach, employing 'complex mixtures of different techniques' (Peres et al., 2008, 150). This raises the issue of what the precise balance between speech, auditory icons and earcons should be. Each of the three has the potential to distract or annoy the user, so special care needs to be taken when recordings of actual sounds, earcon motives and synthetic speech coincide.

The capabilities of soundcards and audio libraries have shot up as 3D games and virtual worlds have entered mainstream computing. Mobile phones and handheld games consoles have capabilities that were beyond the reach of Gaver's Sonic Finder system implemented in 1989, working on state-of-the-art Apple Macintosh computers. Begault and Center. (2000) (a re-issue of a 1994 study) stands out as an authoritative guide to head-related transfer functions, distance cues, headphone distance errors, and reverberation in the context of 3D sound for virtual reality applications. Many of Begault's recommendations find a direct equivalent in the OpenAL specification (Hiebert et al., 2005) and the effects extension that accompanies it.

2.2.1. Auditory map systems

This work predated the release of the browser-based AccessibleMap (currently in open Beta; see Höckner et al. (2012) and http://www.accessible.map.at/), which would have offered an instructive point of comparison, not least because AccessibleMap too derives its core map and point of interest data from OpenStreetMap.

The BATS (Blind Audio Tactile Mapping System) project (Parente and Bishop, 2003) pioneered the use of auditory maps (albeit on the scale of whole countries rather than cities) for use by visually impaired people and so helped create a context for later studies. The researchers asked to what extent spatial sound can aid spatial understanding and learning, and they highlight two techniques in particular: 'spatial auditory icons' (e.g. bubbling water for rivers) and 'callouts' (spoken labels positioned in the auditory scene, e.g. 'Raleigh'). Using tactile trackballs, gamepads and keyboard commands,

users move a virtual listener across the map whose bearing never changes. The BATS developers were keen to ensure that the project should use 'standard computer hardware and open source software' (Parente and Bishop, 2003, 1). Haptic feedback devices should be 'consumer grade' and the authors stress that the Kensington, Logitech and Microsoft controllers used are readily available from consumer electronics stores.

It appears that BATS development was discontinued after the two prototypes presented in the paper. The first was written specifically for one individual (whose feedback was that the interface was somewhat complicated) and evaluation of the second was limited to 'a number of users, sighted and blind, informally interacting with the system' (Parente and Bishop, 2003, 5). The 'system for public release and plans for more rigorous evaluation' then in development – presumably this is related to the 'SonOfBATS' branch in the project's source repository – does not seem to have led to further publications (Parente and Bishop, 2003, 1).

(Heuten et al., 2007) developed an 'auditory torch' system that bears some resemblance to BATS, but one specialising in city maps. The user moves the listener across a map of Brussels using an 'absolute input device', a customized digitizer tablet with raised borders that correspond to the borders of the virtual map. Auditory torch 'walkthroughs' are not entirely immersive experiences in the sense of first-person navigation. Once more, users cannot change their bearing, for example. The project team conducted two formal evaluations, the first with eleven blind participants and the second with six untrained and presumably sighted participants. The results were impressive: participants were able to reproduce maps with considerable accuracy, that is, proportions and landmarks corresponded closely with the underlying map (Heuten et al., 2007, 1964). Heuten went on to describe a first person application of the auditory torch system in Heuten (2008).

The Audio-based Environment Simulator takes a very different approach. It seems that the project's final findings have not yet been published, and what is known about it comes from a work-in-progress paper published in April 2010 (Sánchez et al., 2010). Although there are obvious differences (this is a project about small-scale indoor spaces, not city maps), there are also some interesting parallels with the present research. Sánchez et al. argue that spatial learning of the kind they describe receives a significant boost when it takes place in the context of a game: 'The activity with the highest spatial representation scores was The Game [. . .], which resulted in much higher scores than those obtained for the other activities' (Sánchez et al., 2010, 3995). Once again, however, the expectation is that evaluations are conducted using a relatively small group of individuals (seven in
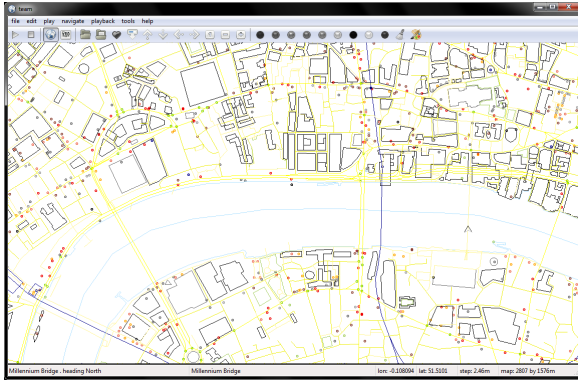
Figure 1. Screenshot

the case of the preliminary evaluation) in controlled circumstances. Here, the evaluation requires access to more than ten rooms, a 'tactile model' of the floor plan with Braille labels, the Audio-based Environment Simulator itself as well as LEGO and plasticine for modeling.

What this research took from these three projects is the need to maximise the number of potential participants by integrating training, evaluation, data logging and feedback in such a way that no supervision is required, and web distribution becomes possible. Paired with web distribution, this approach mitigates the risk of recruiting only small numbers of visually impaired participants in a limited geographical area. One further restriction of these projects was that maps had to be prepared especially for evaluations. By pulling in raw data feeds from the global OpenStreetMap database, this research encouraged participants to explore maps in their local area.

3. DESIGN

3.1. System design

As the literature reviewed in Section 2.1 suggests, the audio map presented here is predicated on an immersive, first-person experience of the map. The choice of perspective is deliberate and reflects the evaluation's emphasis on route knowledge. If the primary goal had been to impart survey knowledge a third-person perspective may have been more appropriate.

It is by design not a 'live' navigation aid: it does not use GPS but instead facilitates map exploration and route learning (with the computer keyboard used as the input device) in preparation of venturing outside. In other words, it is a 'virtual reality' application, not an 'augmented reality' tool (see Azuma (1997) for a definition of the latter). For a screenshot of a map see Figure 1.

3.1.1. Spatial knowledge and wayfinding

It is rarely possible to roam freely through a city. Instead, we tend to navigate a network of public roads and paths. Both modes are available here, but navigation is weighted in favour of roads and paths by ensuring that users do not inadvertently stray from a given path. In our virtual parks as in the real world, users are free to turn and walk as they wish, but as soon as they reach a path or street, movement is restricted to a network of ways and decision points.

The user's direction of travel is automatically adjusted as streets bend left or right. The sound that accompanies each step changes slightly when this happens to give audible feedback.

This means that it is possible to trace a wide oval shape simply by instructing the system to walk straight ahead. As Christopher D. Wickens and Yeh (2005, 391) argue, 'maps need not rotate continuously with the direction of travel, or may be rotated slightly more slowly than changes in the actual ego-orientation or movement, as long as the offset is never greater than 45 degrees.' When approaching sharp turns (that is, turns greater than 45 degrees), however, the user is asked to confirm the change of direction. At decision points, the user is given a range of options corresponding to the options available at that point in the physical environment.

All turns are 45-degree turns. This choice was supported by Klippel et al. (2004, 5), who demonstrate that humans distinguish between eight turn 'concepts' (four 'standard directions' plus four 'modified directions'). Klippel et al. further propose a slight adjustment where half-right turns centre on an angle just below 45 degrees, but here it seemed sensible to ensure that $360 \bmod n = 0$ and so guarantee that repeated turn actions will add up to a complete 360 degree turn.

Citing a 1996 study by Taylor and Tversky, Schneider and Taylor (1999, 416) note that individuals mix route and survey perspectives flexibly, stressing that other studies have corroborated Taylor and Tversky's research. We have used cardinal directions in order to help the player establish an absolute frame of reference and facilitate the acquisition of survey knowledge.

At any point, users are able to request the name of the street they are on, a rough absolute position on the map and their bearing. They can also trigger a sonar function, which gives the user an overview of the landmarks and points of interest in close proximity. Foregrounding connections between ways, landmarks and decision points helps users by reducing the number of salient features to be borne in mind. The near right angle at which the Millennium Bridge juts out northward from the South Bank is relevant here in a way that the round

shape of the reconstructed Globe Theatre is not. The number of distinct coordinates is greatly reduced.

To avoid sensory overload, the application allows users to filter out certain sound sources (e.g. restaurants or playgrounds). The volume of ambient sounds (e.g. market), non-ambient sounds (e.g. turn around) and synthetic speech can be controlled individually. However, these settings are fixed when taking part in the evaluation, as muting certain sound sources (e.g. train lines) would make particular challenges impossible to overcome.

The system eschews haptic input devices in favour of basic keyboard commands. In the absence of supervised training sessions on hardware set up for the purpose, the safest assumption one can make is that most users will own and have some experience using a keyboard. The essential consideration here is that keyboard commands should be as intuitive as possible and use large and/or easily reachable keys wherever possible. The cursor keys are clearly invaluable here as they serve to make walking and turning somewhat intuitive: cursor up is mapped to 'step ahead', cursor down to 'turn around' and right/left to 'turn right/left'. Space ('where am I?') and Return ('sonar') have the advantage of being large and prominently placed. Escape is a convenient choice for interrupting lengthy verbal descriptions and dismissing dialogs. The goal was to make the system as easy to learn as possible without unduly hampering the user's ability to navigate the map and so contribute to the evaluation.

One key feature of the system is that it allows the user to explore any location on the planet, from Alcatraz to Brooklyn, New York. For example, entering 'Camden' and selecting 'United Kingdom' in the download dialog will take the user straight to North London. From there, they can choose a start point and an end point from a self-voicing list. If a point is missing (e.g. a home address), it can be added in the map editor.

For the evaluation, it was necessary to limit the number of routes under investigation to ensure comparability of results, but it is worth stressing that the project is designed to help any visually impaired user navigate their own village, town or city. In essence, any location covered by the OpenStreetMap project is accessible to users of the system.

3.1.2. Auditory display

For the evaluation, users were encouraged to use headphones. It was possible to complete the tasks using appropriately placed loudspeakers at sufficient volume, however. To take part, users had to complete a tutorial level which tested their stereo setup while familiarising them with the controls. No attempt was made to tweak or add to the core attenuation behaviour of the OpenAL library.

Mixing auditory icons, earcons and synthetic speech clearly poses a risk of confusing or, worse, annoying the user. To manage that risk, we took three steps. First, we ensured that all auditory icons fade out when the voice begins to speak (creating a 'ducking' effect). Second, we ensured that auditory icons and earcons serve completely separate functions: auditory icons represent the streets, parks, rivers, landmarks and points of interest that make up the audio map; earcons offer auditory feedback when the user steps ahead, turns, reaches a decision point or walks into an obstacle. Finally, the earcons occupy a very narrow band of the frequency spectrum, which means that they are unlikely to mask or be masked by the field recordings used as raw material for the auditory icons. They are based on a warm but recognisably synthetic synthesizer waveform. Sample recordings are available at http://team.sourceforge.net.

A number of considerations determined the choice of looped samples that make up the auditory icons. To start with, the samples had to sound sufficiently expressive to allow the user to associate them with real objects intuitively and with ease. To quote the phrase Gaver borrows from Hutchins et al., they had to possess 'articulatory directness' (Hutchins et al. (1985) cited in Gaver (1986, 172)). The loops had to focus on a single episode or event: a bus idling noisily before pulling out into the road; a cyclist ringing the bell; the constant hum of traffic on a busy road. It was important to avoid very short loops that draw attention to the fact that they are loops by forming a repetitive rhythm, but very long loops segmenting into distinct episodes were not suitable either. In the absence of visual cues, strong family associations between auditory icons are a hindrance rather than a help. The seaside, lakes and rivers had to be distinctive. This was achieved by selecting samples that foreground seagulls (seaside), ducks (lakes) and the sound of a fast-flowing stream (rivers). Similarly, restaurants had to be differentiated from bars and cafés. Here, the focus was on the sound of dishes (restaurants), music (bars) and animated conversation (cafés). Being able to tell them apart is important not least because they tend to appear in clusters.

Of the remaining auditory icons, some have fairly intuitive representations (e.g. the accentuated rumble of a train), whereas others depend on a particular sound for recognition (e.g. shops, here represented by a recording of the intermittent beep of a supermarket scanner). There is scope for broadening the range of auditory icons to represent landmarks on a truly global scale: clearly supermarket scanners are commonplace in many but by no means all countries. One key factor is the availability of a suitable loop point: the frequency distribution and amplitude at the start of the sample and at the loop point have to be as close as possible to avoid an abrupt

transition that betrays the artificial construction of the 3D environment. The samples are taken from across the globe. The field recording of the café happens to be in English – the principal voice clearly refers to an 'interview' – but many of the other recordings are not. As the vast sound archive of http://www.freesound.org shows, a café in Turkey sounds much like a café in the United States.

The earcons serve two clearly defined purposes: on the one hand, they offer feedback when the cursor keys are pressed, that is, when the user walks straight ahead, turns 45 degrees left or right, or performs a 180 degree turn; on the other hand, they convey information about the options available at decision points. Both earcon families are based on short, discreet synthesizer waveforms that are unlike any of the auditory icons. The navigational earcons are derived from just three notes taken from the minor scale. In keeping with Blattner et al.'s guidance, the three notes – C^4, Bb^3 and D^4 – do not risk dissonance even when they are played together (Blattner et al., 1989, 25–6).

At decision points, a very different kind of earcon is played. The function is twofold: first, the earcon alerts users to the fact that they have reached a decision point. In all other cases, the system announces the title of dialogs as they open and offers guidance on navigation. Here, the priority was to make the interruption of the flow of movement as light as possible, but something was needed to signal to the users that they are expected to make a choice, that a temporary mode switch has taken place. Second, the earcon seeks to prepare the user for the decision by signalling how many options there are in total, and also in which directions they are able to turn.

The inclusion of synthetic speech (the system is self voicing) was prompted by three requirements. First, the evaluation has to be exactly the same for blind and sighted users. Second, it was not possible to ask sighted users to install a screen reader, partly because screen readers by their very nature have to interact with low-level system libraries and typically require administrator privileges. Finally, general-purpose screen readers do not handle graphical canvas views such as the one used here well; at best they would offer a steady stream of redundant information, at worst they would say nothing at all.

3.2. Evaluation design

To make the evaluation available to the largest possible number of participants, ideally spread over a wide geographic area, we relied on the astonishing service that open source hosting companies offer free of charge. SourceForge.net hosted the project home page, the server and database that underpinned the evaluation, provided download mirror servers and version control for the source code. As long as the software is easy to install on personal computers (pressure of time ruled out a Mac port),

does not require on-site training, and does not call for peripherals other than a keyboard and headphones (or loudspeakers), this distribution channel alone opens the project up to a huge number of potential participants.

Wayfinding success was defined as the participants' ability to complete the four levels. Keeping on track (the third of the four wayfinding stages as described in Downs and Stea (1977)) requires them to combine what they can hear along the way with the instructions offered at the start of the level.

The game aspect was kept as simple as possible. Players set out to reach a predefined destination, collecting coins along the way. The level of difficulty quickly ramped up to genuinely challenging sets of instructions. No attempt was made to introduce opponents or obstacles. In the background, each step and action was timed and recorded in the project database.

There were two game modes: the first offered full 3D audio, earcons and synthetic speech, whereas the second replaced 3D audio with spoken descriptions of the player's immediate vicinity, in effect a cut-down version of the 'sonar' function. The non-ambient mode had the advantage of announcing points of interest by name, but also the considerable disadvantage of doing so sequentially. Earcons and general instructions were unchanged. The expectation was that non-ambient mode would lead to significantly higher cognitive load and lower evaluation scores (see Section 4).
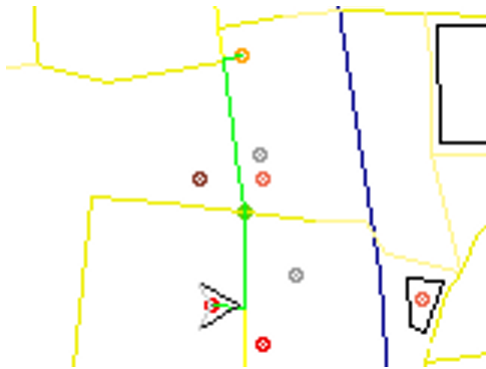
Participants took part in the evaluation from a computer of their choice, at a time that suited them. There were four levels (plus a tutorial level) in total. Sighted participants took part 'blindfolded' in the sense that the map was hidden and the status bar blank. The evaluation used a counter-balanced within-subject design. There were two groups, A and B, which were assigned to participants on the basis of the username. Group A played levels 1 and 3 with 3D audio switched on, and levels 2 and 4 without; Group B played levels 2 and 4 with 3D audio, and levels 1 and 3 without. This arrangement ensured that (a) all participants played all levels, (b) no participant played a given level twice, (c) all levels were played in ambient and non-ambient mode in equal measure, and (d) all participants experienced ambient and non-ambient levels in the same order.

The level instructions were communicated verbally and rather tersely. The lack of detail was intentional (to reduce the amount of information to be remembered) and players encountered contextual cues when navigating the map. One important characteristic of these instructions was that they did not amount to full turn-by-turn instructions. When fully sighted participants came to sketch the path in the doodle, or when partially sighted or blind participants described the route they had taken in free text, it was vital that they did so on the basis of what they had learned

Figure 2. Level 1: Leave Shakespeare's Globe Theatre and turn right along the river. Walk on until you reach your destination, Pizza Express.
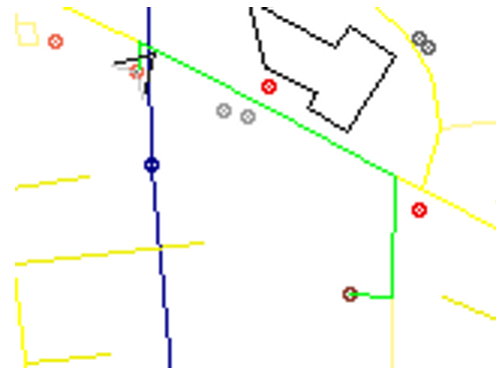

Figure 4. Level 3: Your starting point is the fast food restaurant 'Subway'. Make your way across the railway line to the pub 'Jamies Wine Bar' in Fleet Place.

Figure 3. Level 2: You start at Blackfriars bus stop. Find the nearest café to the North of you.

Figure 5. Level 4: Walk along the canal towpath and cross the canal at the first opportunity. Having crossed the canal, stay as close to the water as possible and make your way to the nearest pub.

about the planar virtual environment, not the instructions they had been given at the start. This learning comprised clues such as different lengths of route segments, the angles at which they connect, map orientation, and landmarks or intersections not mentioned in the level instructions. See Figures 2, 3, 4 and 5 for the four maps and level instructions.

In the general application, users can adjust the stride length, but for the evaluation each step (corresponding to one pixel on the map) was fixed at 1.85m. This length offered a good balance of spatial resolution (the player can negotiate complex junctions) and speed of travel (the evaluation can be completed in under half an hour).

The level of difficulty ramped up in stages. The tutorial level was intended as a gentle introduction. The first two levels pivoted on a single turn – turning right along a river and turning into a street bearing north, respectively – followed by a fairly direct approach to one's destination. The last two levels required two turns, both of them entirely dependent on contextual cues – crossing a railway line and a canal, respectively.

Completion of a given level was followed by a retrieval phase during which participants were invited to draw or describe the map (giving as much detail as possible) and provide feedback using a free-text field and four five-point Likert items: the first two about the impact of the auditory clues (helpful, distracting) and the last two about the player's experience of navigating the map (feeling lost, being confident of heading in the right direction). The former pair formed the sonification index and the latter the wayfinding index. It was assumed that the two items that form each index are inversely proportionate on intrinsic grounds. A 'distracting' rating for the auditory reduced the sonification index and finding them helpful raised it. In the same way, feeling lost reduced the wayfinding index, whereas being confident of heading in the right direction raised it. Higher values were better.
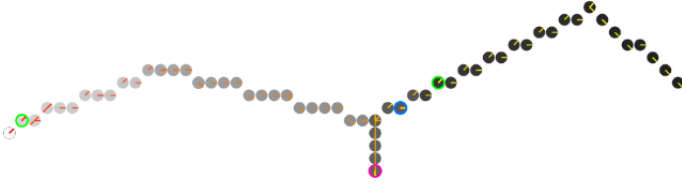
Figure 6. Visualization of the player's movements and actions through time

3.3. Database design

Before the start of evaluation sessions, all participants were alerted to the fact that their actions during the game were logged for research purposes, at which point they were free to press Cancel or Escape to exit. The success of the evaluation hinged on the application's ability to record and transmit what was happening to the project database. The database table contains ten columns – user, level, mode, step number, x, y, bearing, date and time, location, action – and each row fully describes a single action (e.g. step forward, turn left, sonar) in such a way that the journey can be recreated by analysing all rows associated with a given user. At the end of the retrieval phase, further rows are added to capture the user's feedback (map doodles, route descriptions, Likert responses, and so on).

All actual movements and most actions can be visualized in the browser using a report that turns all records associated with a given user and level into an HTML5 canvas trace with colour coding to convey the direction and speed of travel (see Figure 6).

Each point visited is represented by a disc. The colour of a given disc signals the point in time at which the player occupied a particular position. At time t_0, the disc is a light shade of grey. At t_N (which corresponds to the user's final logged action in the level) it is a very dark shade of grey. In between, the shade is determined by the time relative to the start and end times. If a player generally moves quickly but hesitates at a particular point, a clear step change from the point at which the player rests to the next will make this plain to the observer. This also reveals where players turned around and potentially took wrong turns. The player bearing is indicated by a thin line drawn from the centre of a given disc to the outside. If a player performs a 360 degree turn, the corresponding disc resembles a wheel with eight spokes. The colour changes from red to yellow to ensure that the bearing indicator is always visible. Certain events are flagged using coloured bands around discs. A blue band indicates a 'where-am-I' request; a green band the sonar function; a magenta band denotes a 'wrong turn' message.

4. EVALUATION

Following pilot studies in April and May 2010, the evaluation ran from 17 July to 24 August 2010. Participants were recruited on forums, mailing lists and also through personal contacts. In some cases, forum members passed the invitation on to further lists and forums. Canada, Germany, Iraq, Kuwait, New Zealand, Russia, Slovakia, Sweden, Switzerland, the UK and the US were all represented. In all, 218 individuals started the game. Not counting the tutorial level, they have between them taken 50,875 steps, collected 854 coins, triggered 441 wrong turn messages, and reached the end of 506 levels.

Not all data gathered was fit to contribute to the formal evaluation. The criteria for taking part in the formal evaluation was completion of the tutorial level and at least two main levels. Furthermore, only the first set of levels by participants who took part several times was considered. As sudden loss of internet connectivity could have led to gaps or data corruption, all data sets that met these criteria were manually checked for completeness and validity. Another validation step was to go through the general feedback to identify any issues that might affect the validity of a given dataset. One participant, for example, noted that they had played all four levels without hearing anything. The corresponding database rows were withdrawn from the evaluation as a result.

72 participants qualified for the evaluation. Of these, 14 were women, 57 were men and one did not indicate their gender. 37 participants stated that they were blind, 2 were partially sighted and 32 were fully sighted. Their stated age ranged from 'below 20' to '61-70'. 52% were below 30, and 80% below 40.

As the evaluation was conducted remotely, it was not possible to validate the demographic data entered. As a result, demographic (age and gender) was excluded from the formal evaluation.

Each participant went through the following steps:

- Download, install and run software
- When prompted, press Enter to start the evaluation
- Complete tutorial level (this includes a test for the audio setup, with special emphasis on correct left/right placement of headphones or speakers)
- Complete level 1
- Provide feedback on level 1 (this includes a doodle drawn from memory for fully sighted users and a written description for everyone else)
- Repeat for levels 2–4

The hypotheses to be tested were based on the premise that the presence of ambient sound can reduce cognitive processing cost and so improve performance in wayfinding challenges: levels played in ambient mode would have (a) higher coin scores; (b) lower wrong turn scores; (c) higher

sonification index; (d) higher wayfinding index. For the qualitative results, the prediction was that sound sources that are spatially located add more context, and that this would be borne out in the sketch maps and descriptions. The principal goal in scoring the participants was to measure overall wayfinding performance on the basis of both quantitative and qualitative data.

5. RESULTS

5.1. Quantitative analysis

In order to determine whether the hypotheses were correct and the evidence statistically significant, multiple one-way ANOVAs were performed. The response variables were the participants' subjective indices and their objective performance scores. In each case, the factor that divided the sample into groups was the presentation mode (ambient versus non-ambient).

The two subjective indices were based on two Likert items each. The wayfinding index reflected the player's experience of navigating the map ('feeling lost' and 'being confident of heading in the right direction'), whereas the sonification index concerned the impact of the auditory cues ('helpful' and 'distracting'). The indices' range is from 0 (worst) to 4 (best). The significance level is set to $\alpha = 0.05\%$.

Subjectively, players' wayfinding and sonification indices were higher in levels with ambient sound. The mean value of the user rated wayfinding index was statistically significantly higher (see Figure 7). The mean value was raised from 1.63 (non-ambient mode) to 2.12 (ambient mode), i.e. ambient sounds improved the users' subjective wayfinding experience by $30.1\% = \frac{2.12-1.63}{1.63}$. The mean value of the user rated sonification index was also statistically significantly higher (see Figure 8). The mean value was raised from 1.93 (non-ambient mode) to 2.58 (ambient mode), i.e. ambient sounds improved the users' subjective sonification experience by $33.7\% = \frac{2.58-1.93}{1.93}$.

Objectively, players achieved higher coin scores and lower wrong turn scores in levels with ambient sound. Coin scores were statistically significantly higher (see Figure 9). The mean value was raised from 72.3% (non-ambient mode) to 90.7% (ambient mode), i.e. ambient sound improved the users' coin score by $24.6\% = \frac{0.907-0.723}{0.723}$. Players also achieved a wrong turn score that was statistically significantly lower (see Figure 10). The mean value was lowered from 19.4% (non-ambient mode) to 13.2% (ambient mode), i.e. ambient sounds improved the users' wrong turn score by $47.0\% = \frac{0.194-0.132}{0.132}$.

All four hypotheses were confirmed. The statistical significance of all four results alongside growth rates of 25% to 47% offers a strong indication that 3D audio has
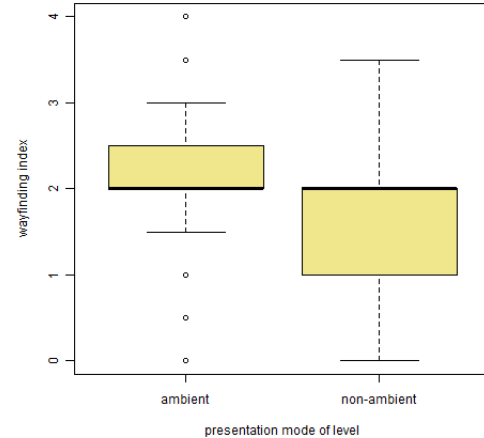

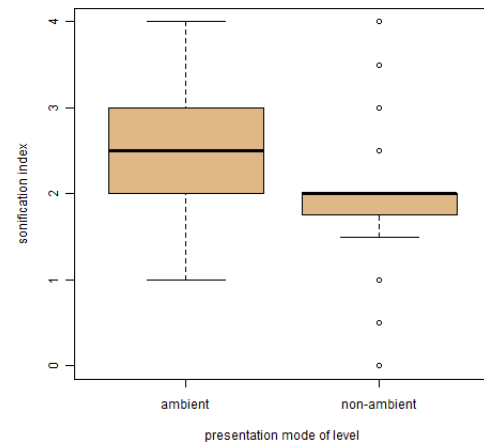**Figure 7.** Subjective indices: wayfinding index

Figure 8. Subjective indices: sonification index

much to contribute to audio maps. That said, it is also true to say that a lot of participants managed to get by remarkably well when ambient sounds were switched off. Unambiguous and strong though the lead of the ambient levels is, maps in verbal mode offered a majority of the participants just enough support to get by.

5.2. Qualitative analysis

The doodles and route descriptions drawn up during the retrieval phase demonstrate that the participants have acquired detailed route and survey knowledge, and this qualitative evidence is backed up by the visual traces of the participants' actual movements and actions in the
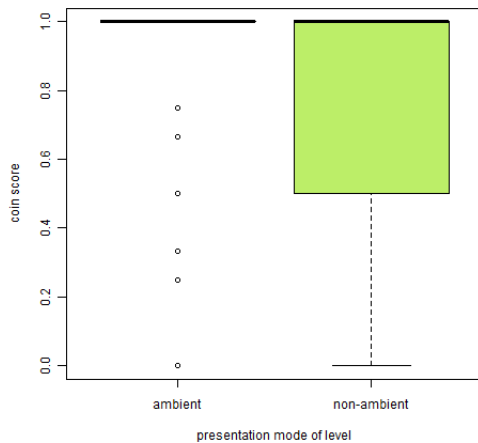
Figure 9. Objective performance scores: coin score

Figure 11. Example 1: trace
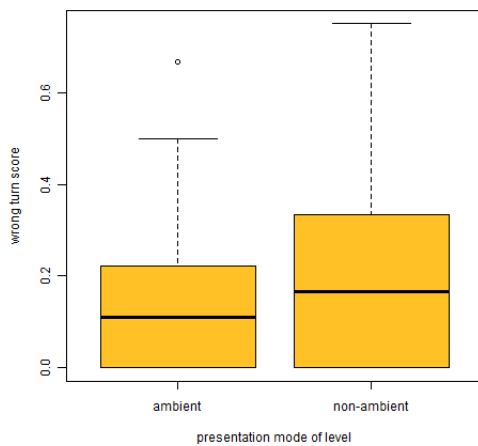

Figure 12. Example 1: drawing

Figure 10. Objective performance scores: wrong turn score

game. Here are two examples covering two of the 526 levels completed, one based on a doodle drawn by a fully sighted participant, the other on a map description written by a blind one.

At nineteen words, the level instructions for the first example are deliberately economical: 'Leave Shakespeare's Globe Theatre and turn right along the river. Walk on until you reach your destination, Pizza Express.' Figures 11 and 12 show the participant's movements and drawing (the map is reproduced in Figure 2).

Both trace and doodle (Figures 11 and 12) show a direct path from start to end point. The participant moves in straight lines, occasionally (especially at the
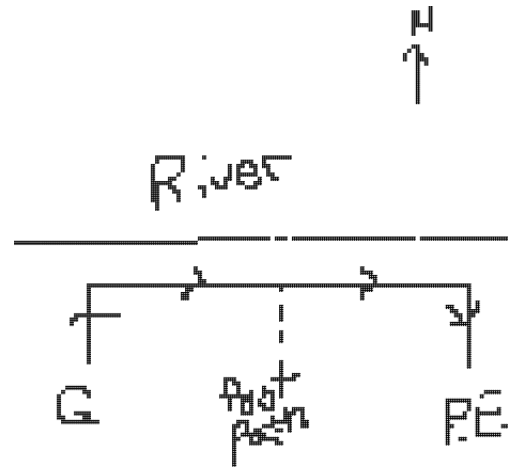
start) pausing to turn and twice requesting a 'where-am-I' message. The doodle reveals an eye for right angles, a simplification of the actual path that is both useful and elegant. Several items recorded in the drawing are not mentioned in the level instructions, especially the cardinal direction, the footpath leading off to the south and the right turn towards Pizza Express. Anyone looking for directions from the Globe Theatre to Pizza Express could do worse than follow this map.

The second example (Figure 13) shows how a blind participant has handled the (considerably more demanding) instructions for level 3. The map is given in Figure 4.

Here is the original level description placed alongside the recollected route description:

Original level instructions: Your starting point is the fast food restaurant Subway. Make your way across the railway line to the pub Jamies Wine Bar in Fleet Place.

Participant's level description (emphasis added) *Right* out of first restaurant *right* at *next junction* across railway *right* into Fleet Place and then *right* into Jamies wine bar.

The participant has turned the highly condensed level instructions into full turn-by-turn directions, correctly
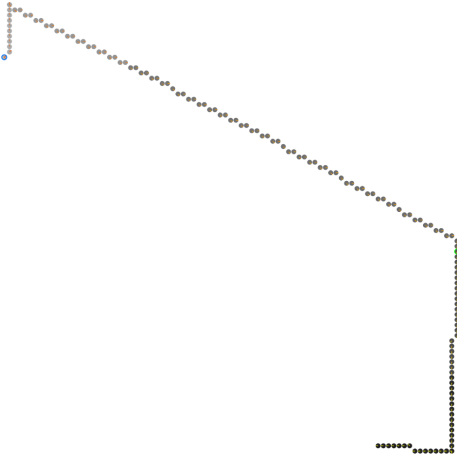
Figure 13. Example 2: trace

describing the four right turns and also, crucially, identifying the most important decision point as the 'next junction'. The route description integrates the level instructions and the contextual cues very effectively and it seems likely that this participant would now be well equipped to go on this particular journey.

6.   CONCLUSION

The results presented in the previous section demonstrate two things. First, they show that ambient 3D audio significantly improves the chances of acquiring detailed spatial knowledge in the absence of visual cues. It is clearly worth spending time and effort refining this aspect of audio maps. Integrating elevation data into the audio model is one enhancement that would make a big difference as the data is readily available from NASA and 3D audio libraries already support it. Second and most importantly, our findings suggest that audio maps built from open source tools and community-created resources such as OpenStreetMap and Freesound have the potential to convey complex city maps in such a way that blind and partially sighted people can look up routes and maps online before they venture outside exactly as sighted people do today.

REFERENCES

Archambault, D., Gaudy, T., Miesenberger, K., 2007. Computer games and visually impaired people. Upgrade 8 (2), 43–53.

Archambault, D., Gaudy, T., Miesenberger, K., Natkin, S., Ossman, R., 2008. Towards generalised accessibility of computer games. In: Pan, Z., Zhang, X., Rhalibi, A. E., Woo, W., Li, Y. (Eds.), Technologies for E-Learning and Digital Entertainment, Third International Conference, Edutainment 2008, Nanjing, China, June 25-27, 2008, Proceedings. Vol. 5093 of Lecture Notes in Computer Science. Springer, pp. 518–527.

Azuma, R., 1997. A survey of augmented reality. Presence: Teleoperators and Virtual Environments 6, 355–385.

Begault, D. R., Center., A. R., 2000. 3-D sound for virtual reality and multimedia [microform] / Durand R. Begault. National Aeronautics and Space Administration, Ames Research Center ; Available from NASA Center for AeroSpace Information, Moffett Field, Calif. : Hanover, MD. URL http://purl.access.gpo.gov/GPO/LPS70089

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

Brewster, S., 1991. Providing a model for the use of sound in user interfaces.

Brewster, S. A., Wright, P. C., Edwards, A. D. N., 1992. A detailed investigation into the effectiveness of earcons. In: In Proceedings of ICAD'92 (Santa Fe Institute, Santa Fe) Addison-Wesley. pp. 471–498.

Canada, C., Vinson, N. G., 1999. Design guidelines for landmarks to support navigation in virtual environments. In: Proceedings of CHI '99, Pittsburgh, PA. May 1999. pp. 278–285.

Christopher D. Wickens, M. V., Yeh, M., 2005. Design applications of visual spatial thinking: The importance of frame of reference. In: Miyake, A., Shah, P. (Eds.), The Cambridge Handbook of Visuospatial Thinking. Cambridge University Press, Cambridge, U.K. ; New York, pp. 383–425.

Dede, C., 2009. Immersive interfaces for engagement and learning. Science 323, 66–69.

Downs, R., Stea, D., 1977. Maps in Minds. Reflection on cognitive mapping. Harper & Row, New York.

Freksa, C., Brauer, W., Habel, C., Wender, K. F. (Eds.), 2003. Spatial Cognition III, Routes and Navigation, Human Memory and Learning, Spatial Representation and Spatial Learning. Vol. 2685 of Lecture Notes in Computer Science. Springer.

Gaver, W. W., 1986. Auditory icons: using sound in computer interfaces. Hum.-Comput. Interact. 2 (2), 167–177.

Gaver, W. W., 1989. The sonicfinder: an interface that uses auditory icons. Hum.-Comput. Interact. 4 (1), 67–94.

Heuten, W., 2008. Non-visual support for navigation in urban environments. Oldenburger Verlag für Wirtschaft, Informatik und Recht.

Heuten, W., Henze, N., Boll, S., 2007. Interactive exploration of city maps with auditory torches. In: Rosson, M. B.,

Gilmore, D. J. (Eds.), CHI Extended Abstracts. ACM, pp. 1959–1964.

Hiebert, G., Gordon, R., Peacock, D., Vogelsang, C., 2005. OpenAL 1.1 Specification and Reference. Creative Labs. URL http://connect.creativelabs.com/openal/Documentation/

Höckner, K., Marano, D., Neuschmid, J., Schrenk, M., Wasserburger, W., 2012. Accessiblemap - web-based city maps for blind and visually impaired. In: Miesenberger, K., Karshmer, A. I., Penaz, P., Zagler, W. L. (Eds.), ICCHP (2). Vol. 7383 of Lecture Notes in Computer Science. Springer, pp. 536–543.

Hutchins, E. L., Hollan, J. D., Norman, D. A., 1985. Direct manipulation interfaces. Hum.-Comput. Interact. 1 (4), 311–338.

Klippel, A., Dewey, C., Knauff, M., Richter, K.-F., Montello, D. R., Freksa, C., Loeliger, E.-A., 2004. Direction concepts in wayfinding assistance systems. In: Baus, J., Kray, C., Porzel, R. (Eds.), Workshop on Artificial Intelligence in Mobile Systems (AIMS'04). SFB 378 Memo 84; Saarbrücken, pp. 1–8.

Klippel, A., Tappe, H., Habel, C., 2003. Pictorial representations of routes: Chunking route segments during comprehension. In: Freksa et al. (2003), pp. 11–33.

Kuipers, B., 1994. An ontological hierarchy for spatial knowledge. In: AAAI Technical Report FS-94-03. AAAI, pp. 93–97.

Kuipers, B., 2000. The spatial semantic hierarchy. Artif. Intell. 119 (1-2), 191–233.

Lawton, C. A., 1994. Gender differences in way-finding strategies: Relationship to spatial ability and spatial anxiety. Sex Roles 30 (11–12), 765–779.

Lynch, K., 1960. The Image of the City (Harvard-MIT Joint Center for Urban Studies Series). The MIT Press.

McCrindle, R., Symons, D., 2000. Audio space invaders. In: Proceedings of the Third International Conference on Disability, Virtual Reality and Associated Technologies. pp. 59–65.

Montello, D. R., 1993. Scale and multiple psychologies of space. In: Frank, A. U., Campari, I. (Eds.), Spatial Information Theory: A Theoretical Basis for GIS. Springer, pp. 312–321.

Montello, D. R., 2005. Navigation. In: Miyake, A., Shah, P. (Eds.), The Cambridge Handbook of Visuospatial Thinking. Cambridge University Press, Cambridge, U.K. ; New York, pp. 257–294.

Oren, M. A., Harding, C., Bonebright, T., 2008. Evaluation of spatial abilities within a 2d auditory platform game. In: Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility. Assets '08. ACM, pp. 235–236.

Parente, P., Bishop, G., 2003. Bats: The blind audio tactile mapping system. In: Proceedings of ACM South Eastern Conference.

Parush, A., Ahuvia, S., Erev, I., 2007. Degradation in spatial knowledge acquisition when using automatic navigation systems. In: COSIT'07: Proceedings of the 8th international conference on Spatial information theory. Springer, Berlin, Heidelberg, pp. 238–254.

Peres, S. C., Best, V., Brock, D., Frauenberger, C., Hermann, T., Neuhoff, J. G., Nickerson, L. V., Shinn-Cunningham, B., Stockman, T., 2008. Auditory interfaces. In: Kortum, P. (Ed.), HCI Beyond the GUI: Design for Haptic, Speech, Olfactory, and Other Nontraditional Interfaces. Elsevier, Ch. 5.

Rinck, M., 2005. Spatial situation models. In: Miyake, A., Shah, P. (Eds.), The Cambridge Handbook of Visuospatial Thinking. Cambridge University Press, Cambridge, U.K. ; New York, pp. 334–382.

Sánchez, J., Lumbreras, M., Cernuzzi, L., 2001. Interactive virtual acoustic environments for blind children: computing, usability, and cognition. In: CHI '01: CHI '01 extended abstracts on Human factors in computing systems. ACM, pp. 65–66.

Sánchez, J., Sáenz, M., Pascual-Leone, A., Merabet, L., 2010. Enhancing navigation skills through audio gaming. In: CHI EA '10: Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems. ACM, New York, NY, USA, pp. 3991–3996.

Schneider, L., Taylor, H., 1999. How do you get there from here? mental representations of route descriptions. Applied Cognitive Psychology 13, 415–441.

Tschander, L., Schmidtke, H. R., Eschenbach, C., Habel, C., Kulik, L., 2003. A geometric agent following route instructions. In: Freksa et al. (2003), pp. 89–111.

Tversky, B., 2003. Navigating by mind and by body. In: Freksa et al. (2003), pp. 1–10.

Wiener, J. M., Büchner, S. J., Hölscher, C., 2009. Taxonomy of human wayfinding tasks: A knowledge-based approach. Spatial Cognition & Computation 9, 152–165.