Technical Report No: 2004/08

Using Spatial Audio in Minimal Attention Interfaces: Towards An Effective Audio GPS Navigation System

(This version revised 2005 to add a courtesy note on final page)

Robert Day
Simon Holland
David Bowers
Anton Dil

11th March 2004

Department of Computing
Faculty of Mathematics and Computing
The Open University
Walton Hall,
Milton Keynes
MK7 6AA
United Kingdom

http://computing.open.ac.uk

Using Spatial Audio in Minimal Attention Interfaces: Towards An Effective Audio GPS Navigation System

Robert Day, Simon Holland, David Bowers and Anton Dil
{r.h.day, s.holland, d.s.bowers, a.dil}@open.ac.uk
Department of Computing
Faculty of Mathematics and Computing
The Open University
Walton Hall
Milton Keynes
MK7 6AA
United Kingdom
Tel: 01908 653778

1. Abstract

This paper builds on previous empirical work (Holland et al, 2002) and discusses the potential for spatial audio to be used in minimal attention user interfaces, and in the context of navigational tasks. Spatial audio in this context is closely related to 3-d audio or virtual audio. The benefits of a spatial audio system for users are highlighted, with particular attention given to in-car systems. In-car systems have been chosen due to their inherent and heightened need for minimal attention interaction, and for the purposes of illustration. Current in-car navigation technology is critically discussed, with attention to potential issues that the proposed spatial audio system may address. Alternative spatial audio systems, such as those developed for the visually impaired, are also highlighted and discussed with reference to the implications of differences (and similarities) to the proposed system. The relative importance of psychoacoustics is suggested in terms of guiding the appropriate spatial audio system design. Current issues surrounding spatial audio technology are discussed, and implications for the proposed system are highlighted. Areas of potential investigation are taken from existing research and suggestions for further work to be undertaken are made.

2. Introduction

Many applications for mobile computing involve providing assistance or information to users engaged in an activity, such as driving a car, where the critical focus of the user's attention needs to be directed towards the task at hand. In addition to constraining user attention, the demands of such tasks often call for the user to be physically involved, resulting in a need for any mobile computer to require minimal prompting or physical interaction from the user. The design goals of minimal attention user interfaces (Pascoe et al., 2000) and ubiquitous computing (Weiser, 1991) offer potential solutions to the requirements of such scenarios. The aim of minimal attention user interfaces (MAUIs) is that they be minimally intrusive to the user, and convey the information required without impairing task performance. Ubiquitous computing, in turn, involves computer systems that are instantly and effortlessly available when required, yet invisible when redundant. It is clear that an ideal interface for many task-critical mobile applications would involve hands-free, eyes-free interaction. In addition, necessary information should be presented using a medium that requires minimal cognitive processing and avoids distraction.

In recent years developments in the technology of Global Positioning Systems (GPS), combined with falling costs involved with implementing this technology have led to a proliferation of systems in the marketplace. However, current GPS navigation systems are typically characterised by small screen displays, which are at best augmented by simple speech audio. Such systems require cognitive processing which place demand on user attention away from the critical task of safely operating a vehicle, or even negotiating an environment on foot. This paper will argue that spatial audio may offer the potential to better meet such demands, and could be implemented effectively as an interface medium for minimal attention mobile applications. Spatial audio involves the binaural processing of sound in order that it can be presented to a user in such a way that it appears to emanate from a particular point in the virtual space surrounding them. Therefore, spatial audio is able to provide users with directional cues, and may lend itself well to the sonification of GPS navigational system interfaces. A spatial audio GPS system of this nature might then be suitable for mobile navigational tasks, minimising demands upon user's visual attention.

This paper begins by discussing the task of navigation, and the task requirements that a navigation information system needs to fulfil. Research into mobile computing is then used to develop the argument for the benefits of such a system employing an audio interface. Research into psychoacoustics is reviewed and described, with reference to any implications it may have upon spatial audio system design. Current systems, either commercially available or being developed by other researchers in a different context, are described and reviewed. Lastly the system design implications drawn from the research are summarised and plans for further research work are detailed.

3. The Task of Navigation

Navigation is defined as the process of directing the movement of a craft from one place to another. The word is derived from the Latin words 'navis' meaning ship and 'agere' meaning to direct or move, although in its modern sense the word is applied to more than just nautical environments. The basic information required to perform the general task of navigation are *current position*, *heading* (orientation), *track* (direction of movement), *bearing* (direction of target relative to heading) and *distance* of a target destination (see figure 1). All this information (bar heading) can be provided by modern GPS systems, which calculate a positional fix by communicating with 24 satellites orbiting the earth. These satellites broadcast a signal from tracked known locations which and GPS receivers calculate the time taken for such signals to reach them and from this deduce the distance of the satellites. When a GPS system receives signals from at least three appropriately aligned satellites it can calculate its position on earth via triangulation (four satellites are required for calculating altitude). Target locations can be stored as 'waypoints' in GPS systems or programmed into them from mapping software. GPS systems calculate heading using either an inbuilt electro-magnetic compass or simply by calculating track when moving by rapidly updating the position fix and assuming that heading is synonymous with the direction of travel. Clearly, a disadvantage of the latter method is that reliable heading information is not available when stationary. Lastly bearings for target destinations are calculated by comparing their position with the GPS systems current calculated position and track. Again however, for systems without an electro-magnetic compass, such information is unreliable when stationary.

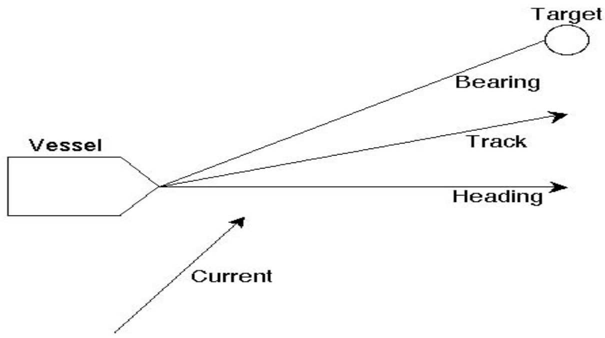

Figure 1. An illustration of the potential difference between *bearing, track and heading* (employing a nautical example).

It is worthy of note that the task of navigation in its various contexts may also involve more complex activities such as collision avoidance and compensation for constraints such as current (waterborne navigation) or fixed routes (i.e. channels, roads or paths). At this investigatory stage however the scope of this research will be limited to the basic navigational task as described, and the speculative potential of spatial audio interfaces in this context.

4. Potential for Spatial Audio Interfaces in Mobile Computing

The vast majority of computing systems developed have opted to focus on a heavily visual graphical user interface (GUI). The benefits of GUI based systems in a variety of contexts (due to our inherent visual 'disposition') have been well established. GUI based systems however, by their very nature, do not lend themselves well to hands free, eyes free interaction. Mobile devices typically have small screens, and therefore extremely limited display capacity. Such limitations potentially make current mobile computing devices difficult to read or see, particularly in the context of undertaking an external attention-demanding task. For these reasons, several researchers have questioned the suitability of GUI interfaces in the context of mobile applications (Kristoffersen and Ljungberg, 1999; Brewster and Walker, 2000; Brewster et al, 2003). An alternative communication medium that offers particular advantages to minimal attention interfaces and ubiquitous computing is the auditory channel. Kristoffersen and Ljungberg (1999) have developed an interaction style specifically for mobile computing they term MOTILE. The MOTILE interaction style was informed by the perceived requirement for mobile computing to place little demand upon users visual attention, and therefore proposes that auditory feedback should be used. In addition Brown et al. (1989) among others have suggested the potential for auditory cues to reduce 'visual workload' in conventional computing environments. By using visual search tasks pre-empted with either a visual or an auditory cue Brown et al. found that complex auditory cues were at least as effective in aiding participant performance as visual cues.

Serious safety considerations surround the use of computer equipment while driving due to the potential for such equipment to distract drivers from attention from the road. In response to research, and a report by the Royal Society for the Prevention of Accidents (ROSPA), UK legislation has been introduced restricting the use of mobile telephones whilst driving. Redelmeier and Tibshirani (1997) conducted a study in Canada based upon billing information of mobile phone users who had been involved in road traffic accidents. They concluded that the risk of having an accident was 4 times greater when using a mobile phone due to the additional demands on physical and cognitive attention. The Redelmeier and Tibshirani study was cited in support of the legislation change discussed, and has therefore been subjected to considerable scrutiny. As a result of such scrutiny the study has been criticised for using potentially inaccurate data derived from mobile telephone bills. Redelmeier and Tibshirani (2001) however, have defended their findings claiming that such limitations in actual fact biased their risk assessment exercise towards finding nothing. Interestingly Trbovich and Harbluk (2003) conducted a study using eye-tracking of participants driving in actual traffic situations. They found that even when using 'hands free' mobile telephones, participants paid significantly less visual attention to traffic signals. One potential criticism of this experiment however was that participants were not merely engaged in conversation using mobile telephones, but were instead asked difficult arithmetic questions and asked to respond with answers. As arithmetic questions were not asked in the control condition (driving without using a mobile phone), potential distraction due purely to the cognitive load of such a task may have confounded any effect due to the mobile phone itself.

Burnett and Joyner (1997) conducted road trial research comparing currently available in-car navigation systems (with visual interfaces) with verbal navigation instructions from an 'informed' passenger. They recorded drivers' visual activity and found that when using the visual interface system drivers were found to spend less time looking at the road ahead and checking mirrors. This was due to the fact that over 18% of participants' visual attention was diverted towards the navigation system itself. In another study based on a test track, Tijerina et al. (1998) found that when participants were asked to input a destination into a commercially available navigation system, the mean time taken was between 40 seconds and 2 minutes (dependent on the age group involved). In addition, and of equal concern, the mean number of accidental lane departures during this procedure was 0.9, which indicated that often participants strayed from their lane. This compared to mean task times of between 15 and 20 seconds, and between 0.1 and 0.2 lane departures for tuning a radio or dialling a number on a mobile phone. With the proven impact of mobile phone use on driving safety, the implication that current visual navigation systems may be more distracting is clearly cause for concern. It would seem therefore, that a minimal attention spatial audio navigation system might offer a viable non-speech, non-visual alternative worthy of investigation.

5. Spatial Audio Versus Speech Output

Spatial audio involves the binaural processing of sound in order that it can be presented to a user in such a way that it appears to emanate from a particular point in the virtual space surrounding them. Moreover, the term spatial audio, in the context of the proposed system, is used to describe non-speech spatially presented sounds. Current in-car navigation systems employ at best speech audio instructions in an attempt to improve safety for drivers by reducing distraction. A clear advantage of spatial audio over standard speech output is the preservation of a clear conversation channel. With a free speech channel users are able to talk to passengers, use a hands free mobile phone or even listen to speech on the radio. Some navigation systems have used spatially processed speech audio (Loomis et al. 1998; 2001; Loomis, 2003 – see discussion below). However comparisons of the effectiveness of spatial audio and spatial speech have demonstrated some differences. Klatzky et al. (2002; 2003) compared the 2 modalities and found that participants' ability to update spatially was comparable. Spatial updating refers to the ability to navigate to a target after moving from their start point in a fixed direction, other than towards the target. However, for multiple locations, mental models were 'learned' faster when using non-speech spatial audio than spatial speech, despite the fact that participants confirmed azimuth for each cue with a head nod in both conditions with comparative accuracy. Klatzky et al. concluded that this effect was due to extra cognitive processing loads imposed by converting language to spatial content. Based on this conclusion it is not unreasonable to suggest that if language requires extra cognitive processing, and therefore cognitive load, it is probably not the most suitable medium for minimal attention interfaces.

6. The Role of Psychoacoustics

Research into spatial hearing and psychoacoustics is, perhaps surprisingly, extensive and long established. Published studies were conducted in the field of spatial hearing as early as the late nineteenth century (e.g. Politzer, 1876; Bloch, 1893). Lord Rayleigh (1907) proposed an early (and subsequently developed yet still widely accepted) theory to explain human perception of the direction of a sound source termed the 'Duplex Theory'. This theory explains that the perceived azimuth angle of a sound source (i.e. the perceived angle of the sound source from being straight ahead of the listener in the horizontal plane) is dependent on cues provided by the Interaural Time Difference (ITD) and the Interaural Level Difference (ILD). ITD refers to the potentially minute time difference for a sound wave emanating from an azimuth angle other than 0° (straight ahead) or 180° (directly behind) to reach a listener's two ears. More precisely a sound emanating from a listeners left will reach their closer left ear fractionally before it reaches their farther right ear. ILD refers similarly to the marginal differences in the amplitude of such a sound volume at each ear.

In addition to these two comparatively simple factors 'spectral filtering' of sounds (before they reach the eardrum) by the head, shoulders and especially the outer ear also contribute to spatial hearing. The outer ear, or 'pinna', is made up of a series of folds and ridges of cartilage which reflect sound waves and cause minute amplitude and time differences. These differences alter the spectral content reaching the eardrum in a manner dependent upon the spatial position of the sound (Begault, 1994). Also, the head acts as a baffle between the two ears and in fact amplifies differences in sound signals (Blauert, 1983). Lastly, Searle et al. (1976) have demonstrated an effect of the shoulders, termed 'shoulder bounce', whereby sound waves are reflected from the shoulders and either amplified or suppressed, dependent upon their frequency, resulting in spatial cues. Such filtering mechanisms of the head and body are termed the head related transfer functions (HRTFs), and are a source of individual difference in perception of localised sound due to human anatomical variability. HRTFs can be measured by inserting small microphones into participants' ear canals, or alternatively by using 'dummy' heads containing microphones. An ideal spatial audio system would employ individually tailored HRTFs to maximise localisation accuracy, however this would obviously make the system impractical to implement. Fortunately Wenzel et al. (1993) found that a generalised set of HRTFs measured from an individual demonstrated to be a good 'localiser' of spatial audio afforded localisation accuracy comparable with free field listening for most participants. This finding indicates that a practical system employing a carefully selected and validated set of generalised HRTFs can yield acceptable localisation accuracy for much of the population.

A great deal of research has been conducted into localisation accuracy in human spatial hearing. In general such research can be divided into two main fields based upon the methodology used and data collected. One field of research has focussed upon the absolute accuracy of localisation in human spatial hearing (Makous and Middlebrooks, 1990; Oldfield and Parker, 1984). Such research requires participants to indicate the perceived azimuth angle (direction) of a controlled sound source. By contrast another field has focussed upon the sensitivity of the human auditory system to small changes in the locality of sound sources, or finding the 'minimum audible angle' or MAA (Mills, 1958; Hartmann and Rakerd, 1989). The findings of such empirical work offer some important considerations to the design of a virtual spatial audio system. In particular localisation blur or MAA has been shown to be greatly variable depending upon the stimulus presented and the azimuth of the sound source.

The absolute minimum audible angle in the horizontal plane (azimuth) has been shown to be approximately 1° when stimuli is straight ahead, or 0° azimuth, (Schmidt et al., 1953) and usually varies between 2° and 3° (Carlile et al., 1997). However it has been widely shown that when a sound source is moved left or right of 0° azimuth (straight ahead) the MAA increases and therefore localisation accuracy decreases. The MAA is between 3 and 10 times larger at 90° in either direction and then, although decreasing once more, is twice as large at the rear (Blauert, 1983). This indicates that human audio localisation accuracy is weaker in the rear hemisphere than the front hemisphere and weakest of all at the flanks of an individual's auditory space. Any potential spatial audio system therefore might consider attempting to provide additional cues as to azimuth angle if high-resolution accuracy is required. In addition several studies have shown that localisation blur is influenced by the nature of the sound stimulus presented, with accuracy reduced by restricting the bandwidth of the stimulus (e.g. Butler, 1986). The implication of such research is therefore that improved localisation can be achieved by using sound stimuli with broad bandwidths.

A common type of error widely reported in studies of localisation accuracy (Carlile, 1996) involves the mirroring of perceived and actual sound locations between the hemisphere in front of the participant and that to their rear. Such errors (often called front-back errors) are probably due to the fact that at the exact mirror opposite positions in the front and back hemisphere, distances between sound sources and each ear (and therefore presumably ITD and ILD) will be more or less the same. In fact ITD and ILD will be the same for any point on a hyperbola around the axis of a line between the two ears of a listener, with the listeners ears as its foci. This phenomenon is often referred to in the literature on spatial hearing as the 'cone of confusion' (see figure 2). Increasing the sound stimulus bandwidth has again been shown to help reduce front-back error frequency (Middlebrooks, 1992). In fact, research into the development of emergency vehicle sirens that are easy to localise, found that by introducing bursts of broadband (or 'white') noise into the siren sound front-back errors were reduced from 56% to 18% (Withington, 1996; 1998). In addition, Plenge and Brunschen (1971) among others have found that accuracy problems with spatial audio, including front-back errors, are generally reduced with exposure to the system (i.e. training and experience). Lastly, it has been shown that spatial hearing accuracy is improved, and front-back errors reduced, when it is possible for listeners to move their head (Thurlow and Runge, 1967; Begault et al., 2001).

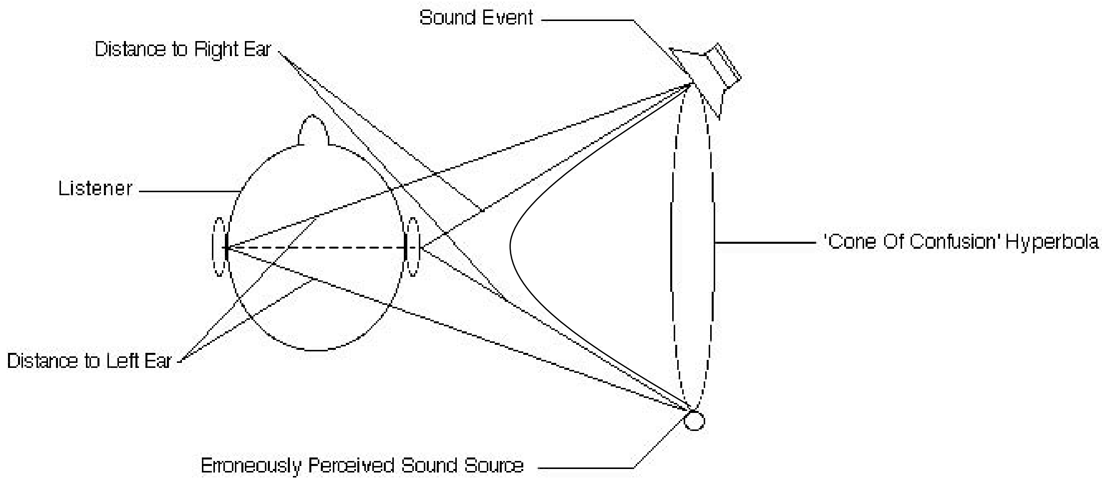

Figure 2. A diagram to illustrate the so-called 'cone of confusion' and an example of a resultant front-back error.

Another research finding is the relatively poor ability for spatial audio to accurately convey distance information. Zahorik (2002) conducted simple trials in which participants were asked to estimate the distance of a sound source. Zahorik found that at distances greater than a metre participants' average estimated distance was substantially less than the actual distance of the source. Conversely, at distances less than a metre, participants tended to overestimate the distance of the sound source. In addition to this effect, termed 'estimate bias', the results of this study also showed that estimates were extremely variable (even within the data for individuals), although some individuals' estimates were considerably less variable than others. Although Zahorik offers guidelines for the improvement of distance perception in auditory displays, he concludes that human auditory distance perception is simply inaccurate, even in real audio environments. Clearly, as one of the key requirements of any navigation system is to accurately convey distance information for target destinations, it would seem unwise to rely on spatial audio to do this. In the case of the proposed system, it is fair to say that the primary objective is to efficiently enable completion of the navigation task and not the authentic reproduction of an auditory environment. Therefore the proposed system would benefit from distance mapping being implemented in a more specific and discernable manner.

A further consideration relating to localisation accuracy emerges from the effect of the listeners' physical environment on sounds heard via reverberation. In an interesting and highly relevant paper, Hartman et al. (2002) demonstrated that the asymmetric speaker positioning of car stereo systems, relative to the driver's position, requires some consideration for spatial audio presentation. More precisely when speakers are asymmetric (in both angle and distance to listeners' ears), the localisation cues provided become conflicted when simply presented as if over a basic stereo system. This conflict can cause stereo 'images' to seem spread out or even drift over time. Hartman et al. recommend that to counteract this effect, correction factors (level adjustments and time delays) should be calculated and applied to each speaker. In this way the system can be balanced, and localisation cues can be accurately presented to users in the driving position.

7. Current Systems

The term navigation, in its broader sense, has also been used to describe the process of moving through information system interfaces or virtual environments. Although these processes may not always have obvious implications on physical navigation tasks, research in this field does offer some insight into the arguments developed in this paper. In particular it is worth noting that spatial audio has been suggested (and successfully implemented) as a suitable output medium to facilitate eyes free interaction (Crispien and Fellbaum, 1996) and especially in the context of mobile, small screen systems (Brewster and Walker, 2000; Brewster et al., 2003). Potentially more specifically relevant studies have also shown that spatial audio can improve navigation performance in immersive virtual environments (Gunther, 1997; Gröhn et al., 2003). These studies show that by augmenting virtual environments with spatial sound attached to objects of interest participants are able to locate such objects quicker and more efficiently.

Several researchers have developed spatial audio navigation systems to aid visually impaired users (Loomis et al., 1998; 2001; Loomis, 2003; Moon, 1998; Makino et al., 1997; Petrie et al., 1996). Although at first sight such systems appear to be extremely similar to the proposed system, there are some fundamental differences both in design goals and resultant implementation. The goal of such systems is not simply to facilitate navigation, but to improve spatial awareness and enhance blind users' mental models of their environment. In addition such systems employ speech audio, due to the increased informational load on the auditory modality. This informational load encompasses feedback, for system programming and initialisation, as well as richer information about the user's environment. By contrast, a minimal attention system is designed to be unobtrusive and facilitate navigation in the context of performing other, often critically important tasks. For this reason the information required from the system is restricted to that necessary to the basic task of navigation, and therefore non-speech audio is preferred to speech audio.

Current in-car satellite navigation aids are increasingly available but as yet typically rely on either visual or synthetic speech output in the form of maps and / or so called 'turn-by-turn' instruction. As discussed however the safety implications of such systems in terms of placing cognitive and attentional demands upon drivers are worthy of concern (Burnett and Joyner, 1997).

Lastly spatial audio has also been suggested as a suitable output medium for several other types of driver assistance systems (Bellotti et al., 2002). Such systems include:

- Lane keeping and blind spot monitoring systems
- Distance keeping and collision avoidance systems

- Parking and reversing aids

It is possible however that the use of spatial audio for such systems in conjunction with navigational information may overload the modality and therefore impair performance.

8. Conclusions and Implications for Proposed System

The research discussed offers a number of implications for the design of a spatial audio system for navigation. On a practical level the limitations of many current GPS systems mean that accurate heading information (and therefore bearings for target destinations) is not available when stationary. To overcome this problem the system will need to include an electro magnetic compass. The research into localisation accuracy suggests that it is maximised when:

- Generalised HRTFs are validated as being from a 'good localiser' (Wenzel at al., 1993)
- Broadband sound stimuli are employed (Butler, 1986)
- Listener head movements are reflected in the presentation of the sound (Begault et al., 2001)

The proposed system should take all these findings into account and use validated HRTFs, broadband sound events and possibly some form of head-tracker to enable head movement cues. However it should be noted that for some applications, such as operating a vehicle, a head-tracker might not be entirely appropriate as user head movements and the inconvenience of attaching hardware to users may result in unwanted distraction.

As accuracy of spatial hearing has been shown to be progressively weaker towards the listener's flanks and slightly weaker in the rear hemisphere than the front (Blauert, 1983) additional azimuth cues might enable finer localisation accuracy. To achieve this the current prototype system (Holland et al., 2002) presents a second 'chase' tone along with the localisation cue, which is at precisely the same pitch as the localisation cue at 0° azimuth (dead ahead), and then progressively diverges in pitch for 90° in either direction. The maximum pitch difference between the two tones, occurring at the 90° and 270° azimuth angles (directly to the right and left respectively), is an octave and the pitch is varied discretely and chromatically. This chase tone was introduced based upon heuristic evaluations in an attempt to facilitate clearer interpretation of azimuth and greater response accuracy. In addition, the prototype system uses different tones for the front and rear hemisphere in an attempt to reduce front-back errors. It is possible however that such cues may well be over complex, and therefore increase the need for user training. Additionally the introduction of broadband noise into the localisation cue may make these current azimuth mappings even more confusing or render them redundant. In any case, as these cues are based upon heuristics, they need to be empirically evaluated to assess any added benefits or distractions they provide.

The identified weakness of human spatial hearing to accurately interpret the distance of sound cues (Zahorik, 2002) places a demand for distance cues external to the spatial audio cue itself. More precisely, the ability for users to extract distance cues from spatial audio alone is limited (even in 'natural' listening scenarios), and therefore additional distance mappings are required. A problem for the proposed system is that to convey precise distance information using non-speech audio is virtually impossible. To illustrate this problem more precisely consider conveying the information that a destination is 2.3 miles away as a simple non-speech sound that requires minimal processing to interpret. However it is not entirely necessary for distance information to be so precise for it to be of use in a navigational context. Current audio reversing monitors in cars, for example, employ a 'Geiger counter' style mapping to represent the distance to obstacles. More precisely, as the distance to an obstacle decreases the number of sections in repeated audio cues increases. It is therefore possible to convey distance information in a relatively familiar way using non-speech audio, albeit in a more discrete manner than potential visual alternatives.

A more elaborate form of such a mapping could easily be used in the proposed system to aid navigation and signify arrival at target destinations. A potential version (currently under investigation in the prototype system) would be to divide the distance to the target into discrete segments. A similar 'Geiger counter' style mapping could then be employed; with an additional audio cue section added each time a segment of the distance to the target destination is covered. Conversely if a navigational error has occurred and the distance to the target has increased a reduction in the number of audio cue segments will provide an additional alert to the user. A further improvement to this mapping would be to make it hierarchical, thereby facilitating more precise distance information when near to the target destination. To illustrate this, a current version being tested divides the total journey into ten distance segments and increases the number of sections in each audio cue incrementally as users move a 'segment' closer to their destination. When users enter the 10^{th} and final segment of distance the system resets providing a further ten segments (and corresponding 'Geiger counter'), which correspond to 100^{th}'s of the total journey distance. In this way precise distance information is provided at the appropriate time, i.e. on the approach to the target destination. A final distinct 'arrival tone' can then be used to indicate arrival at the target destination.

9. Further Work

Clearly a great deal of research is required to assess some of the implications of existing research and maximise localisation accuracy in the proposed system. As part of a doctoral research studentship on behalf of the first author this research is already underway. Initially further research into the safety issues will focus on comparisons of the prototype system with existing visual and speech based navigation interfaces. Current research 'on the road' aims to compare user satisfaction, workload (using the NASA TLX inventory) and route efficiency between a visual and prototype spatial audio interface. A further study will employ driving simulation software and eye-tracking hardware to compare users' visual attention distribution and driving behaviour when using such systems. In order to compare spatial audio with speech, to assess the impact of any added cognitive processing loads, studies might again employ a driving simulation scenario. By measuring response times to a series of hazardous driving situations, as well as comparing overall route efficiency and driving behaviour, it may be possible to demonstrate that added cognitive processing detracts from task safety and efficiency.

In terms of developing and refining the system, research into optimising and validating the HRTFs employed will be conducted. In addition, research will be conducted into finding the best possible audio cues (or combinations of audio cues) to optimise localisation accuracy. Abstract comparisons of localisation accuracy for a variety of audio frequencies, intensities and bandwidths will be empirically tested. The current mappings for additional distance and azimuth cues will be thoroughly evaluated (in terms of both user satisfaction and effect upon task), and alternative mappings will be explored. Furthermore, for in-car use, careful tuning of the system balance will be attempted, in order to minimise the effect of asymmetrical speaker positioning. Lastly, in an attempt to provide head movement cues, the use of head tracking hardware will be assessed in terms of its impact upon localisation accuracy and practical suitability for in-vehicle scenarios.

Finally upon further development of the proposed system investigations will be conducted into potential extensions of its scope. Formal requirements analyses will be conducted to better understand the needs of specific tasks, users and environments. Subsequently more complex navigational tasks, user groups and environments will be addressed. The added navigational requirements of waterborne scenarios for example include compensating for 'cross-track' error. Cross-track error is essentially a discrepancy between heading (direction of movement through water) and track (absolute direction of movement relative to target destinations). Such discrepancy is due to the impact of tidal flow or current upon the sailing vessel. To compensate for cross-track error indications of deviation from the 'ideal' path, between start point and target destination, would be useful. A further extension of the system would be to augment it with some optional speech output in order to facilitate programming and set-up of the system by the visually impaired and thereby widen access to the system. In doing so the system may provide a usefully functional, yet unobtrusive, navigational aid accessible to blind users. It is important to note however that when considering such extensions of the system further research work will be required, in terms of requirements analysis, to ensure that the needs of potential users will be fully met.

10. References

Begault, D. R. (1994). *3-D Sound for Virtual Reality and Multimedia*. Cambridge MA: Academic Press Professional.

Begault, D.R., Wenzel, E.M. and Anderson, M.R. (2001). Direct comparison of the impact of head tracking, reverberation and individualized head-related transfer functions on the spatial perception of a virtual speech source. *Journal of the Audio Engineering Society*, 49, 904-916.

Bellotti, F., Berta, R., De Gloria, A. and Margarone, M. (2002). Using 3D sound to improve the effectiveness of the advanced driver assistance systems. *Personal and Ubiquitous Computing*. 6, 155-163.

Blauert, J. (1983) *Spatial Hearing*. Cambridge MA: MIT Press.

Bloch, E. (1893). Das Binaurale Hören [Binaural Hearing]. Z, Ohren-Nasen-Kehlkopfheilk. 24, 25-83.

Brewster, S. A. and Walker, V. A. (2000). *Non-Visual Interfaces for Wearable Computers.* IEE Workshop on wearable Computing (00/145). IEE Press

Brewster, S. A., Lumsden, J., Bell, M., Hall, M. and Tasker, S. (2003). *Multimodal 'Eyes-Free' Interaction Techniques for Wearable Devices.* Proceedings of CHI 2003

Brown, M. L., Newsome, S. L. and Glinert, E. P. (1989). *An Experiment Into the Use of Auditory Cues to Reduce Visual Workload*. Proceedings of CHI 1989

Burnett, G. E. and Joyner, S. M. (1997). An assessment of moving map and symbol-based route guidance systems. In Y. I. Noy (Ed.), *Ergonomics and Safety of Intelligent Driver Interfaces*, 115-137, Mahwah NJ: Lawrence Erlbaum Associates.

Butler, R. A. (1986). The bandwidth effect on monaural and binaural localization. *Hearing Research*, 21, 67-73.

Carlile, S. (1996). Auditory Space. In S. Carlile (Ed.), *Virtual auditory Space: Generation and Applications*, 1-25, Austin, Texas: R. G. Landes Company.

Carlile, S., Leong, P., Hyams, S. and Pralong, D. (1997). The nature and distribution of errors in the localization of sounds in humans. *Hearing Research*, 114, 179-196.

Crispien, K. and Fellbaum, K. (1996). *A 3-D Auditory Environment for Hierarchical Navigation in Non-Visual Interaction*. In Proceedings of the 1996 International Conference on Auditory Display, Palo Alto, California, USA.

Gröhn, M., Lokki, T. and Takala, T. (2003). *Comparison of Auditory, Visual and Audio-Visual Navigation in a 3D space.* In Proceedings of the 2003 International Conference on Auditory Display, Boston MA, USA.

Gunther, R. T. C. (1997). *The Use of 3D Sound as a Navigational Aid in Virtual Environments*. Unpublished Ph.D. Thesis, University of Waterloo, Ontario, Canada.

Hartman, R., Pohlmann, K. and Heber, K. (2002). High frequency effects on localization and sound perception in a small acoustic space. SAE Technical Paper 2002-01-0117. In Proceedings of the SAE 2002 World Congress, Detroit, Michigan.

Hartmann, W. M. and Rakerd, B. (1989). On the minimum audible angle: A decision theory approach. *Journal of the Acoustical Society of America*, 85(5), 2031-2041.

Holland, S., Morse, D. R. and Gedenryd, H. (2002). Audio GPS: Spatial audio in a minimal attention interface. *Personal and Ubiquitous Computing*, 6(4), 253 – 259.

Klatzky, R. L., Lippa, Y., Loomis, J. M. and Golledge, R. G. (2002). Learning directions of objects specified by vision, spatial audition, or auditory spatial language. *Learning & Memory*, 9, 364-367.

Klatzky, R. L., Lippa, Y., Loomis, J. M. and Golledge, R. G. (2003). Encoding, learning and spatial updating of multiple object locations specified by 3-D sound, spatial language, and vision. *Experimental Brain Research*, 149, 48-61.

Kristoffersen, S. and Ljungberg, F. (1999) *"Making Place" to Make IT Work: Empirical Explorations of HCI for Mobile CSCW*. In Proceedings of International ACM SIGGROUP Conference on Supporting Group Work (Group'99), ACM Press.

Loomis, J. M., Golledge, R. G. and Klatzky, R. L. (1998). Navigation system for the blind: Auditory display modes and guidance. *Presence*, 7(2), 193-203.

Loomis, J.M., Golledge, R. G. and Klatzky, R.L. (2001). GPS-based navigation systems for the visually impaired. In W. Barfield and T. Caudell (Eds.), *Fundamentals of Wearable Computers and Augmented Reality*. Mahwah NJ: Lawrence Erlbaum Associates.

Loomis, J. M. (2003). *Basic and Applied Research Relating to Auditory Displays for Visually Impaired People*. In Proceedings of the 2003 International Conference on Auditory Display, Boston MA, USA.

Makino, H., Ishii, I., Nakashizuka, M. (1997). *Development of Navigation System for the Blind using GPS and Mobile Phone Combination.* In Proceedings of the 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 18, Pts 1-5.

Makous, J. C. and Middlebrooks, J. C. (1990). Two-dimensional sound localization by human listeners. *Journal of the Acoustical Society of America*, 87, 2188-2200.

Middlebrooks, J. C. (1992). Narrow-band sound localization related to external ear acoustics. *Journal of the Acoustical Society of America*, 92, 2607-2624.

Mills, A. W. (1958). On the minimal audible angle. *Journal of the Acoustical Society of America*, 30, 237-246.

Moon, G. (1998). Personal navigation systems for the blind. In *Institute of Engineering Surveying and Space Geodesy*, pp. 328. Nottingham: University of Nottingham.

Oldfield, S. R. and Parker, S. P. A. (1984). Acuity of sound localisation: A topography of auditory space I. *Perception*, 13, 581-600.

Pascoe, J., Ryan, N. and Morse, D. R. (2000). Using while moving: HCI issues in fieldwork environments. *ACM Transactions on Computer Human Interaction*, 7, 417-437.

Petrie, H., Johnson, V., Strothotte, T., Fritz, S., Michel, R. and Raab, A. (1996). MoBIC: Designing a travel aid for blind and elderly people, *Journal of Navigation 49* (1), 45-52.

Plenge, G., and Brunschen, G. (1971). *Signalkenntinis und Richtungsbestimmung in der Medianebene bei Sprache [A priori knowledge of the signal when determining the direction of speech in the median plane*]. In Proceedings of the7th International Congress on Acoustics, Budapest. 19 H 10.

Politzer, A. (1876). Studien über die paracusis loci [Studies of the paracusis loci]. Arch. Ohren-Nasen-Kehlkopfheilk. 11, 231-236.

Lord Rayleigh. (1907). On our perception of sound direction. *Philosophical Magazine*, 13(74), 214 – 232.

Redelmeier, D. A. and Tibshirani, R. J. (1997). Association between cellular-telephone calls and motor vehicle collisions. *The New England Journal of Medicine*, 336(7), 453–58.

Redelmeier, D. A. and Tibshirani, R. J. (2001). Car phones and car crashes: Some popular misconceptions. *Canadian Medical Association Journal*, 164, 1581-1582.

Schmidt, P. H., Van Gemert, A. H. M., De Fries, R. J. and Duyff, J. W. (1953). Binaural threshold for azimuth difference. Acta Physiol. Pharmacol. Nederl. 3, 2-18.

Searle, C.L., Braida, L.D., Davis, M.F. and Colburn, H.S. (1976). Model for auditory localization. *Journal of the Acoustical Society of America*. 60, 1164-1175.

Thurlow, W.R. and Runge. P.S. (1967). Effect of induced head movements on the localization of direction of sounds. *Journal of the Acoustical Society of America*, 42(2), 480-488.

Tijerina, L., Parmer, E. and Goodman, M. J. (1998). *Driver Workload Assessment of Route Guidance System Destination Entry while Driving: A Test Track Study.* In Proceedings of the 5th ITS World Congress, Seoul, Korea.

Trbovich, P. and Harbluk, J. L. (2003). *Cell Phone Communication and Driver Visual Behaviour: The Impact of Cognitive Distraction*. Proceedings of CHI 2003.

Weiser, M. (1991). The computer for the 21st century. *Scientific American*, 265(3), 94-104.

Wenzel, E. M., Arruda, M., Kistler, D. J. & Wightman, F. L. (1993) Localization using non-individualized head-related transfer functions. *Journal of the Acoustical Society of America*, 94, 111-123.

Withington, D. J. (1996). The quest for better ambulance sirens. *Ambulance UK*, 11 (2), 20-21.

Withington, D. J. (1998). Siren sounds - Do They Actually Contribute to Traffic Accidents? *IMPACT, the Journal of the Institute of Traffic Accident Investigators*. Spring 1998.

Zahorik, P. (2002). *Auditory display of sound source distance*. In Proceedings of the 2002 Conference on Auditory Display, Kyoto, Japan.

Retrospectively (this note added in 2005) we have been asked to point out that that a 1996 French patent 96.05980 (and its extension in other jurisdictions), of which we were unaware, covers some aspects of personal direction finder systems. It is worth noting that the application of the basic spatial audio metaphor to navigation tasks started considerably earlier than the mid-nineties. The patent does not consider the more refined mappings noted in this technical report, and the patent appears to have little or bearing on many diverse technological means by which the basic metaphor, and more refined mappings could be put to use.