AudioGPS: spatial audio in a minimal attention interface

SIMON HOLLAND, DAVID R. MORSE & HENRIK GEDENRYD

Computing Department, The Open University, Walton Hall, Milton Keynes, MK7 6AA, United Kingdom.

Email: {S.Holland, D.R.Morse, H.Gedenryd} @open.ac.uk
Phone: +44 (0) 1908 653148, +44 (0) 1908 858463, +44 (0) 1908 659542
Fax: +44 (0) 1908 652140

Abstract

In this paper we describe a prototype audio user interface for a Global Positioning System (GPS) that is designed to allow mobile computer users to carry out a location task while their eyes, hands or attention are otherwise engaged. Audio user interfaces for GPS have typically been designed to meet the needs of visually impaired users, and generally (though not exclusively) employ speech-audio. In contrast, our prototype system uses a simple form of non-speech, spatial audio. This paper analyses various candidate audio mappings of location and distance information. A variety of tasks, design considerations, technological opportunities and design trade-offs are considered. The findings from pilot evaluation experiments are reported. Finally, opportunities for improvements to the system and for future empirical testing are explored.

Keywords: *Audio, Non-speech audio, Spatial audio, Global Positioning System, GPS, Sound, Navigation, Minimal Attention User Interface*

Introduction

Recent empirical work on mobile computer usage has argued that the GUI-based interaction style currently embodied in most mobile devices is inappropriate in many situations in which mobile computers are used [1]. Such situations are often characterised by:

This paper appeared as Simon Holland, David R. Morse and Henrik Gedenryd: AudioGPS: Spatial Audio Navigation with a Minimal Attention Interface. Personal and Ubiquitous Computing 6(4): 253-259 (2002).

- interactions with the real world being more important than interactions with the computer, so the user has a strictly limited attention capacity for the computer;
- the user's hands being used to manipulate physical objects;
- the user being involved in tasks that demand a high level of attention (to avoid danger to the individual as well as to engage with the task at hand);
- users being highly mobile, and adopting a variety of positions and postures during the task;
- the user's interactions with the environment being context dependent [2];
- interactions with the computer being high speed and driven by the external environment [1, 3].

In order to overcome some of the perceived drawbacks of the Graphical User Interface, both Kristoffersen [4] & Pascoe [5] proposed alternative user interaction styles aimed at reducing the amount of user attention, but not necessarily the amount of interaction, that is required to perform a particular task. The intention of both interaction styles was to transfer interactions to modes that take less of the user's attention away from their current task. Kristoffersen [4] proposed a technique called MOTILE that is based on the three principles of (i) little or no visual attention, (ii) structured, tactile input, and (iii) the use of audio feedback. In contrast, Pascoe [5] proposed a Minimal Attention User Interface (MAUI), linked to context-awareness that successfully reduced the amount of attention that a fieldworker had to pay to their mobile computing device [5].

Both MOTILE and MAUI were developed to support fieldwork applications, where a principal requirement is data capture through making observations on the local environment, be it on a ship [4] or a game reserve in Kenya [5]. A second situation in which a MAUI would be useful is in navigation. Whatever mode of transport is used, frequent reference must be made to the navigation aid — usually a map or written instructions. This takes one's attention away from the main task (moving through the local environment), which can result in accident, particularly when driving a vehicle. Therefore, one of the primary requirements for navigation aids is that they be low-attention, which has driven a long tradition of research into electronic navigation aids, particularly for the visually impaired [6].

In practical terms, a MAUI navigation aid would generally use audio as a primary means of communication with the user. Since the deployment of satellite-based navigation networks (the American Global Positioning System (GPS) and the Russian GLONASS networks), and with recent advances in speech technology, the majority of navigation aids have been based on these two technologies. Such navigation aids include the Personal Guidance System [6, 7], the Personal Navigation System [8], the Mobility of Blind and Elderly People Interacting with Computers system (MOBIC) [9] and a navigation system for the blind that uses a GPS and a mobile telephone connection [10]. These devices have the dual function of supporting navigation whilst also informing the user about relevant landmarks of interest to them. Similar applications that have been developed with sighted people in mind — electronic tourist guides — have become the canonical context-aware demonstration application [11, 12].

The AudioGPS

In recent years, handheld GPS systems (such as those manufactured by Magellan, Trimble and Garmin) have ceased to be the preserve of amateur mariners and are now being used in all walks of life. For most people, GPS systems are used to find out where you are now. The small screen of most handheld GPS units present users with the same usability problems that were found in the fieldwork applications quoted earlier that led to the development of the Minimal Attention User Interface concept.

In GPS systems, the core information to be communicated to the user typically takes the form of time-varying spatial information and relatively simple quantitative information. Given this, GPS systems are good candidates for investigating the design issues arising in minimal attention user interfaces. As noted above, GPS navigation systems for the visually impaired have been studied by several authors [6, 7, 8, 9, 10]. By contrast, in this paper, we consider a minimal attention user interface for the sighted person who is simultaneously involved in other demanding tasks. The two situations have somewhat different needs. For example, an audio GPS system for visually impaired users needs to be able to communicate a wide range of information through the auditory channel to allow the GPS to be programmed, and to communicate many kinds of information to the user about their surroundings. Given the potential complexity of this information, speech audio is almost essential. For the sighted but minimally attentive user, a simpler interface which requires less attention can be considered, namely a non-speech spatial audio interface. We do not wish to overload the user by using a communications medium that places a large processing and attention burden on the user. In addition, we do not want to compete with the voice channel, which might be used for other communication tasks. For example, while walking, the user might be engaged in conversation with companions, either in person or via a cell phone. Such a system could be a useful component of a more general mobile user interface — albeit for sighted users — in hazardous or demanding situations; for example, when it is dark, raining, when hands are full of shopping, or when there are children to be supervised. In this paper we consider some of the issues involved in the design of such a system and report on a prototype implementation.

The basic concept that has guided us relies on a virtual acoustic display as the user interface. This display takes an audio signal and transforms it into a binaural signal that the user listens to through headphones. The result is a sound that appears to emanate from a given environmental location. Such a system would be most useful in assisting a user to navigate from their current position towards their destination. One assumption is that the route is navigable by proceeding more or less in the direction of the destination. For destinations where access is impossible without finding unique intermediate access points (such as bridges), sequences of intermediate waypoints have to be established - typically before undertaking the journey. Yet more sophisticated ideas have been considered by Moon [9].

Audio representations of direction and distance

When using a GPS system for navigation, there are generally two essential pieces of information that have to be communicated to the user. Firstly, the distance to the destination (or intermediate waypoint). Secondly, the direction in which the destination lies, relative to the current direction of movement or relative to the direction in which the user is facing. With a graphical user interface, the distance typically would be displayed numerically and the direction would be displayed by a compass bearing or arrow. In our prototype AudioGPS, we use the following audio representations of direction and distance.

Direction

To a first approximation, the bearing to the next waypoint from the current position (relative to the current direction of motion) is indicated by the simple means of panning a sound source representing the destination across the stereo sound stage. Later prototypes are planned to incorporate a more capable 3D audio rendering system but the current prototype has the positive benefit of permitting an investigation into the usefulness of a simple, computationally inexpensive audio rendering resource. The sound source used is a briefly repeated tone, which we will describe in more detail in a moment. With simple panning, it is not hard to distinguish between sound sources at the extreme left, extreme right, straight ahead and at some intermediate points. However, panning the sound does not help distinguish between a sound source placed in front versus one placed behind the listener. Indeed, even relatively capable 3D audio systems employing more precise generalised Head Related Transfer Functions can suffer from similar confusion. As a pragmatic way around this perceptual problem, we use a sharp tone (for the purposes of the prototype a harpsichord sound) when the destination is in the semicircle in front of the user, and a relatively muffled tone (a muted trombone sound) when the destination is located in the semicircle behind the user.

Ideally, the bearing of a destination should be computable even when the user is stationary, and should be heard to pan around the head if the user revolves on the spot. This is realisable using a flux magnetometer (an electronic compass) in addition to a GPS. Currently the AudioGPS does not incorporate an electronic compass. Hence, the AudioGPS, just like most visual GPSs, has no way of knowing which way the user is facing. However, provided the user is moving, or has recently been moving, the direction of motion can be calculated. Just as with its established visual equivalent, the AudioGPS uses the convention that if the user faces in the direction of travel, the destination will be rendered in the correct direction.

By these means, the location of a destination at the four cardinal points left, right, ahead and behind is relatively straightforward to communicate. In addition, it is generally possible to perceive if a sound source is at some intermediate point, giving at least eight compass points. If a better angular precision than eight cardinal points is required then more affordances are needed. As an experimental way of addressing this problem, we have employed a ‘chase tone’, which works as follows. A repeated tone at a fixed pitch, but appropriately panned in space, always indicates the destination. A second tone, which we call the chase tone, coincides exactly in pitch with the first tone if the destination is straight ahead, but progressively diverges in pitch to a maximum pitch difference of one octave as the angle of the destination reaches a maximum of 90^o from the path angle. Both tones are always spatially located together. Rather than using a chase tone whose pitch varies continuously, we have used a chase tone that varies its pitch discretely and chromatically. We considered using two tones on a continuous scale but in pilot experiments, we found that the chromatic scale was easier to interpret than the continuous scale. More extensive laboratory and field experiments are needed to find out the usefulness of this mapping, taking into account individual differences between users.

Distance

We have been experimenting with several representations of distance. Our primary representation is a “Geiger counter” or “hot / cold” metaphor, where the number of pulses of sound, together with their rapidity, gives an indication of how far it is to the next waypoint. If the user is a long way from the waypoint then few, widely spaced sound pulses are emitted. If the user is close to the waypoint, then the user hears many, closely spaced, sound pulses. This metaphor is easy to assimilate using the familiar Geiger counter idea. In addition, on arrival, a distinctive arrival tone is emitted. This is the mapping we have used most to date. It is similar to the system used on some vehicle reversing aids, which are intended to avoid collisions when reversing a vehicle; as an obstacle is approached, the warnings become more strident and insistent.

At a predetermined distance from the destination, the system emits an arrival tone to indicate that the user has reached the destination — within the limits of the precision of the GPS system that is being used to support navigation. The radius of the arrival zone can be set by the user; for example, different radii are applicable to users moving at different speeds, such as pedestrians and motorists. If a differential GPS system is used (which gives a more precise estimate of location than an uncorrected GPS system) then the arrival zone can be smaller than if an uncorrected GPS system is used (as in the current prototype).

Alternative encodings of distance

An alternative approach to representing distance to a destination could use a discrete analogue mapping to an appropriate scale, such as five clicks in each run of clicks indicates fifty metres, whereas one click indicates ten metres. Martyn Cooper [pers. comm.] has referred to this as an “inverse Geiger counter” mapping because when the user is a long way from the destination, clicks are emitted more rapidly. Different scales are appropriate to different situations - pedestrians versus car users, for example. As in the Geiger counter metaphor, a different scale might be needed for the last part of a very long journey. Hierarchical versions of such a mapping are possible to permit greater precision over a wider range of distances.

Alternative direction mappings

The direction mapping in the current prototype encodes equal changes of angle with equal pitch differences in semitones. An alternative mapping has been considered that encodes angles of particular significance, such as 30° 45° and 60° to pitch intervals considered (by musicians) to be of particular prominence: the perfect fourth (five semitones), tritone (six semitones) and perfect fifth (seventh semitones) respectively. This non-linear mapping might help some users to recognise those particular angles, but seems unlikely to help other users, or to help with other angles. However, its effectiveness is a matter for empirical investigation.

As already noted, an alternative mapping to the discrete pitch mapping that we use at present would be to use continuous frequency mapping, within an octave, to encode direction. In theory, this might offer higher resolution than the current mapping, but it is not clear that a higher resolution would be perceptually available to most listeners. Similarly, it is not known whether this mapping makes it easier to discern the direction of change of angles, since changes in pitch at any given time tend to be smaller. The constraint of mapping within an octave could be dropped, but then it would be harder to find a 'natural' mapping for a 90° pan. This is not a mapping we recommend, but finding the most useful mappings requires further experiments in the laboratory and in the field. The above discussions emphasise the extent to which good audio mappings of direction and distance depend upon the task and the situation in which they are used. In conclusion, we anticipate that the most suitable sonification will vary between people, tasks and situations.

Issues for spatial audio GPS

Timbre choice

In the current prototype, conventional musical instrument timbres are employed to represent the target. In a more refined system, timbres more specifically tailored for the purpose could be used. However, some characteristics of the timbres in the current prototype were chosen according to principles applicable whatever the timbres, as we will now explain.

A harpsichord tone was chosen to suggest destinations ahead, and a trombone for destinations behind the user. The harpsichord timbre has a relatively high amount of energy in the higher frequencies compared to the trombone timbre, for a tone of a given pitch. In addition, the amplitude onset envelopes of the harpsichord timbre is much sharper than that of the relatively muffled trombone timbre. These relative characteristics were chosen to suggest the 'rear' tone as a version of the 'forward' tone muffled by the user's own head. Users appeared to be able to distinguish between the forward and rear tones easily and not to confuse the two.

'Silence as no news', versus 'silence as success'

The current system uses silence when it has nothing useful to indicate, such as when there is no reliable signal or no current computed direction of movement. One interesting alternative design approach would be to use silence to indicate success. For example, silence might be used to indicate that the user is travelling in exactly the right direction towards the next waypoint, or, depending on the application, has arrived at the waypoint. This mapping has the merit that when the user is doing the right thing (i.e. moving in the right direction or arriving at the right place) there is nothing in the audio display to distract the user. Such a mapping might be useful in situations where the user is free to move in exactly the right direction, since when the user is doing the right thing, the system would be minimally distracting. Unfortunately, in many situations users may be unable to move directly towards their target for long periods, since the direct path is inaccessible. In such situations, the ‘silence as success’ mapping does little to minimise obtrusiveness. Another problem is that the mapping is prone to ambiguity in various situations, such as loss of GPS signal, loss of computed position or temporary masking of the audio signal. Therefore, ‘no news’ is liable to be confused with success and vice versa, unless new sounds are adopted specifically to signify the lack of news. Such new sounds might themselves be obtrusive when there is frequently no news to report.

In situations where a GPS signal will rarely be lost, when an electronic compass is available, where listening conditions are typically good for the user, where the reliability of all parts of the system are high and a direct path may be made towards the target, ‘silence as success’ may be an attractive strategy. In other situations, silence as no news seems more immediately useful, provided measures are taken to address obtrusiveness, as discussed in the next section.

Obtrusiveness

When travelling short distances, the information provided by the audio GPS typically changes rapidly, and provides rapid audible navigation towards the target. In such situations, the tones tend to be perceived as useful rather than obtrusive. In certain other situations, measures may be required to avoid, or minimise possible irritation.

When the path is long, straight and some distance from the destination, neither distance nor direction information is needed continuously. A simple volume control can be used in this situation, as in many similar cases, to avoid irritation. By contrast, in obstacle-strewn situations, information about distance to destination may not be required urgently, but information about direction may be required continuously. In such situations, the directional mapping may be maintained, while the distance information is temporarily suppressed by using a single tone per burst, irrespective of the distance to the target.

Implementation

A primary requirement for any navigation aid is portability. Notwithstanding that, our implementation platform was chosen for its development facilities and flexibility rather than the convenience with which it can be carried. As we explain in the Further Work section of this paper, we intend to experiment with a wide range of peripheral devices such as headtrackers, and to use extensive recording and logging software. Hence, we took a deliberate decision to use standard development environments, where the necessary specialised drivers are available, rather than the more obviously appropriate wearable or mobile computers. Therefore, we are using VisualWorks Smalltalk and Java, running on Apple Macintosh and Microsoft Windows laptop computers. We have equipped specialised rucksacks with active cooling to allow both hands to be kept free, and all of the necessary peripherals and data collection to be organised conveniently. Sound synthesis is carried out using software sound synthesis controlled via MIDI, which is available via a QuickTime API for both software platforms. We used ‘open’ stereo headphones that do not exclude background environmental noise because such noise provides important navigational cues and warnings to users (figure 1).

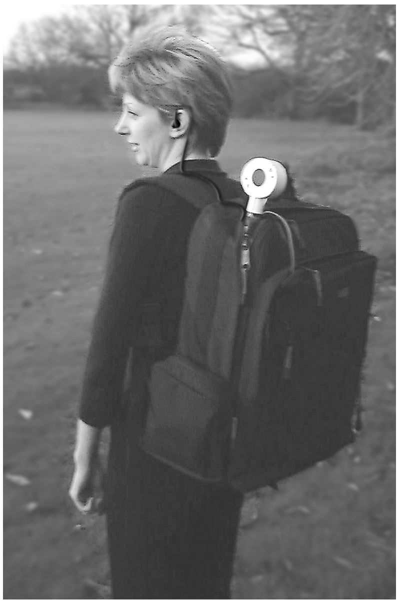

Figure 1. The backpack for running the prototype AudioGPS system and collecting experimental data. The fan-cooled, temperature-monitored backpack contains a laptop connected to a GPS system. The snorkel sticking out of the backpack doubles as a filtered air inlet for the cooling fan and a mounting point for the GPS antenna to optimise GPS signal quality. The laptop is wireless networked to allow it to be monitored remotely by an experimenter when appropriate.

User trials

We have performed two informal pilot user trials of the AudioGPS, which were designed to evaluate it in the following areas.

i. **Audio representations of direction and distance.** We have carried out a formative evaluation of the audio representations of direction and distance both with musicians (who might be expected to have a good grasp of the principles upon which the audio representations are based) and with non-musicians. We found that the distinction between being in front, behind, to the right and to the left of the user (represented by changes in the tone and placement of the sound on the stereo stage) was easy to discern, as were the intermediate four compass points. In other words, it is possible to perceive the direction of the sound source.

ii. **Does the system work under field conditions?** Informal tests of whether the system works under field conditions were carried out on foot at night and in a car. In the former experiment we were primarily concerned with finding out how long it took the AudioGPS to provide usable navigational information after having been switched on. (It takes time for any GPS system to acquire “fixes” from the orbiting satellites and use these to calculate its location, velocity and direction of movement.) Secondly, we were interested in how responsive the system was to directional changes in the path that the user was taking. Typically, it took approximately twenty seconds (or twenty metres while walking) for the system to initialise and start reporting usable bearing information. Subsequent changes in direction took about half that time.

In summary, the system was useable for finding targets on foot. The spatial audio interface was used for the navigation part of the task, not for setting up destinations. Our preliminary experiments with in-car use of the AudioGPS demonstrated that the system was not responsive enough to cope with direction changes faster than every 10 seconds or so (while driving around a car park). However, it worked adequately once the GPS system had "settled down" while driving along the open road.

Further work

With the system as currently implemented, we intend to find out how individual differences affect the effectiveness of the audio mappings we have adopted. We will also find out how well the system can be used by sighted users to find unknown locations on foot at varying distances. We will also experiment with other forms of locomotion. At the same time, we are upgrading the system to use a flux magnetometer (electronic compass) and a rapidly reacting head tracker. The head tracker is expected to improve the discrimination between front and rear, and angular resolution, which we plan to test in field trials. Finally, the system will be augmented sparsely with speech audio.

Summary

Audio GPS systems have a very important role to play for visually impaired users, but they have a slightly different role for mobile sighted users in situations of minimal attention, and where conversations may compete with speech audio information. In contrast to many GPS systems, we have focussed on spatial non-speech audio rather than speech audio. Preliminary trials have suggested that very simple and computationally inexpensive spatial mappings are surprisingly good at helping users to find locations, irrespective of musical training.

Acknowledgements

Thanks to Steve Brewster for hiking from Willen and many tips, Jim Ballas for some basics, Dave Haniff for drivers, Kristina Höök for silvan elks; Sharon Powell, Nicky Kempton and Mike Richards for the photo session; and John Winberk and Georg Strom for useful comments.

References

[1] S. Kristoffersen and F. Ljungberg, “"Making place" to make IT work: empirical explorations of HCI for mobile CSCW,” presented at Proceedings of the International ACM SIGGROUP conference on Supporting group work (GROUP'99), Phoenix, AZ USA, 1999.

[2] M. Weiser, “The Computer For the 21st-Century,” *Scientific American*, vol. 265, pp. 66-75, 1991.

[3] J. Pascoe, N. Ryan, and D. Morse, “Using While Moving: HCI Issues in Fieldwork Environments,” *ACM Transactions on Computer Human Interaction*, vol. 7, pp. 417-437, 2000.

[4] S. Kristoffersen and F. Ljungberg, “Designing Interaction Styles for a Mobile Use Context,” in *First International Symposium on Handheld and Ubiquitous Computing (HUC 99)*, vol. 1707, *Lecture Notes in Computer Science*, H.-W. Gellersen, Ed. Karlsruhe, Germany: Springer-Verlag, 1999, pp. 208-221.

[5] J. Pascoe, N. Ryan, and D. Morse, “Issues in developing context-aware computing,” in *First International Symposium on Handheld and Ubiquitous Computing (HUC 99)*, vol. 1707, *Lecture Notes in Computer Science*, H.-W. Gellersen, Ed. Karlsruhe, Germany: Springer-Verlag, 1999, pp. 208-221.

[6] J. M. Loomis, R. G. Golledge, and R. L. Klatzky, “GPS-based navigation systems for the visually impaired,” in *Fundamentals of wearable computers and augmented reality*, W. Barfield and T. Caudell, Eds. Mahwah, New Jersey: Lawrence Erlbaum Associates, 2001, pp. 429-446.

[7] J. M. Loomis, R. G. Golledge, and R. L. Klatzky, “Navigation system for the blind: Auditory display modes and guidance,” *Presence-Teleoperators and Virtual Environments*, vol. 7, pp. 193-203, 1998.

[8] G. Moon, “Personal Navigation Systems for the Blind,” in *Institute of Engineering Surveying and Space Geodesy*. Nottingham: University of Nottingham, 1998, pp. 328.

[9] H. Petrie, V. Johnson, T. Strothotte, A. Raab, S. Fritz, and R. Michel, “MOBIC: Designing a travel aid for blind and elderly people,” *Journal of Navigation*, vol. 49, pp. 45-52, 1996.

[10] H. Makino, I. Ishii, and M. Nakashizuka, “Development of navigation system for the blind using GPS and mobile phone combination,” presented at Proceedings of the 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol 18, Pts 1-5, 1997.

[11] K. Cheverst, N. Davies, K. Mitchell, A. Friday, and C. Efstratiou, “Developing a Context-aware Electronic Tourist Guide: Some Issues and Experiences,” presented at CHI 2000: The Future is Here, The Hague, The Netherlands, 2000.

[12] G. D. Abowd, C. G. Atkeson, J. Hong, S. Long, R. Kooper, and M. Pinkerton, “Cyberguide: A mobile context-aware tour guide,” *Wireless Networks*, vol. 3, pp. 421–433, 1997.