Extending the Auditory Display Space in Handheld Computing Devices

Ashley Walker and Stephen Brewster

Glasgow Interactive Systems Group, Department of Computing Science,
University of Glasgow, Glasgow, G12 8RZ, UK.
Tel: +44 (0)141 330 4966 Fax: +44 (0) 141 330 4913
Email: stephen, ashley@dcs.gla.ac.uk Web: www.dcs.gla.ac.uk/~stephen/

ABSTRACT

An important challenge that presents itself to designers of mobile computing devices is: how can interfaces for small screens be structured to present users with opportunities for efficient interaction with a plethora of (visually overwhelming) materials such as telecommunications, calendars and contact information, and the Web? We propose a low-cost and immediately implementable solution to the problem in the form of three-dimensional (3D) audio display. This solution introduces an expansion of the notion of "interface" by extending the user's perception-action space beyond the dimensions of the display to encompass a virtual 3D auditory space surrounding the user. In this paper we present a high level outline of our 3D audio windowing system and its associated suite of utilities for exploiting 3D space in information display.

INTRODUCTION

With the rapid growth of networked and integrated computing devices, a user's interaction with his/her computing device is becoming increasingly multi-tasking: while immersed in a foreground task, the user is typically engaged in monitoring multiple simultaneously running background tasks with widely varying response times. For the small screen mobile computer user, multi-tasking (and utilisation of remotely networked resources) offers the possibility of much richer computing experience; however, existing interface architectures do not support efficient interactions of this sort. Presently, information-access rates are strongly limited by screen size and, as most manufacturers of modern computing devices are aimed at minimising device size, this (visual) information bottleneck looks likely to tighten.

The prime motivation underlying our work is to overcome this information-access rate barrier through the development of alternative modality interfacing tools. This paper takes a high level look at a new project which combines rapidly developing 3D audio tools with tried and tested graphical user interface (GUI) techniques for exploiting space in information representation.

Current research

Our work so far has focussed on how the limitations of display size can be minimised by the addition of structured non-speech sounds [1, 2]. One problem with mobile devices is that they have a limited amount of screen space: the screen cannot be large as the device must be able to fit into the hand or pocket to be easily carried. As the screen is small it can easily become cluttered with information as designers try to cram on as much as possible. In many cases desktop widgets (buttons, menus, windows, etc.) have been taken straight from standard graphical interfaces (where screen space is not a problem) and applied directly to mobile devices. This has resulted in devices that are hard to use, with small text that is hard to read, cramped graphics and little contextual information.

One way to solve the problem is to substitute non-speech audio cues for visual ones. Sound can be used to present information about widgets so that their size can be reduced. This would mean that the clutter on the display can be diminished and/or allow more information to be presented. Results from two studies on the 3Com PalmIII [1, 2] showed that the usability of large and small on-screen buttons can be improved by the addition of simple sounds. This includes more data being entered and workload reduced when using the sonically-enhanced buttons.

The limitation of this approach is that the auditory display space on mobile devices is small. Devices usually have just a single loudspeaker so sounds can only be presented as coming from a single point in space. The problem arises that the auditory display space can itself become cluttered if too many sounds are presented at the same time. One way to solve this would be to incorporate more speakers or allow the user to wear headphones. Another speaker would add cost and weight, and could not be positioned far enough from the other to give good stereo separation. Headphones can provide stereo and full three-dimensional (3D) sound. The disadvantages are that users are tied to the device by a cable and their ears are covered. However, users commonly wear headphones with personal stereos, 'hands-free' kits are also common with mobile telephones and wireless headphones are becoming more widely available. We are using this latter approach in our investigations of sound in mobile devices.

BACKGROUND: SPACE IN THE GUI INTERFACE

Windows into the global information space

GUI interfaces evolved to their present level of sophistication over several generations. Much of the richness of the interactions that they afford are so natural and simple that we take them for granted (until forced to compare them with the vastly less sophisticated interfacing tools available to the other senses). Perhaps the most powerful exploitation of space in a modern GUI interface lies in the use of spatial partitioning to facilitate multi-tasking. By presenting distinct sources of information in distinct visual packages, or windows, the user is able to interact easily with multiple sources of simultaneously available information. Users can arrange these windows spatially so as to reduce information clutter and offer themselves affordances for particular interactions. A commonly occurring example of the latter is the positioning of windows to mirror the priority level of their contents: low priority tasks may be placed in iconized windows, medium priority (e.g., monitoring) tasks may be placed in one or more windows around the periphery of the screen, while a select few high priority tasks are placed in the centre foreground to continually draw the user's attention back to their contents.

The use of space in window utilities

The GUI's use of space does not stop at the level of window partitioning of distinct information sources. Within windows a variety of carefully evolved spatial mappings are used to facilitate interaction with window content. For example, at the heart of a GUI windowing display is the window scrollbar whose spatial layout allows users to determine rapidly the size and structure of information contained within a window. Spatial representations are exploited in messaging semantics as well. The progress bar, for example, efficiently communicates the status of multiple download parameters (e.g., file size, transfer rate, percentage transferred, etc.) in the movement of a simple icon across a 2D visual axis.

The principle underlying these mappings is that a spatial representation of temporally extended information affords a more natural interaction with that information because search and recall can be performed more easily on spatial --- as opposed to temporal --- memory. Whether the context of space is real or virtual, space is frequently recalled and intimately associated with the recognition of events, and, furthermore, enhances the memory of those events [10, 13].

No remotely equivalent audio windowing interface exists. To date, most audio rendering tools such as text-to-speech translators (currently the fastest and most natural facility for making text/graphical information perceivable) collapse information from a variety of concurrently operating windows into a single serial stream of sound. This imposes a usability bottleneck which, like the visual bottleneck imposed by overcrowded screens, must be overcome in order for users to enjoy efficient multi-tasking interactions. We advocate utilising rapidly developing 3D audio tools to increase the display bandwidth of mobile computing devices.

Audio windows into the global information space

The solution which we are pursuing utilises the spatialisation of sound sources to expand and repartition a single audio stream into multiple spatially segregated streams of information --- i.e., *acoustic windows* --- which, like visual windows, each present information from a unique spatial position. In the same way that the position of a visual window can be used to disambiguate its contents from that of other windows, so can the position of a sound source be exploited to disambiguate its contents from other temporally overlapping audio streams in a 3D auditory display.

In the acoustic domain, this perceptual phenomena, known as the *cocktail party effect* [4], has been used successfully in several specialist audio applications underlying multi-media communications [5, 6, 7, 8, 9, 12, 15].

To facilitate this audio equivalent of gaze (or attention) control, audio windowing systems must provide the user with facilities for controlling the audibility of a source. We facilitate attentional control by wrapping volume adjustment features into the spatial metaphor itself. As outlined below, listeners can manipulate the relative position, orientation and size of sources in their sound field in order to change the perceptual saliency of these sources. This technique has been proved successful in several other systems [6, 7].

Because mobile device users are increasingly likely to be used for browsing activities (e.g., simultaneously skimming several broadcasts or Web information), we also provide simple attention-focusing facilities. In this scheme, the signal strength enhancement of the preferred source fades over time to allow other sources to gain the user's interest. Similar facilities --- along with facilities for marking story interest-boundaries with non-speech signals --- have been shown to be effective in the *AudioStreamer* audio-only browsing environment wherein users selectively listen to one of three simultaneously presented news broadcasts arrayed in a horizontal plane about their head [12]. The authors of AudioStreamer report that such a feature is necessary because when focused on one source, users are conscious of very little sound leaking through from the other channels. A single-minded user who does not want to be tempted to browse in this way can overcome this effect by making consecutive indications of interest in a particular source. This action bumps up the fade delay constant and volume of the preferred source.

Manipulating audio windows

Within an audio windowing interface, users must be given the same sort of environmental control available in GUIs: the ability to spatially position windows in ways that facilitate efficient interaction with their contents. Within desktop VR/communication systems, this is most commonly accomplished via gesture control (e.g., head movements [8, 12] or hand movements [6]). In small, lightweight system such as ours --- where hardware support for these facilities cannot be provided --- graphical tools will be given to the user.

Users can employ their normal graphical interaction tool (e.g., a mouse or stylus) to move, re-orients and/or re-size sound sources within a miniature egocentric graphical display of the soundscape. Moving and/or reorienting the icon associated with a particular sound source alters the volume of that source in the same way as walking and turning towards it would in the real world. (Source directionality is under the user's control such that each source can be made to radiate sound omni-directionally or in a tightly focused beam which can be best heard when the user is positioned orthogonal to it.) Users can adjust source volume without changing the spatial layout of sources by resizing individual sources. Similarly, users can adjust the overall system volume by enlarging/decreasing the iconic representation of their own head. Variants of these features have proved successful in *MAW* (multidimensional audio windows) --- an interactive teleconferencing front-end which exploits the cocktail party effect by allowing users to graphically manipulate the directional characteristics of multiple sources, or callers [6].

As most users will find graphical manipulation facilities to be less natural than gestural control, it is important that mouse/stylus actions are tightly coupled with changes in the soundscape. To achieve this effect, sound sources are filtered while undergoing direct graphical manipulation in order to make changes more salient. Amongst the tools that will be used for this purpose is a *spotlight filter* and a *muffler* [6]. The former works like a visual highlight and shifts source pitch without affecting its volume and is used to confirm, for example, selection of a window (as a prelude to invoking some action). Similarly, a source being moved is muffled, or low pass filtered, to distinguish it as the one being handled.

The use of audio space in window utilities

A host of tools have already been developed for rendering visual and text based information in audio [11]. Moreover, a plethora of audio source materials (e.g., news and entertainment broadcasts) exists on the Internet. In order to afford users a more natural interaction with this information, audio content visualisation tools are required. Furthermore, because a network-based windowing system must support indirect interactions (such as task monitoring) as well as direct creation and consumption of content, tools for sonifying background/monitoring tasks are needed. Here we discuss two tools currently under construction.

Foreground tasks: browsing, scrolling, reading

In specialist domains, audio browsing tools exist to facilitate interaction with audio-format data. For example, the *Audio Browser* is a hierarchical sound file navigation and audition tool which allows a user to move virtually through an archive of sound clips whilst simultaneously listening to the clips nearest to the present position [15]. Kobayashi and Schmandt's *Dynamic Soundscape* [8] --- an audio browsing tool --- re-maps temporally extended data (e.g., a news broadcast) onto a spatially extended audio display. Here a virtual newsreader orbits a user's head such that different topics are played at different spatial positions. If a user wants to review any segment of the broadcast, s/he need not rely on temporal recall of its sequence in the audio stream but, rather, simply the position where the topic of interest was heard.

Along these same lines, we are currently developing an acoustic scrollbar. It works by assigning each audio source (e.g., a series of audio rendered documents) to a user-defined spatial axis, and playing each subtopic from a spatial position proportional to its temporal position within the data source stream (see Figure 1). As in Dynamic Soundscape, users can rapidly revisit material and select material in the audio stream by simply indicating the corresponding position along the spatial display axis. As an extension to these scroll facilities available in Dynamic Soundscape, we will allow users to control the movement resolution such that a source may continuously move during playback (i.e., each new utterance is interpreted as a new topic) or jump through a series of N discrete positions during playback --- where N corresponds to the number of subtopics in a document. In the latter case, N may be set to one to anchor playback to one position.

Background tasks: monitoring progress

We have developed and tested a spatialised audio tool for monitoring limited-lifetime processes (e.g., a downloads) [14]. Its design was borrowed from that of the visual progress bar whose simplicity and clarity is derived from the use of two basic components: (i) a progress indicator which moves along (ii) a fixed reference axis. The corresponding spatialised audio progress bar was, therefore, built using only two spatialised sound cues --- the first provides the reference (i.e., it is played from a fixed target position located in front of the user) and the second component is spatialised to indicate the percentage of the task complete by its angular position within a circular orbit centred on the user's head (see downloads 1 and 2 in Figure 1). As with the visual progress bar, a full set of delay affordances are salient as the movement rates and relative position of these two components. Tests with this spatialised, audio-only versus a visual progress bar showed that the former lead to improved accuracy in a background progress monitoring task as well as improved efficiency in a simultaneously conducted foreground task.

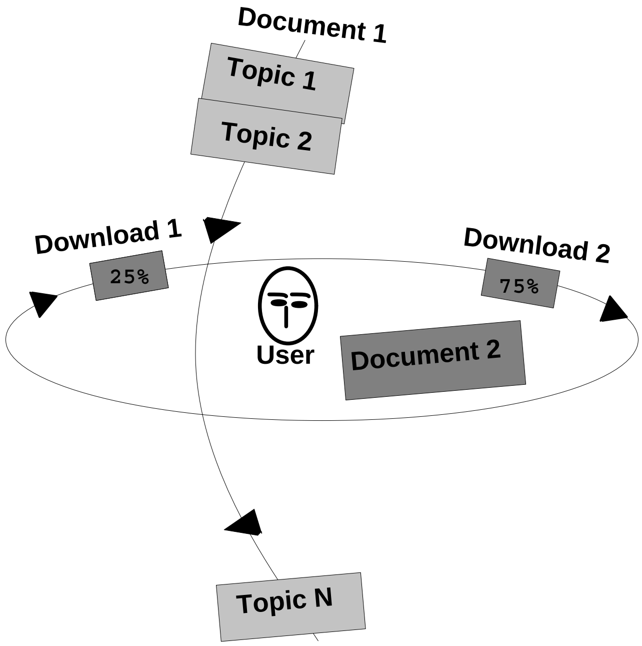

Figure 1: Auditory display space. Here two progress (e.g., download) monitoring windows and two content windows are shown as small and large boxes, respectively. The percentage complete of the two download processes are shown as the spatial position of their audio windows within the circular orbit centred on the user's nose (starting and ending in front of the user's nose). The contents of Document 1 is scrolled in an alternative plane --- as illustrated by the large, light grey boxes which play the topical contents of Document 1 from a series of different elevations. (Once the whole broadcast has been played, the user can return to, e.g., Topic 2, by pointing up 60 degrees.). Document 2, currently being generated by the user (large, dark grey box), is played back from a constant front-left position.

Spatialisation: creating the 3D audio space

The facilities discussed here are general in that they are not tied to any particular form of sound spatialisation technique (sapatialisation is the technique used to simulate 3D sound through headphones). As tools for spatialising sound improve in speed and resolution, so will the facilities discussed here. Our pilot applications utilised a suite of low-cost digital convolution filters based on head-related transfer functions measured at San Jose State University [3]. Currently, we are interfacing audio windowing software with Microsoft's DirectX spatialisation routines. As the technology matures, we will investigate resource conservation solutions which seek not to allow the maximum number of spatialisable sources to limit the number of actual sources available to the user. In this case, users can aggregate sources so that that together they utilise one spatialisation filter and share a unique timbre.

CONCLUSIONS

As screen size drops and interactions become increasingly network-based, mobile device users will experience increased difficulty in extracting the information they need to enjoy efficient interactions. We have presented a method for increasing the bandwidth of mobile device "displays" by presenting information in the 3D auditory display space surrounding the user. Whereas most existing audio interface tools simply present information as a temporally extended stream of sound, we advocate the use of multiple spatially segregated streams, or windows, of sound. The output bandwidth of the unspatialised approach is very low because it is hard to attend to more than one synthesised voice or sound stream at a time --- even if speech is presented at a rapid rate, such an acoustic display does not provide anything like an equivalent speed and ease of use to the graphical counterparts. Moreover, temporally extended representations of acoustically rendered information are difficult to search, skim and recall. By contrast, a 3D audio windowing system --- consisting of both local and global window managing facilities analogous to those existing in a GUI --- can support much of the same efficiency of interaction as a modern GUI. In summary, an audio user-interface of this sort has a number of benefits, including:

Increase in display bandwidth: A spatialised audio display provides a mapping of information from a single, temporally-extended sound channel onto a more naturally recalled and indexed spatial display consisting of multiple spatial channels.

Increase in display area: A spatialised audio display affords a large and salient information space which extends beyond the dimensions of a monitor to encompass the full 360 degree acoustic sphere surrounding a user. This will allow complex information to be presented in sound without cluttering the auditory display space.

Compatible with existing graphical user interfaces Spatialised audio windows could exist as part and partner to existing window-based graphical user interfaces. Such a multimodal interface could be configured by each user to suit his/her individual needs and preferences for audio versus visual presentation.

*Readily implementable using existing technolog*y: The maturation of technology for generating spatialised audio has far out-paced its commercial use outside the entertainment industry. The novel use of 3D audio proposed here provides a unique example of how the information bearing capabilities of 3D audio can be employed in interface design.

ACKNOWLEDGEMENTS

This work was funded by EPSRC Grant GR/L79212, GR/L66373 and a Microsoft Active Accessibility Grant.

REFERENCES

All references by the authors can be found at: http://www.dcs.gla.ac.uk/~stephen/publications.shtml

1. Brewster, S.A. (1999). Sound in the interface to a mobile computer. *To appear in Human-Computer Interaction International'99*, Munich, Germany.
2. Brewster, S.A. & Cryer, P.G. (1999). Maximising Screen-Space on Mobile Computing Devices. *In Summary Proceedings of ACM CHI'99*, Pittsburgh, PA: ACM Press Addison Wesley, pp. 224-225.
3. Brown, C.P. (1996). *Modelling the elevation characteristics of the head-related impulse response* (Technical Report No. 13). San Jose State University.
4. Cherry, E.C. (1953). Some experiments on the recognition of speech. *Journal of the Acoustical Society of America*, 25, pp. 975-979.
5. Cohen, M. (1993). Integrating graphical and audio windows. *Presence: Tele-operators and Virtual Environments*, 1(4), pp. 468-481.
6. Cohen, M. (1993). Throwing, pitching and catching sound: audio windowing models and modes. *International Journal of Man-Machine Studies*, 39, pp. 269-304.
7. Cohen, M. & Ludwig, L.F. (1991). Multidimensional audio window management. *International Journal of Man-Machine Studies*, 34, pp. 319-336.
8. Kobayashi, M. & Schmandt, C. (1997). Dynamic soundscape: mapping time to space for audio browsing. In S. Pemberton (Ed.), *Proceedings of ACM CHI'97*, Atlanta, GA: ACM Press Addison Wesley, pp. 194-201.
9. Ludwig, L.F., Pincever, N. & Cohen, M. (1990). Extending the notion of a window system to audio. *IEEE Computer*, August, pp. 66-72.
10. Mandler, J.M., Seegmiller, D. & Day, J. (1977). On the coding of spatial information. *Memory and Cognition*, 5.
11. RNIB (1999). *RNIB Homepage* (http://www.rnib.org.uk/wedo/research/).
12. Schmandt, C. & Mullins, A. (1995). AudioStreamer: Exploiting simultaneity for listening. In I. Katz, R. Mack, & L. Marks (Ed.), *ACM CHI'95 Conference Companion*, Denver, CO: ACM Press Addison Wesley, pp. 218-219.
13. Schulman, A.I. (1973). Recognition memory and the recall of spatial location. *Memory and Cognition*.
14. Walker, V.A., Brewster, S.A., Trading space for time in interface design. *Submitted to Brewster,S., Cawsey, A. and Cockton, G. (Eds.), INTERACT'99*, Volume II, Edinburgh, UK.
15. Whitehead., J.F. (1994). An audio database navigation tool in a virtual environment. *In Proceeding of the International Computer Music Conference*, pp. 280-283.