SONICALLY EXPOSING THE DESKTOP SPACE TO NON-VISUAL USERS: AN EXPERIMENT IN OVERVIEW INFORMATION PRESENTATION

Louise Valgerður Nickerson and Tony Stockman

Department of Computer Science,
Queen Mary College, University of London,
Mile End Road, London E1 4NS, United Kingdom
{lou,tonys}@dcs.qmul.ac.uk

ABSTRACT

The vast majority of computer interfaces do not translate well onto non-visual displays (e.g. for blind users, wearable/mobile computing, etc). Screen readers are the most prevalent aural technology to expose graphical user interfaces to the visually impaired. However, they eliminate many of the advantages of direct manipulation and WYSIWYG applications. While the use of sound in interfaces has become more prevalent due to advancement in sound cards for computers, it is still primarily for alerts and status-reporting. The use of sound can be expanded to enhance or replace a GUI by providing a 3D auditory environment. However, users of this environment would need a reliable and effective method of navigation. Little is known of the usability of a system based on sound identification and localisation. In this work, we describe an experiment which will examine users' ability to navigate a 3D auditory environment based on these concepts.

1. INTRODUCTION

Current computer interfaces lack methods for providing general information to users in a non-visual manner. A visual user gets generalised information by glancing at their screen and thus are able to navigate more efficiently to accomplish goals. For the visually impaired, sound is crucial to providing access that the sighted take for granted. Screen readers and keyboard navigation of a Graphical User Interface (GUI) provide specific information about files and objects within applications. Currently, a user must already know what objects are present in the interface or spend considerable time exploring in order to navigate to a file. Our research will explore methods of providing overview information using 3D sound, focusing on the desktop.

We are investigating if 3D sound can be used to provide quick, generalised information about a user's system status. We try to make it easier to get an idea of what is there. That is, in the case of the desktop, where objects are located in the desktop space and what their general type is. It is intended to be a brief synopsis that does not provide specific information but rather enables the non-visual user to know where to look for information. In current tools for the visually impaired, there are few instances of overview display, the emphasis is on the specific rather than the general. We aim to shift this focus in order to lead to better navigational tools. The target audience is not only blind users but partially-sighted users, who are more likely to use tools such as screen-magnifiers, and users who may not always have access to a visual display.

Section 2 will discuss research in the area of auditory display that relates to providing non-visual interfaces. Section 3 will discuss the proposed work in more detail.

2. BACKGROUND

2.1. Interfaces for the Visually Impaired

2.1.1. Auditory Interfaces

Screen readers are the most common tool for computer interaction for blind users. The technology allows a user to navigate a computer via the keyboard and vocalises the contents of the screen. The support of non-speech sound in consumer products is relatively new (JAWS for Windows 2004). However in academic research, more progress has been made. The Mercator Project [1] developed an X Windows auditory interface that can be navigated with a mouse. Windows and widgets are monitored and provide feedback to the the user. The translation of the GUI to an auditory interface is two-dimensional and reactive: the user must explore or be familiar with the interface. More recently, Zhao et al. [2] describe the idea of "gists" in information seeking tasks. They outline an Auditory Information Seeking Principle where "gists" would give the user an impression, prior to interaction, of, in their case, American census data displayed on a map of the USA states, thus allowing the user to choose an geographical area to explore rather than wandering through the interface. This principle is similar to the overview idea presented in our work.

2.1.2. Auditory Icons and Earcons

Other research has been more specifically oriented towards working with objects within the interface. In order to navigate, Pitt and Edwards [3] used tone to show proximity to targets. The work showed that it was possible to lead users close to a target but they had a hard time fixing it. Directional as well as distance information is crucial to maximum efficiency. Earcons have also been shown to facilitate navigation [4]. The navigation in our work is of hierarchical systems: we will use earcons to represent files on the desktop but not as targets for navigation, though this may be addressed in future work.

2.2. 3D Spatial Audio

2.2.1. Sound Localisation

Spatial audio can be a powerful tool. Wenzel [5] provides a good overview of how spatial audio is interpreted and how it can be taken advantage of. Localisation of multiple sounds improves with larger delays between the sound to be located and distractor sounds [6]. Concurrency allows for greater efficiency as more information can be presented quickly, however, concurrent earcons are difficult to distinguish. It has been shown that if they are staggered in time and have differentiating timbre, identification is more accurate [7]. Additionally, locating the earcons spatially also improves identification [8]. For this reason, we are choosing to attempt serial representation of the desktop as well as diverse spatial location of objects.

2.2.2. *Auditory Environments*

Research into auditory environments has explored the use of 3D sound for navigation and localisation. In particular, Savadis et al. [9] explored the use of spatial audio for computer interfaces using 3D audio output and gestural and speech input. Walker et al. [10] evaluated three types of sound beacons for performance in guiding users in an auditory virtual environment: noise beacons were found to be the most effective. Serafin and Serafin [11] use sound to increase the sense of presence in virtual reality by using environmental sounds.

3. THE EXPERIMENT

3.1. Overview

As mentioned previously, the aim of this research is to explore providing a brief overview of desktop. We will be studying how well this method conveys the desktop information using sound. This would be as an equivalent or an alternative to a user's visual glance at a graphical screen.

We aim to use readily available technology; thus, we will be providing 3D sound over headphones using generalised Head-Related Transfer Functions (HRTF). The use of headphones also makes it possible to easily adapt this interface for mobile and wearable technology.

3.1.1. *The Sound Space*

To give the user a snapshot of the current state of their desktop, the computer screen will be translated into a 3D sound space. The 2D graphical representation of the desktop will map to a horizontal sound plane that is angled slightly towards the user. The angle used for the base sound plane is derived from the idea of it being like a drafting table as opposed to the traditional desktop/table metaphor. The sound space will only extend in front of the user to give the impression that the objects to be interacted with are within easy reach.

The desktop objects to be presented are located on an axis perpendicular to the base sound plane dependent on their normalised file size. We set an expected minimum and maximum size for each object type and objects of average size are presented at the junction of their axis and the base sound plane while larger and smaller files are presented above and below the user respectively. This is depicted in Figure 1. Size is not necessarily the most crucial information to be presented but this gives us the opportunity to evaluate the discernibility of this sort of information. Other information could be imparted with use of timbre and pitch.

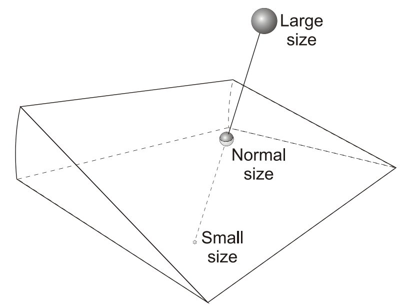


Figure 1: *Normalised file size.* A file is located along the axis perpendicular to its location on the base plane. Larger files are located above the base plane while smaller files are located below the base plane. The relative sizes are based on normalised sizes for that file type.

3.1.2. *The Auditory Icons*

The auditory interface presents each object present by playing a sound at its location. Our experiment will evaluate a variety of presentation methods: concurrent presentation and three types of serial sweeps of the audio space. In order to keep the interface simple, each desktop object type will have a particular sound associated with it to provide basic information. The following types have been defined:

Documents such as PDF, word processing or presentation files

Plain text and code files

Markup Language files with associated viewers

Folders and Archive files

Organisational such as calendars, to-do lists and email

Spreadsheets

Media such as image, music and video files

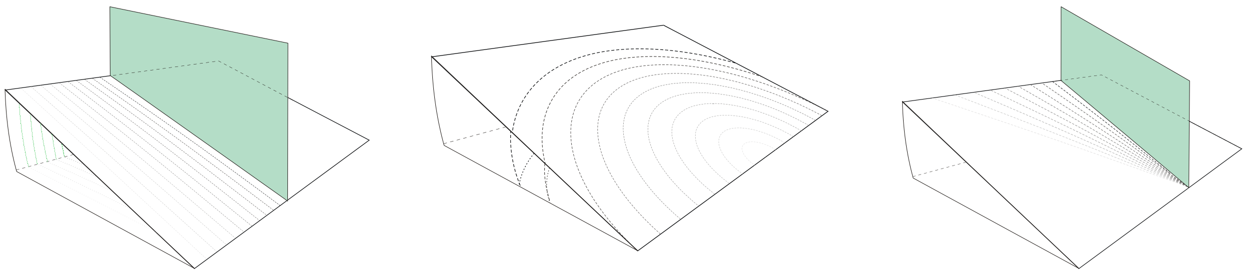
Communication such as instant messaging, shell programs, VoIP

Applications will be associated with their type and thus will not have a category of their own. A specific concurrently played sound will indicate if the object refers to an application or link, while files stored on the desktop would get no modifier.

3.1.3. *Training*

Each participant will be familiarised with the sounds being used to identify each type of desktop object as described above. This will be done in a short session where the participant will make associations between the sounds and the targets. They will be asked to listen to each sound and select in a graphical interface the correct target type. The training interface will provide positive or negative feedback. The rate of association will also be tracked.

Additionally, there will be series of trial runs of listening to the desktop and seeing the graphical representation so that the participants can gain familiarity and simulate confidence that would usually come with experience.



(a) *Condition 2: Serial Horizontal Sweep*. A virtual plane sweeps horizontally across the sound space and plays the objects it encounters.

(b) *Condition 3: Serial Radial Sweep*. A virtual sphere with increasing radius sweeps from the bottom centre of the sound space at the level of the base plane and plays objects as it passes them.

(c) *Condition 4: Serial Angular Sweep*. A virtual plane hinged on the bottom centre of the sound space sweeps clockwise, playing objects as they are encountred.

Figure 2: *The Serial Sweep Experimental Conditions*.

3.1.4. The Method

Each participant will be asked to listen to several desktop configurations and draw out what they heard. They will pinpoint what they assume is the location of the earcon as well as their impression of the size of the file or whether it is an application or link. The drawing will be compared to the actual layout to determine how effective a means of communication this overview presentation of desktop objects is. At the conclusion of the experiment, we will conduct a structured interview to document the experience.

3.1.5. Participants

There will be 24 participants for the experiment. Each experimental condition will be tested by six participants. We will be using non-novice computer users to avoid having unfamiliarity with computers and the graphical desktop interface interfere with the results. Each participant group will be made up of an equal number of male and female participants.

3.2. Experimental Conditions

What is described below are the four experimental conditions that participants will be asked to test. At this stage in the work, these conditions are simulated.

Condition 1: Concurrent Presentation In concurrent presentation, a sound announcing that the desktop objects are to be presented are played and then all the objects are presented concurrently. With this condition it is presumed that the user will get a impression of the busy areas of the interface if not very much granular information.

Condition 2: Serial Horizontal Sweep The serial horizontal sweep, shown in Figure 2(a), presents objects serially as a virtual plane sweeps the audio space from left to right. This presentation method simulates reading. The implications of this presentation is that all objects on a column in the graphical equivalent would be presented concurrently. As this is a serial presentation, each desktop object on the column would be offset slightly from each other to make them distinct from each other.

Condition 3: Serial Radial Sweep The serial radial sweep, shown in Figure 2(b), presents objects as a virtual semi-sphere expands from the user with increasing radius. This means that the closest objects would get presented first. This is not necessarily ideal as most window managers automatically place objects beginning in the upper corners of the graphical interface.

Condition 4: Serial Angular Sweep The serial angular sweep, shown in Figure 2(c), presents objects as a virtual vertical plane, hinged on the user and orthogonal to the base sound plane, sweeping clockwise. This operates rather like a sonar. Like the serial radial sweep, there is a disconnect with normal desktop object placement.

3.2.1. Other Variables

We are experimenting with a variety of variables to the sound in order to optimise our method. Informal pre-experiment simulations will set these variables and the structured interviews will help gauge their effectiveness.

Sound Length The length of the sounds representing desktop objects can have a great impact on the localisation and identification of the object. However, having longer sounds can negatively impact the length of the serial sweeps.

Sweep Speed The total length of each presentation condition varies with the distance to be travelled. Assuming a common speed, the angular sweep is the slowest method, horizontal is the second slowest, radial the second fastest, and concurrent the fastest. The speed the virtual plane or semi-sphere travels through the sounds space can be modified so that, for example, the same length of presentation is achieved for all the sweeps.

Decay The use of decay where a sound slowly fades while another sound is playing may provide additional cues to a user and give more of an impression of persistence.

Movement Sound localisation is easier to achieve with a moving sound source or when the user may move their head to better triangulate the location of the source. As the sound space is effectively a grid, we will be looking at the effect of having the desktop object move around in its sound cube or slide up/done its axis to base sound plane.

4. DISCUSSION

4.1. Future Work

The poster to be presented at the conference will describe the four experimental conditions and the results of our study. It will also include examples of the auditory data and describe more fully the condition that proved most effective.

There are several directions our work may take in the future. One such direction is allowing certain objects to be more audible than others. For example, desktop objects whose targets have been used recently could be emphasised. Alternatively, objects that have not been used in some time could be made more noticeable so that the user can clean up their desktop.

Another possibility is to present the objects based on patterns in the desktop. That is, the area that the most dense objects could be presented first, possibly more slowly, then other less sparse areas could be presented in turn. This would be a more user-oriented way of conveying the important aspects of the the system status.

Yet another avenue of interest is increasing the sense of presence by simulating echoes against the limits of our 3D space. The space would then become enclosed and the interface would become an environment.

Elements of user control could also be added such that they could redefine file types and sounds, filter for certain types, hone in on an area of the interface for more granularity, etc.

All of these avenues of future work lead to permitting better navigation and mouse interaction without a visual display. Overview information provides the possibility for creating better mental models of the interface and thus should make it possible to navigate more effectively. By allowing the visually impaired access to mouse navigation, we are attempting to open doors to improved collaboration with other users as well as improve interaction for all users. Using the same input devices as non-impaired users should make it easier to work with others and discuss how to accomplish tasks. If an auditory interface is successful, it will also provide another layer of feedback to a non-impaired user and potentially increase understanding of their interface.

4.2. Conclusion

The purpose of our work is to explore methods of providing general information about the desktop interface. This sort of information could lead to improved navigation of computer interfaces without use of a screen. The assistive technology for the visually impaired currently does not provide an effective method of navigation of the interface that rivals the direct manipulation of the mouse. Providing overview information aurally over a short period allows the user to interact with a system without necessitating that the objects contained therein constantly announce their presence. The overview presentations described in our work also place the control back in the hands of the user; that is, it is user-initiated and therefore less distracting and obtrusive. Research of this nature is increasingly important as computers becomes more ubiquitous. In order to move forward, we need to develop other means of interaction; sound is an appropriate medium for portable devices as well as accessibility technology.

5. REFERENCES

[1] Elizabeth D. Mynatt and W. Keith Edwards, "Mapping GUIs to auditory interfaces," in *UIST '92: Proceedings of the 5th annual ACM symposium on User interface software and technology*. 1992, pp. 61–70, ACM Press.

[2] Haixia Zhao, Catherine Plaisant, Ben Shneiderman, and Ramani Duraiswami, "Sonification of geo-referenced data for auditory information seeking: Design principle and pilot study," Sydney, Australia, July 6-9 2004, International Conference on Auditory Display.

[3] I. J. Pitt and A. D. N. Edwards, *People and Computers VI: Proceedings of HCI '91*, chapter Navigating the Interface by Sound for Blind Users, Cambridge University Press, 1991.

[4] Stephen A. Brewster, "Using nonspeech sounds to provide navigation cues," *ACM Trans. Comput.-Hum. Interact.*, vol. 5, no. 3, pp. 224–259, 1998.

[5] Elizabeth M. Wenzel, "Spatial sound and sonification," in *Auditory Display - Sonification, Audification, and Auditory Interfaces*, Gregory Kramer, Ed., vol. XVIII, pp. 127–150. Addison-Wesley, 1994.

[6] Russell L. Martin, Ken I. McAnally, and J.P. Watt, "Localisation of multiple sounds as a function of the duration of the inter-sound gap," Sydney, Australia, July 6-9 2004, International Conference on Auditory Display.

[7] David K McGookin and Stephen A Brewster, "An investigation into the identification of concurrently presented earcons," Boston, Massachussetts, United States, 6-9 July 2003, International Conference on Auditory Display.

[8] David K McGookin and Stephen A Brewster, "Space, the final frontearcon: the identification of concurrently presented earcons in a synthetic spatialised auditory environment," Sydney, Australia, July 6-9 2004, International Conference on Auditory Display.

[9] Anthony Savidis, Constantine Stephanidis, Andreas Korte, Kai Crispien, and Klaus Fellbaum, "A generic direct-manipulation 3D-auditory environment for hierarchical navigation in non-visual interaction," in *Assets '96: Proceedings of the second annual ACM conference on Assistive technologies*. 1996, pp. 117–123, ACM Press.

[10] Bruce N. Walker and Jeff Lindsay, "Effect of beacon sounds on navigation performance in a virtual reality environment," Boston, Massachussetts, United States, 6-9 July 2003, International Conference on Auditory Display.

[11] Stefania Serafin and Giovanni Serafin, "Sound design to enhance presence in photorealistic virtual reality," International Conference on Auditory Display, July 2004.