Sonic Grid: An Auditory Interface for the Visually Impaired to Navigate GUI-based Environments[1]

Deepak Jagdish
Microsoft
Bangalore, India
deepak.jagdish@gmail.com

Mohit Gupta
Microsoft Research
Bangalore, India
v-mgupta@microsoft.com

Rahul Sawhney
Indian Institute of Information Technology
Hyderabad, India
rahul.sawhney@gmail.com

Shreyas Nangia
Indian Institute of Management
Calcutta, India
shreyas.nangia@gmail.com

ABSTRACT

This paper explores the prototype design of an auditory interface enhancement called the Sonic Grid that helps visually impaired users navigate GUI-based environments. The Sonic Grid provides an auditory representation of GUI elements embedded in a two-dimensional interface, giving a '*global*' spatial context for use of auditory icons, ear-cons and speech feedback. This paper introduces the Sonic Grid, discusses insights gained through participatory design with members of the visually impaired community, and suggests various applications of the technique, including its use to ease the learning curve for using computers by the visually impaired.

Author Keywords

Auditory Interface, Accessibility, Touch based devices

ACM Classification Keywords

H.5.2. User Interfaces: Auditory (non-speech) feedback, Input Devices and Strategies.

INTRODUCTION

The modern world has been designed with certain assumptions about the inherent capabilities of humans. In engineering our world, we have often ignored minorities with certain physiological or mental incongruities that make it hard for them to interact with modern environments – both real and virtual.

Our project called Sonique, aims to create auditory interfaces, and in the process explore technological and design issues in developing navigational solutions for the visually impaired. We looked at two broad areas where navigation for the visually impaired is difficult – *navigating through interior and urban spaces* and *using graphical user interfaces (GUIs)*.

This paper focuses on the latter area of the project, which tries to tackle the challenges that visually impaired users face while interacting with graphical interfaces on computers and mobile devices. In particular, we present the Sonic Grid, an auditory representation of two-dimensional space that provides visually impaired users with a spatial context while navigating GUIs.

We employed a participatory design approach, by having numerous sessions with students at the Blind People's Association, Ahmedabad, India. Preliminary results indicate that the concepts behind the Sonic Grid show promise by providing effective global spatial context for sound and speech feedback annotations.

MOTIVATION

Though command line interfaces are relatively accessible by the visually impaired (through the use of existing tools), icon and pointer based Graphical UIs still pose significant barriers to adoption. Accessing the different elements of a GUI becomes a cumbersome task simply because the visually impaired are not spatially aware of the location of the different elements on screen.

Accessibility problems are exacerbated due to the growing use of stylus and touch based devices that have interaction elements spatially dispersed over an interaction surface. Visually impaired users cannot correlate the movement of a pointing device to the corresponding location on the screen. Consequently they are unable to perceive the spatial layout and lack the context to be able to effectively navigate the GUI. Finding

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IUI'08, January 13-16, 2008, Maspalomas, Gran Canaria, Spain.

Copyright 2008 ACM 978-1-59593-987-6/ 08/ 0001 $5.00

[1] The work presented here was undertaken by the authors as students of Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar, India.

ways to provide a global context through auditory means has been a central goal of the Sonique project.

Several other factors motivated our work:

1. Audio interfaces can be customized and evolved much more easily than haptic feedback devices, which require specialized hardware and are thus limited by hardware capabilities.

2. Feedback during the participatory design sessions with the visually impaired community indicated that auditory landscapes – through increased engagement - could encourage computer adoption by blind children, especially in developing countries such as India.

3. Such interactions can encourage the visually impaired to learn more about navigating spatial environments. There is little spatial training for blind students in India.

4. Past work on *sonification* is explored primarily for complex data [1]. Also, work relating to auditory icons and ear-cons [2] take an object-based approach to provide audio-cues about GUI elements, but have not focused on how to provide a global spatial context.

THE SONIC GRID

The Sonic Grid is an interface enhancement that allows the visually impaired to access the elements of a GUI based system in its intended non-linear spatial representation. In order to facilitate effective use of the spatial pointing devices, a visually impaired user is presented with an *interactive grid layer* on the screen, which gives *sound feedback* conveying the active position on the screen using non-speech audio cues (see Figure 1). This virtual audible grid covers the entire viewing area of the screen making sure that every pixel is accessible to the user via audio feedback. Unlike auditory icons and ear-cons, the Sonic Grid feedback is not icon or application dependent.

Prototype Implementation

The current prototype of the Sonic Grid, developed on the Microsoft Windows platform is annotated with both speech and non-speech information. The annotations provide audio-cues to the user about objects, operating system events, results of user action etc. The annotations were implemented through extensive experimentation with operating system hooks[2] to create rule based filters that captured, modified and translated operating system

[2] In the Microsoft® Windows® operating system, a hook is a mechanism that allows intercepting messages, mouse actions, keystrokes before they reach an application. http://msdn2.microsoft.com/ms997537.aspx

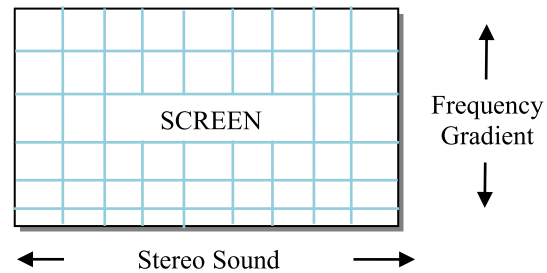

Figure 1. An abstract representation of Sonic Grid

and input device (keyboard, mouse and stylus) events. Further details of software implementation are not discussed here because of space constraints. Different styles and compositions of audio feedback were also explored during our research, as we explain in later sections.

AUDIO REPRESENTATION OF SPATIAL LOCATION

We explored several ways to encode sound "coordinates" of any location on screen. We present some of the encoding schemes in this section. The next section presents feedback from participatory design sessions with the visually impaired community.

The lateral (left-right) positional aspect can be expressed when the sound output is in *stereo* mode. The relative volume of the left and right earphones are adjusted so as to provide lateral cues: when the mouse pointer is to the extreme left, only the left speaker is activated, and as the pointer moves towards the right, the right speaker is given increased prominence, thus simulating a pan effect (Figure 1). Sound is perceived as being on a plane between the ears and 'within' the head [1]. To increase the detail of feedback for the horizontal position, we experimented with chord progressions, and keyboard triggered speech feedback.

Encoding *vertical* position is more challenging. We employed a variety of techniques that encode the vertical position as frequency and beat variations. In the basic implementation, frequency is mapped on the vertical axis of the grid such that, at the bottom of the screen, it is very low and as the pointer moves up, the frequency gradually increases as per the gradient.

The elementary sounds used with the Sonic Grid range from about 100Hz to 5500Hz as higher frequency sounds are very unpleasant [2]. The sounds are a composed of a complex of frequencies that produce beats, which makes identification easier. The variation in frequency with vertical pixel shifts on screen is linear, and change in frequency is calibrated by dividing the screen into equal sized grids.

FEEDBACK FROM PARTICIPATORY DESIGN

Using an early prototype we engaged in participatory design sessions with students at the Blind People's Association, Ahmedabad. During our interactions with them, we observed both free exploration and completion of predefined tasks. These sessions were documented on video and the following observations and required changes were noted, for further system iterations.

The spatial information provided by the Sonic Grid was initially not appreciated by the users, and required an explanation of GUI organization. This was because of a lack of awareness about GUI-based systems, and the nature of the sounds used in the system. Our efforts to deal with these issues are discussed in later sections.

The students gained more confidence as they spent more time understanding the system and felt at ease using a speech engine that was similar sounding to the one used by JAWS (a popular Windows-based screen reader).

Several issues relating to the dynamics of the Sonic Grid were identified. A need was felt to play audio alerts when the user moved rapidly over objects. This gave them a better 'picture' of the screen. The alerts were followed by speech feedback if the user hovered over a particular GUI element. Truncation of the speech feedback occurred only if the user moves to a new element requiring speech. In other cases the user is notified of '*moving away'* by reducing the amplitude of feedback.

The background sound of the grid was given a separate volume control so that the grid could be a faint sound while keeping speech volume higher. We also included the option to disable the continuous sound so that the grid sound is only heard on specific intervals or on a change of a set number of pixels. This meant that the user will be provided with the Sonic Grid feedback only if he/she travels 10 pixels from the current position, thus cutting down on the continuous background 'noise'. Also a volume gradient was incorporated where the sound of the grid would reduce when the speech engine was active and then rise back to the set volume after it had finished.

In earlier versions, our system relied largely on pointing device input. Based on strong feedback from the user group, we integrated more keyboard support, including shortcuts and keyboard-based navigation (*mousekeys*). For specific scenarios like web browsing and media handling, we felt that specialized menus leveraging the Sonic Grid could bring value. Thus, a prototype called the *in-focus menu* was designed and is later described briefly.

Spatial training for the visually impaired

Our interactions revealed that spatial training is not imparted to the visually impaired students in India. Looking into the methods of spatial training by institutes such as BECTA, UK we realized that the Sonic Grid can help students gain spatial skills, but would require development of a training module. Tactile models, apart from audio games such as AudioQuake [5] can help in the understanding of virtual space and sound, although we did not explore this extensively.

CHOOSING THE RIGHT SOUNDS

Our experimentation with sounds was undertaken at two levels – (a) to design easily discernible, but aesthetically pleasing sounds and (b) use sounds to help train and encourage blind children to adopt computers.

Using simple tones was found to be monotonous and distracting; therefore we explored using sounds that appealed more to users. An initial list of experimental soundscapes that were selected for design of the Sonic Grid were:

1. Sounds inspired from the environment such as the gradient (intensity & speed) of (a) raindrops falling and (b) winds – the natural flow of air
2. Sounds inspired from 'ragas[3]' and Western classical scales.

Sounds inspired from the environment

Through manual recording and external sources, an attempt was made to build a sound library that is easily discernable and can be used to represent the pixels on screen. For a clear representation of the average screen at a resolution 1024x768, with toolbar icon heights approximately 15-20 pixels about 50 sounds are required in a gradient. A gradient is essential as its absence negatively affects the user's ability to identify position on the screen. To create smooth gradients for sounds such as rainfall and winds require sound modeling that was not explored during the course of our project.

Sounds based on scales or `ragas`

Sound collections are inherently limited as they are very hard to generate dynamically, and therefore we explored the use of sound octaves to create pleasant sounding Sonic Grid feedback. The plain use of musical notes does not produce sounds pleasant enough and therefore the use of multiple harmonics was essential [2].

The scales we identified for use with the system based on their different cultural background and supposed emotional response for testing were the Chromatic Scale, Natural (pure) Minor Scale, Raag Kiravani (C D D# F G G# B C), the Major Pentatonic scale and the Folk Scale (C C# D# G G# Cg). The Folk Scale (pelog) was particularly interesting as this scale is used in meditation and is said to be inherently rhythmic which could help in the notes being identified easily.

Landscape inspired sounds for children

It is challenging for the visually impaired to start learning computers as it requires a huge amount of memorizing and training. It was also observed that blind children do not appreciate computers because of similar reasons.

---

[3] A *raga* (*raag*) can be broadly defined as a series of musical notes which are systematically organized within a scale with different ascending and descending versions of itself.

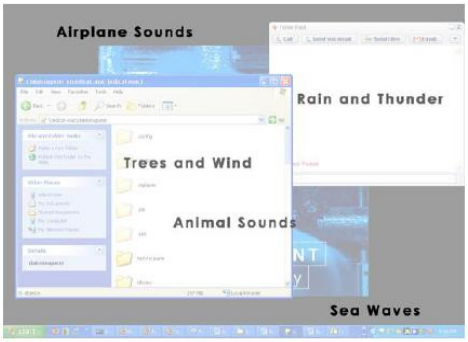

Figure 2. Using familiar sounds for easier understanding and initiation of users

We collected audio samples that represent a landscape to make computers fun for children and at the same time making them comfortable with using a computer, keyboard and mouse. Some of the familiar sounds we explored were aircraft sounds, natural sounds comprising of animals, wind, trees etc. (Figure 2). This configuration was very popular with students, and we are planning to include it as part of their computer training process.

FURTHER WORK

The work presented in this paper is preliminarily and there is a further need for systematic user tests to refine the system. The user studies carried out with people who were already well versed with keyboard-based systems such as JAWS revealed a certain frustration when trying to learn to use the system. They could complete the given task easily if they used the keyboard. We are therefore, refining the system to allow for a gradual introduction of the Sonic Grid alongside traditional screen readers. We also hope to experiment with the incorporation of musical structure beyond the use of musical scales [6].

We are excited about using the Sonic Grid to allow for spatial expression by the visually impaired. We did some preliminary experiments in this direction using a paint application and vOICe [7] but it needs further exploration.

Additionally we are also exploring the design of interfaces that leverage spatial organization with specially designed visual elements that aid people with partial vision. In particular our prototype called *in-focus menu* is a high contrast radial menu that presents commonly used actions in an application. This was designed to augment the underlying layer of the Sonic Grid. The user moves around the point where he/she first clicked, over the different segmented options to receive speech feedback about each action. The radial menu is coupled along with a sub-vertical menu as and when required. Such a vertical listing would occur in scenarios like the play-list content in a media player or a listing of hyperlinks that are present in a webpage. A pie menu is used so the user can rapidly access actions, as he/she learns their directions on the menu.

It was felt that the greatest contribution of our system was to ease learning and initiation of computer use by visually impaired users. We interacted with teachers and instructional designers who expressed their concern about the use of the Sonic Grid to support the curriculum without further testing. We intend to continue with their guidance to a build a well defined training module for spatial understanding of GUIs.

CONCLUSION

Auditory Interfaces are crucial for the visually impaired to access the same computer applications that the visually-abled use with ease. We introduced the Sonic Grid, which provides users with spatial cues, allowing them to keep a global spatial context. Furthermore, with addition of proper speech feedback and audio cues for notification of events, the Sonic Grid opens up a new dimension of interaction with computers for the visually impaired. There is a certain amount of learning required over a period to be able to use the Sonic Grid fluently, but since it capitalizes on auditory perception to enable two-dimensional interaction, this will be a useful step towards the accessibility of pointer and touch based devices. In fact we have seen that providing an audio landscape encourages visually impaired children to become interested in using computers. We hope that the ideas and experiences described here will stimulate further research in this area.

ACKNOWLEDGEMENTS

We express our gratitude to the visually impaired students and Mr. Ranchod Bhai, at the Blind Peoples' Association, Ahmedabad for their patience and candid feedback. We would like to thank Prof. Binita Desai from DA-IICT and Joseph Joy at Microsoft Research for their guidance; and our friends, Soumya Manivanan and Praneet Koppula for their help.

REFERENCES

1. Zhao, H. Interactive Sonification of Abstract Data, *PhD Thesis 2006 University of Maryland*, pp. 2-3
2. Brewster, S. A. Providing a Structured Method for Integrating Non-Speech Audio into Human-Computer Interfaces. *University of York, HCI-G, 1994.*
3. Pitt, I. & Edwards, A. (1991). Navigating the interface by sound for blind users. In D. Diaper & N. Hammond (Eds.), *Proceedings of HCI'91*, pp. 373-383.
4. Massof, RW. Auditory Assistive Devices for the Blind, *Proceedings of International Conference on Auditory Display 2003,* icad.org
5. http://www.agrip.org.uk/AudioQuake
6. Leplatre, G. and Brewster, S. A. (1998). Perspectives on the Design of Musical Auditory Interfaces. *Focus conference on Anticipation, Cognition and Music.*
7. Jung-Kyong Kim & Robert J. Zatorre (2006) Visual-to-auditory substitution learning: Behavioral findings and neural correlates, *IMRF 2006*.