JVIB

February 2005 • Volume 99 • Number 2

Representing Graphical User Interfaces with Sound: A Review of Approaches

Dan Ratanasit and Melody M. Moore

Abstract: The inability of computer users who are visually impaired to access graphical user interfaces (GUIs) has led researchers to propose approaches for adapting GUIs to auditory interfaces, with the goal of providing access for visually impaired people. This article outlines the issues involved in nonvisual access to graphical user interfaces, reviews current research in this field, classifies methods and approaches, and discusses the extent to which researchers have resolved these issues.

This article is based on research that was conducted for the Rehabilitation Engineering Research Center on Workplace Accommodations, Center for Assistive Technology and Environmental Access, Georgia Tech, and was supported, in part, by Grant H133E020720 from the National Institute on Disability and Rehabilitation Research, U.S. Department of Education.

A significant advance in computer applications in the past two decades has been the development of the graphical user interface (GUI) (Shneiderman, 2003). GUIs are pervasive in almost every workplace domain, including business information systems, design-support software, financial management, and decision support software. Most of the interfaces to the World Wide Web, a critical source of information in the business world, are graphical. Although GUIs have greatly improved the convenience of computer use for sighted people, computer users who are visually impaired (that is, those who are blind or have low vision) are at a disadvantage because of the largely graphical nature of the web and most application interfaces (Alty & Rigas, 1998; Asakawa & Itoh, 1998; A. D. N. Edwards, 1989).

The ubiquity of GUIs is a considerable challenge for computer users who are visually impaired. Without access to GUIs, their opportunities for employment, advancement, education, and even leisure activities may be limited (Gerber, 2003). Providing people with visual impairments with access to GUIs is a critical concern (Mynatt & Weber, 1994), particularly since computer literacy is a critical skill in the workplace (Tobias, 2003). Fortunately, researchers have examined approaches to transforming or representing GUIs with auditory interfaces. Related work in accessibility for persons who are visually impaired has been under way for decades and has included ULTRA, an apparatus that assisted visually impaired students in chemistry laboratories (Lunney & Morrison, 1981) and the Optophone, a hand-operated mechanical device that represented printed text with musical notes (Beddoes, 1968; D'Albe, 1920). This article discusses the issues inherent in transforming GUIs into auditory interfaces, reviews research in the field, classifies the approaches, and evaluates the issues that still remain.

Background

Character-oriented user interfaces

Before the advent of GUIs, information systems were implemented with character-oriented interfaces, which required typed commands and provided synchronized responses from the computer. These user interfaces were fairly straightforward to represent in an auditory manner because the text on the screen could be read out loud through screen readers and speech synthesizers (Mynatt, 1997). GUIs introduced difficulties to visually impaired computer users because graphical components, such as buttons and icons, could not be represented by early screen readers (Donker, Klante, & Gorny, 2002; Mynatt & Edwards, 1992a). With text-based screen readers, the text on the screen was captured, and the contents were spoken aloud by a speech synthesizer. With graphical components, the items in question are not text characters but pixel values, making auditory representation much more difficult (Mynatt, 1992).

Characteristics of GUIs

Although some older business systems are still implemented with character-oriented interfaces, most have incorporated some form of direct manipulation afforded by GUIs (Mynatt & Edwards, 1992a). GUIs provide visually based representations of computer objects on different levels from the operating system (files, directories, and hard disks) to applications (edit boxes and scroll bars) (Mynatt & Edwards 1992a). The typical implementation of a GUI is the WIMP (Windows Icons Menu Pointer) paradigm (van Dam, 1997), which is organized primarily by containers (windows, frames, and dialogues). These interfaces make use of the mouse pointer and icons to allow users to manipulate items in their computer environment.

Multitasking or *multiprocessing*, the ability to work with multiple tasks at one time, is another powerful advantage provided by GUIs. That more than one window may be open concurrently, each one accessed through its own interface, allows users to select and switch between windows of focus (W. K. Edwards, Mynatt, & Stockton, 1994; Mynatt & Edwards, 1995). Users have the ability to arrange their windows spatially in a manner that provides convenient access to all (W. K. Edwards et al., 1994; Ludwig, Pincever, & Cohen, 1990).

In addition to the technical differences between graphical and auditory interfaces, there are also numerous design differences. The challenge is to develop an auditory interface that provides the same advantages as GUIs do (Mynatt & Edwards, 1995). Some aspects of GUIs that constitute this challenge are as follows:

1. Organization: Visual information is frequently used to present the logical structure of content spatially (Asakawa, Takagi, Ino, & Ifukube, 2002), or users may arrange items to grant convenient access (W. K. Edwards et al., 1994). For example, icons on a desktop that represent related documents can be grouped together on the screen.

2. Graphical information: Visual representations (icons) of applications, files, or other objects (Mynatt & Edwards, 1992b) allow users to locate and identify desired items rapidly. Visual interfaces allow many icons to be viewed simultaneously, reducing search time.

3. Multitasking: Users may easily switch among applications that are running in concurrently open windows (Mynatt & Edwards, 1995; W. K. Edwards et al., 1994). For example, a user may have a spreadsheet application and a presentation application open at the same time, to move or copy data from one to the other.

4. Occlusion: Graphical windows can be superimposed over each other or over icons to hide information. In a graphical interface, the hidden window has not disappeared, but occlusion

renders it inaccessible to a screen reader (Mynatt & Edwards, 1992b).

5. Spatial semantics: In GUIs, information is presented by position through groupings, tables, lines, and spacing (Asakawa et al., 2002). For example, a dialogue box may group a text box for a product name with checkboxes indicating pricing options.

6. Graphical semantics: Semantic data are conveyed through visual elements like font size, style, face, background colors, and foreground colors (Asakawa et al., 2002). Items, such as links on a web page, can be highlighted.

7. Two-dimensional structure: GUIs can present information on a two-dimensional screen, facilitating layout and organization. However, this type of organization is not easily translated to the serial nature of speech, which can present information only sequentially (Nielsen, 2003).

Methods

We searched the online literature databases CiteSeer, Portal, Open WorldCat, and GALILEO and the publications of ASSETS (ACM SIGCAPH Conference on Assistive Technologies), SIGCAPH (Special Interest Group on Computers and the Physically Handicapped), and SIGCHI (Special Interest Group on Computer-Human Interaction) to retrieve articles that were relevant to the subject of this article. The online search engine Google allowed us to search for specified keywords or researchers' names. The keywords searched were *auditory interface, auditory icons, visually impaired, blind users, earcons,* and *nonvisual access.*

Articles on voice recognition and voice interfaces that resulted from the search using the above keywords were discarded because they did not focus on audio output in user interfaces. The articles that we found presented different methods and approaches to the problem of representing GUIs with sound, including providing complete audio interfaces for a computer system, for web browsing only, or exploring the use of nonspeech audio.

Findings

The studies and projects described in this review used different techniques in representing GUIs with sound. These techniques can be divided into the following categories: use of nonspeech sound cues, audio representation of tables and diagrams, audio representation of backgrounds and visual effects, audio representation of spatial information and graphical differences, and the use of three-dimensional audio. The research for each of the categories is outlined in the following sections.

Use of nonspeech sound cues

Three types of nonspeech sound cues were discussed in the works covered by this review: *auditory icons*, *earcons*, and *hearcons*.

Auditory icons

Gaver (1986) defined *auditory icons* as everyday sounds that are used to represent graphical objects. With an auditory icon, a direct analogy exists between the sound and the object it represents. Some examples are using the sound of tapping on glass to represent a window or the sound of a typewriter to represent a text edit box (Gaver, 1993).

Mercator (Mynatt, 1997), created by Mynatt and Edwards (1992b), makes use of auditory icons. The primary objective of Mercator is to provide transparent access to GUIs on the X-Windows operating system for users who are visually impaired, which means that auditory access to a graphical application would not require modifications to the application. Mercator implements a software agent that collects information on the application as it runs. The agent observes the application's behavior and graphical items that it draws to the screen. The results from the agent's analysis form what is called an *off-screen model*. The off-screen model provides information about the GUI's components (such as buttons and scroll bars) to construct the auditory interface. For input and events, Mercator translates objects for an auditory interface by communicating with different X-Windows libraries

(Mynatt, 1992).

In Mercator, the audio equivalents of GUI objects, such as buttons and menus, are called *audio interface components,* represented by auditory icons (Mynatt, 1992). To convey attributes on audio interface components, auditory cues, called *filtears*, are used. Filtears are effects, such as pitch, inflection, reverb, and muffling, that are applied to auditory icons to help identify the state or attributes of an audio interface component.

The Audio Rooms project, another project spearheaded by Mynatt and Edwards (1995), also incorporated auditory icons to provide graphical functionality in auditory interfaces. The authors used the metaphor concept (Marx, 1994) for Audio Rooms, which uses knowledge in one domain to increase familiarity with another. Some examples of sounds that were used in the Audio Rooms environment were a creaky door to signify entering a room, a copy machine to signify copying a file, or a laser printer to signify printing a file (Mynatt & Edwards, 1995).

Roth, Petrucci, Pun, and Assimacopoulos (1999) developed an auditory interface to help visually impaired users use graphical Internet browsers that also uses auditory icons. This tool was designed to help users who are visually impaired access web browsers through auditory and tactile means. Some auditory icons that are used by this browser are typewriter sounds to represent text and the sounds of a camera shutter to represent images.

Earcons

Blattner, Sumikawa, and Greenberg (1989) described earcons as another example of audio cues that are used to convey information about computer objects and operations. An earcon is an abstract sound that is not necessarily semantically connected to the object it represents. Earcons can be broken down into components called motives (different rhythms, pitches, intensity, timbre, and register). They may be used to represent operations or objects that are not ordinarily associated with everyday sounds. Examples include a musical sound that is played when files, programs, or menus are opened or closed. These actions and objects can be assembled into different combinations. Musical instruments may be used to represent actions. For example, a violin may be used to represent opening a file, or a flute may be used to represent deleting a file.

Brewster, Wright, and Edwards (1993) described an evaluation of the effectiveness of earcons in auditory interfaces. The rhythms used for the earcons in their experiments were combinations of quarter notes and eighth notes, and the pitches used were different notes on the musical scale. The intensity of the earcons was simply a difference in volume. The different timbres used in the experiments were a piano, brass instruments, a marimba, and pan pipes. Different registers were notes at different octave positions on the musical scale.

The first phase of Brewster et al.'s (1993) experiment included write, paint, and draw operations, as well as manipulations of system objects, such as files, folders, and applications. These researchers separated the operations and objects into different families and assigned a motive to each. For example, all paint objects used the same instrument, and all application objects used the same rhythm. Items in the same category, such as two separate paint files, were distinguished by different pitches (C and G). The second phase of the experiment used menus with a variety of selections, such as Open, Close, Delete, Save, Copy, and Undo. Each menu had its own timbre, and the items were differentiated by rhythm, pitch, and intensity. The third and fourth phases experimented with combinations of Phases 1 and 2.

Another important aspect of this study was testing the earcons against unstructured sounds for their ability to convey information. The unstructured sounds lacked rhythm, simply lasting one second, but shared some of the timbres of the earcons. The participants were asked to describe the items represented by the earcons and unstructured sounds to see how well the earcons communicated information. Brewster et al. (1993) found that the musical earcons were significantly more effective than were the unstructured sounds. In a second experiment, the earcons were enriched to improve their sound. The rhythms were redesigned with more notes, the pitch patterns were made more complex, and the earcons were expanded to two timbres. Brewster et al. tested the enriched earcons with the same set of participants to compare the results with those of the first experiment. They found that the changes resulted in increased levels of recognition by the participants.

Brewster, Raty, and Kortekangas (1995) tested the use of earcons to represent menu hierarchies. A tree structure was used to represent a hierarchy in which a node on the tree was occupied by an earcon. Each earcon inherited attributes from earcons above it, such as rhythm, pitch, timbre, register, tempo, stereo position, effects, and dynamics. The example that Brewster et al. (1995) presented illustrated a hierarchy of errors. The earcon at the root node, named "ERROR," consisted of simple motives. The earcon used middle-register pitch, occupied a central stereo location, and used a flute as its timbre. The children of this earcon, "OPERATING SYSTEMS ERROR" and "EXECUTION ERROR," had different timbres, different registers, and different stereo positions from each other and from the parent node, but used the same note, A. The earcons were made more complex as they occupied further nodes down the tree. The children of "EXECUTION ERROR" had more complex rhythms and intensity as additional parameters.

The participants in this experiment were all familiar with the hierarchy that was used and were given full explanations of the earcons' structures. They were tested on how well they remembered what the earcons represented in a four-level hierarchy. The participants achieved an average success rate of 82%, indicating that the earcons represented hierarchies well. They had the most difficulty remembering items at the bottom level, which indicates that increasingly complex earcons that were built upon inherited attributes were not as effective. Brewster et al. (1995) surmised that this difficulty was due to the increasing amount of audio information to be remembered but suggested using different, rather than fewer, motives at the lower level for better success.

Hearcons

Donker et al. (2002) developed another type of sound cue for their auditory browser, ZIB. They named these sound cues hearcons and categorized them into two distinct groups: nature sounds and musical works or musical instruments. Hearcons are similar to earcons when there is no natural metaphor with the objects they represent. They differ in that earcons are formed by separate audio components (rhythms, pitches, intensity, timbre, and register), whereas hearcons may consist of completed sounds, such as those produced by birds or a running river, rather than being pieced together.

ZIB uses hearcons to represent the different components that are typically found on web pages, such as links, images, headings, and paragraphs. The sounds of horns represent page headings, xylophones playing news-show ticker rhythms represent paragraphs, passages from musical pieces signify images, and synthesized sounds represent hyperlinks. Although the participants in Donker et al.'s (2002) study were trained on the semantic mappings represented by the hearcons in ZIB, the results showed that the hearcons were not effective. Donker et al. determined that the hearcons did not sufficiently represent semantic relationships.

A. D. N. Edwards (1989) discussed a word processor with an audio interface, Soundtrack, designed as part of an experiment to model the interactions of visually impaired persons using a mouse with audio interfaces. Unlike the other works discussed in this review, rather than convert an already existing graphical interface, Soundtrack was an application that used both a graphical and audio interface. The tool emitted different musical tones when the mouse pointer encountered a different object on the screen and spoke the name of the object if it was clicked. The experimenters measured the time it took for users to choose a target, plan a route among targets, move to a target, and click on the target to create the model.

Audio representation of diagrams and tables

Bly (1981) performed an experiment that showed that sound may reveal relationships in data in much the same way as a two-dimensional graph does. Bly mapped different groups of data to different sounds and to different points on a graph. The participants were able to classify samples into groups on the basis of the audio information just as well as they could on the basis of the visual information.

Kennel (1996) developed a tool called Audiograf with the specific purpose of assisting users who are visually impaired to read diagrams. The sounds that Audiograf uses may be classified as earcons, since no natural relationship exists between the sound and the object it represents. For example, a plucked string represents lines on a diagram. The user navigates to different diagrams on the screen with a finger, and the auditory output provides information on the diagram that is displayed. Tests with participants with visual impairments showed that Audiograf performed well and that the participants correctly described the displayed graphs within three minutes.

Brewster (2002) and Ramloll, Brewster, Yu, and Riedel (2001) developed a tool that uses earcons to improve access to two-dimensional tables of numerical information. They found that using only speech to convey information that is contained in tables became overwhelming with increasingly larger tables. Ramloll et al. stated that using speech alone was too time-consuming to deliver numerical data from a table and made it too difficult to uncover trends in data because of the limitations of human memory. They identified three critical issues in this scenario. The first is knowledge of the current location within the table; users frequently want to know their current location in the table and become discouraged and feel lost if this information is not readily available. The second is overloaded speech feedback from navigation; typically, information on the current row, column, and contents is given when navigating from one table cell to another. With speech, this information is excessive and not necessary at all times. The third is the lack of information on the size of the table; users benefit from knowing the bounds of the table and just how much data they are contending with.

Ramloll et al.'s (2001) interface uses MIDI (Musical Instrument Digital Interface) sounds with differences in pitch as nonspeech cues. Lower numbers are associated with lower notes, and higher numbers are associated with higher notes. Differences in the numerical values are not directly mapped to differences in the pitch. It would be extremely difficult for a user to determine that a pitch that is twice as high as another represents a number that is twice as large as another. The tool may operate in three modes: label, value, and pitch. The label and value modes give the user information on the table using speech and the pitch mode using nonspeech (that is, sounds with differences in pitch). The speech outputs of the label and value modes are easily distinguished by a male or female voice. The tool also makes use of stereo sounds to give users a better sense of their current position in the table.

Ramloll et al. (2001) conducted an experiment with their tool with 16 visually impaired participants who were aged 23–57. The goal of the study was to test the effectiveness of using speech alone versus using speech combined with nonspeech. The participants were given data tables on London crime rates and students' performance and sets of questions that required them to refer to the tables to answer. Ramloll et al. found that tests with the speech-pitch combination yielded the better results. Through the use of the tool, the workload and time taken to complete the given tasks both decreased. The participants also had greater success rates with their tasks using the speech-pitch combination. Ramloll et al. envisioned the tool as a particularly good complement to screen readers and systems that provide access to spreadsheets, tables, graphs, and data plots (see also Brewster, 2002).

Audiograph (Alty & Rigas, 1998) is another audio tool for representing graphs and diagrams. It uses music to represent graphical objects. Audiograph was used in a set of experiments to see if music alone could successfully represent graphical objects to users who are visually impaired. The coordinates on a graph were represented with pitch, whereby a higher pitch signified a greater distance from the origin. The X-axis and Y-axis were represented with organ and piano timbres, respectively. Geographic shapes were drawn on an X-Y graph and were represented using the described differences in pitch. Control actions were represented with earcons. For example, the command “Expand” was conveyed with an earcon of a particular melody and rhythm (timbre not indicated), and its inverse was used to represent the command “Contract.”

Alty and Rigas (1998) drew conclusions on the role of context in their methods of converting graphical interfaces. They stated that not only should the musical structures be distinct, but that the user will have some expectation of what the music is meant to represent and that the design of the musical metaphor depends, to some extent, on this expectation. Alty and Rigas reported good results from experiments with the tool. The participants who were visually impaired could use the tool to identify shapes and objects, move them, change their size, and save and retrieve them. They had positive reactions to the tool, but criticized the lengthiness of the sound cues and the effort taken to interpret them.

The TeDUB project (Technical Drawings Understanding for the Blind), coordinated by the University of Bremen, Germany, is a system designed to present technical diagrams, such as electronic circuit diagrams, UML (Unified Modeling Language) diagrams, and architectural drawings, to users who are visually impaired (Petrie et al., 2002). Five types of interfaces for the system—text, static tactile overlay, two-dimensional sound, force-feedback joystick, and three-dimensional sound—will be implemented and tested. The text interface will present the diagram as text, and the user will navigate and interact with the diagram with the help of a screen reader. The static tactile overlay interface will use a touch-sensitive keyboard to work with a tactile presentation of a diagram. The two-dimensional interface will enhance the text interface by providing nonspeech sound cues to assist the user with the diagram. The force-feedback joystick will make use of various tactile force effects to allow the user to explore a diagram. The three-dimensional interface will attempt to convey spatial information to the user. The three-dimensional and joystick interfaces were tested in 2002. Nearly all the participants made favorable comments about the joystick, saying it provided easy navigation and a good sense of the user's current location. More negative comments were made about the three-dimensional audio. The participants said that spatial information was not adequately provided and suggested that it should be used with the joystick (FNB, 2004).

Meijer (1992) discussed an experimental system for representing graphical images, rather than diagrams or tables, as sound patterns. The system translates an image pixel by pixel using three different sound characteristics. The row of the pixel is represented by a corresponding pitch, the column by a "time after click," and the level of brightness by volume. The images that Meijer described were that of a parked car and a human face. Meijer reported that the sound patterns that were produced matched the authors' expectations, resulting in successful conversions and at a good resolution.

Itoh and Yonezawa (1990) worked on a system for visually impaired users that assisted with handwriting. This system did not translate a GUI into an auditory interface, but demonstrated the importance of a user's knowledge of the current position of his or her pen. The user would write on a tablet that emitted different frequencies and amplitudes of sounds, depending on the position of his or her pen. The system was tested with three visually impaired and sighted (blindfolded) students with good hearing. It was found to be effective in assisting the visually impaired students to write, showing that audio feedback is beneficial in an interface.

Audio representation of backgrounds and visual effects

Asakawa et al. (2002) worked on representing visual effects on the World Wide Web through auditory and tactile methods. They described a macroapproach that focused on representing page organization and overview and a microapproach that focused on information conveyed through textual differences. The macroapproach is described here, and the microapproach is described in the next section.

Asakawa et al. (2002) found that most web pages use primarily background colors (for tables or full pages) to signify organizational groupings and fragmentations. They used music to represent these colors and groupings. They associated different colors with different instruments; for example, blue was mapped to strings, red was mapped to synthesizers, and a piano playing John Lennon's "Imagine" represented white.

Asakawa et al. (2002) identified links, text, and images as conveying semantics by visual effects. They described the use of earcons to these visual effects and chose sounds that were significantly different from the background music to prevent confusion. A gunshot represented a link, a bagpipe's low tone represented text, and a bagpipe's high tone represented images.

Asakawa et al. (2002) conducted an experiment to evaluate their macroapproach using a simple page (a few groupings: menu, heading, and main content) and a complex page (many groupings: menu, heading, main content, sports, weather, business, video, links, and so forth). They conducted their experiment with five visually impaired participants who were already familiar with using auditory browsers for Internet access. Asakawa et al. found that the users who were most familiar with web browsing performed best with their system. The users differentiated different groupings on the simple page much more quickly (within three trials) and accurately than with the complex page (small adjacent groups were difficult to distinguish). They commented that choosing their own songs would be preferable, which Asakawa et al. found quickened recognition. Since a large number of colors can be displayed, Asakawa et al. determined that using a song to represent a color group, rather than an individual color, would be more feasible.

Audio representation of graphical differences

Asakawa et al.'s (2002) microapproach focused on levels of emphasis that were conveyed through different fonts. The authors applied two different levels of emphasis to their earcons: stronger and weaker. The tinkling of a small bell signified a weaker emphasis (different font styles to express emphasis or larger fonts relative to the size of ordinary text on the page), and the ringing of a large bell signified a stronger emphasis (even larger fonts).

The microapproach was tested with a simple page containing different sizes of text and bold text. All the participants were able to determine the correct number of emphasis levels, some with fewer trials than others. Asakawa et al. (2002) deemed the microapproach successful, since it dealt with independent units of text.

Asakawa and Itoh's (1998) home page reader also makes distinctions between regular text and hyperlinked text with different voices. A male voice is used to read ordinary text, but the browser switches to a female voice whenever the user navigates to a hyperlink. This noticeable auditory change quickly notifies the user that information of a different context has been reached. Roth et al.'s (1999) browser also uses differences in voice to indicate the presence of hypertext links to the user. However, rather than use a completely different voice, the browser uses a voice with a different tone and adds simple sounds, such as beeps. The ZIB browser made use of hearcons to differentiate between regular text and hyperlinks. The hearcons that Donker et al. (2002) chose for this purpose were of synthesized sounds consisting of different motives.

Audio representation of spatial information

Asakawa et al.'s (2002) article, on research with music and earcons, also addressed auditory representations for spatial layout. The authors found that HTML provided no way to indicate groupings organized by background colors. As we mentioned earlier, they used music to represent colors and earcons to represent text, images, and links. They chose highly contrasted sounds (a piano versus a gunshot) to reduce a listener's confusion. The changes in music would help the user differentiate the context of content during navigation, whether it came from the same group or a different group.

The table readers described earlier—Audiograph (Alty & Rigas, 1998) and the one developed by Ramloll et al. (2001)—both work to preserve spatial layout. Both tools use pitch to help the user determine the distance from a particular reference point. Rather than use songs, as Asakawa et al. (2002) did, Audiograph used earcons that were exclusively musical in nature. Piano or organ sounds were used on the graph, and different rhythms were used for the control actions (Alty & Rigas, 1998).

Roth et al.'s (1999) audio browser uses a three-dimensional auditory space to preserve spatial layout and graphical objects. The developers created a proxy server that inserts scripts into Internet documents to obtain their attributes to determine the properties of the objects (such as text, images, and links) that are contained in the document. Doing so enables the tool to get information on the spatial content of the page and to relay this information to the user through three-dimensional audio.

Three-dimensional audio

Three-dimensional audio is another approach taken by some researchers. Perrett and Noble (1997) performed two experiments to test the effects of head motion on detecting the locations of sound sources. Their work demonstrated that listeners could indeed differentiate sounds originating from different heights, showing that three-dimensional audio can convey meaningful information to a listener.

An important aspect of Audio Rooms (Mynatt & Edwards, 1995), mentioned earlier, was three-dimensional audio. This Audio Rooms scenario was similar to a desktop but used a three-dimensional room, rather than a two-dimensional desktop. The room acted as a container for similar applications the way a window groups together similar tasks. The contents of a room could be files, data, or “doorways” to other rooms. The authors proposed that the “rooms” metaphor is attractive to visually impaired users who are more spatially aware of their everyday surroundings than are sighted users. Mynatt and Edwards believed that the use of spatialized sound would help users who are visually impaired determine the layout of the rooms, as well as the locations of objects in the environment.

Savidis, Stephanidis, Korte, Crispien, and Fellbaum (1996) described a three-dimensional auditory environment that they developed, presenting their tool as a generic reusable environment. They maintained that their tool stands apart from others because it uses three-dimensional pointing and voice input and is reusable. Savidis et al. found that previous related works were designed for more specialized tasks, whereas since their environment is more generalized, it is reusable in different environments. They arranged hierarchical objects within a horizontal circular plane with the user centrally located. Three-dimensional stereo audio let the user know where the objects were located. The user navigated and selected objects using a glove for three-dimensional hand gestures and voice input. This architecture granted users with visual impairments the ability of direct manipulation. At the time that their article was published, the authors had yet to determine suitable sound cues, hand gestures, or voice commands. They stressed the characteristics of three-dimensional audio, voice input, and three-dimensional hand gestures as strengths of their tool.

Roth et al.'s (1999) browser maps an item on the screen to a location in its three-dimensional audio space. Users of the browser explored the web document in two phases: a macroanalysis phase and a microanalysis phase. The macroanalysis phase, performed first, is where the overall document structure and objects, such as text, images, and forms, are analyzed. The microanalysis phase is where information is gathered on the individual objects.

During the macroanalysis phase, the user runs a finger across the touch-sensitive screen. Feedback is given to the user in the form of nonspeech sounds to indicate the type of interface object encountered. The sound is projected into a three-dimensional sound space to give the object's type and location. During the microanalysis phase, the user gains information on the objects revealed in the macroanalysis.

The auditory browser ZIB (Donker et al., 2002) attempts to maintain the information conveyed by two-dimensional graphical layouts. ZIB uses stereo sound to deliver hearcons in a three-dimensional auditory interaction realm, which users can distinguish spatially from one another. The system also makes use of a pointing device or a joystick to assist in navigating from one hearcon to another in much the same way as a sighted person uses a mouse.

Donker et al. (2002) evaluated ZIB in 16 trials with sighted (blindfolded) and visually impaired users ranging in age from 18 to 40. They found that the users who were visually impaired did not perform better with ZIB than they did with familiar browsers. Using ZIB, the participants did not correctly identify all objects in a layout and also had trouble identifying a page’s layout.

The participants were asked to reconstruct the layout of a web page after they used ZIB. The sighted users attained better results overall, whether or not they were blindfolded, which led Donker et al. (2002) to hypothesize that sighted individuals form different mental models than do visually impaired individuals.

Results and discussion

Generating an audio interface from a GUI is much more complex than simply representing graphical objects with sound. The articles we reviewed illustrated that nonspeech audio has been a major focus of this field. Brewster’s et al.’s (1993, 1995) experiments with earcons showed they are effective when not overly complex, whereas Donker et al.’s (2002) experiments with hearcons proved the opposite. However, these articles did not evaluate auditory icons.

A quick audio cue can be used to impart immediate information about an object to a user much more quickly than can a spoken word, as demonstrated by Audiograf (Kennel, 1996); Audiograph (Alty & Rigas, 1998); and Ramloll et al.'s (2001) table reader. Even different voices used by Ramloll et al.'s table reader and Asakawa and Itoh's (1998) home page reader, although still speech, impart different meanings to the user by simply being distinct. The user can learn to associate a given voice with a given meaning, in much the same way as he or she can learn to associate a nonspeech sound with a particular meaning. A person can differentiate among several different audio signals but can attend to only one or two at a given moment. These techniques require well-designed audio cues and training of users (Buxton 1989), but bring the visually impaired user closer to being able to scan a document or interface quickly or to decide quickly what to discard and what to focus on. Thus, we conclude that nonspeech audio provides more of the advantages of a GUI than does speech alone.

It is imperative that in the workforce, visually impaired users have access to the same functionality and content as do sighted users through GUIs (Mynatt, 1992). Achieving this goal while maintaining consistency between audio interfaces and their related GUIs presents a challenge. Designers must consider the need not only to convert graphics to text or sound, but to convert a two-dimensional layout to a serial layout.

One issue that was not thoroughly addressed in any of the articles that we reviewed was *serialization*: converting a two-dimensional interface (graphical) into a one-dimensional interface (generally required for an auditory interface) (Mynatt, 1992). Some of the projects aimed to maintain the semantics behind two-dimensional graphical layout and organization. This is quite a challenge, since ultimately the information is conveyed to the visually impaired user in a sequential manner. A sighted user may easily perform a quick scan of the interface with a simple glance. This task can be achieved with an audio interface, but it still requires information to be relayed serially.

Table 1 summarizes the studies reviewed in this article and categorizes them by their approaches, methods, and results. Many of the studies addressed more than one of the issues in auditory interfaces. A quick scan of the columns reveals the amount of research that each issue received; for example, almost all the studies addressed nonspeech sounds, but representation of graphical differences received much less attention. Some prominent issues, such as serialization, were not addressed at all. Some of the studies focused on design issues, such as determining the optimal sounds for representing icons, while other studies focused on transformation issues, such as representing graphical tables and diagrams with auditory outputs. Each study made a unique contribution to the state of the art of assistive technology for people with visual impairments.

Conclusions and future work

Building upon conclusions drawn from these projects, we devised a list of requirements for transforming GUIs to auditory interfaces. These requirements incorporate both nonspeech sounds and serialization for two-dimensional graphical representations. We are currently designing and implementing a software tool set, called AudioMORPH, to facilitate the adaptation of workplace graphical interfaces into auditory interfaces for people who are visually impaired. The specific aim of AudioMORPH is to provide automated configurations for existing screen readers to assist in customizing proprietary software. Business applications typically require a programmer or technician to adapt a GUI manually, specifying auditory representations of screen layouts. AudioMORPH is designed, on the basis of recommendations and results reported in the literature, to automate this process and to incorporate nonspeech sounds as cues. Our goal is to provide a quick, straightforward method of producing a usable, comprehensible auditory interface from a business system GUI without the aid of a programmer. We hope that AudioMORPH will provide more opportunities for people with visual impairments to be optimally productive in the workforce.

References

Alty, J. L., & Rigas, D. I. (1998). Communicating graphical information to blind users using music: The role of context. *CHI '98: Conference on Human Factors in Computing Systems* (pp. 574–581). New York: ACM Press.

Asakawa, C., & Itoh, T. (1998). User interface of a home page reader. In *Proceedings of the Third International ACM Conference on Assistive Technologies, 1998* (pp. 149–156). New York: ACM Press.

Asakawa, C., Takagi, H., Ino, S., & Ifukube, T. (2002). Auditory and tactile interfaces for representing visual effects on the web. *Proceedings of the Fifth International ACM Conference on Assistive Technologies, 2002* (pp. 65–72). New York: ACM Press.

Beddoes, M. P. (1968). An inexpensive reading instrument with a sound output for the blind. *IEEE Transaction on Biomedical Engineering, 15,* 70–79.

Blattner, M., Sumikawa, D., & Greenberg, R. (1989). Earcons and icons: Their structure and common design principles. *Human-Computer Interaction, 4,* 11–44.

Bly, S. (1981). Presenting information in sound. *Proceedings of the 1982 Conference on Human Factors in Computing Systems* (pp. 371–375). New

York: ACM Press.

Brewster, S. A. (2002). Visualisation tools for blind people using multiple modalities. *Disability and Rehabilitation Technology, 24,* 613–621.

Brewster, S. A., Raty V., & Kortekangas, A. (1995). *Representing complex hierarchies with earcons* (Working paper ERCIM-05/95R037 of the European Research Consortium for Informatics and Mathematics Research Reports) (Online). Available: ftp://ftp.inria.fr/associations/ERCIM/research_reports/pdf/0595R037.pdf

Brewster, S. A., Wright, P. C., & Edwards, A. D. N. (1993). An evaluation of earcons for use in auditory human-computer interfaces. *Proceedings of the INTERCHI '93 Conference on Human Factors in Computing Systems* (pp. 222–227). New York: ACM Press.

Buxton, W. (1989). Introduction to this Special Issue on Non-Speech Audio. *Human-Computer Interaction, 4*, 1–9.

D'Albe, F. (1920). The Optophone: An instrument for reading by ear. *Nature, 105*, 295–296.

Donker, H., Klante, P., & Gorny, P. (2002). The design of auditory user interfaces for blind users. *Proceedings of the Second Nordic Conference on*

Human-Computer Interaction (pp. 149–155). New York: ACM Press.

Edwards, A. D. N. (1989). Modelling blind users' interactions with an auditory computer interface. *International Journal of Man-Machine Studies, 30*, 575–589.

Edwards, W. K., Mynatt, E., & Stockton, K. (1994). Providing access to graphical user interfaces—Not graphical screens. *Proceedings of the First Annual ACM Conference on Assistive Technologies* (pp. 47–51). New York: ACM Press.

FNB. (2004). *TeDUB: Technical drawings understanding for the blind* [Online]. Available: http://www.tedub.net

Gaver, W. W. (1986). Auditory icons: Using sound in computer interfaces. *Human-Computer Interaction, 2*, 11–44.

Gaver, W. W. (1993). Synthesizing auditory icons. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 228–235). New York: ACM Press.

Gerber, E. (2003). The benefits of and barriers to computer use for individuals who are visually impaired. *Journal of Visual Impairment & Blindness, 97*, 536–550.

Itoh, K., & Yonezawa, Y. (1990). Support system for handwriting characters and drawing figures for the blind using feedback of sound imaging signals. *Journal of Microcomputer Applications, 13,* 177–183.

Kennel, A. R. (1996). Audiograf: A diagram reader for the blind. *Proceedings of the Second Annual ACM Conference on Assistive Technologies* (pp. 51–56). New York: ACM Press.

Ludwig, L. F., Pincever, N., & Cohen, M. (1990). Extending the notion of a window system to audio. *IEEE Computer, 23*, 66–72.

Lunney, D., Morrison, R. C. (1981). High technology laboratory aids for visually handicapped chemistry students. *Journal of Chemistry Education, 58,* 228–231.

Marx, A. N. (1994). Using metaphor effectively in user interface design. *Conference Companion on Human Factors in Computing Systems* (pp. 1–2). New York: ACM Press.

Meijer, P. B. L. (1992). An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering, 39,* 112–121.

Mynatt, E. (1992). Auditory presentation of graphical user interfaces. *Proceedings of the 1992 International Conference on Auditory Display* (pp. 1–18). Santa Fe:

Addison-Wesley.

Mynatt, E. (1997). Transforming graphical interfaces into auditory interfaces for blind users. *Human Computer Interaction, 12,* 7–45.

Mynatt, E., & Edwards, W. K. (1992a). Mapping GUIs to auditory interfaces. *Proceedings of the Fifth Annual Symposium on User Interface Software and Technology* (pp. 61–70). New York: ACM Press.

Mynatt, E., & Edwards, W. K. (1992b). The Mercator environment: A nonvisual interface to X Windows and UNIX workstations. *Proceedings of the ACM Symposium on User Interface Software and Technology* (pp. 92–105). New York: ACM Press.

Mynatt, E., & Edwards, W. K. (1995). Metaphors for nonvisual computing. In A. Edwards & J. Long (Eds.), *Extraordinary human-computer interaction: Interfaces for users with disabilities* (pp. 201–220). New York: Cambridge University Press.

Mynatt, E., & Weber, G. (1994). Nonvisual presentation of graphical user interfaces: Contrasting two approaches. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating Interdependence* (pp. 166–172). New York: ACM Press.

Nielsen, J. (2003). Alternative interfaces for accessibility. *Jakob Nielsen's Alertbox* [Online]. Available: http://www.useit.com/alertbox/20030407.html

Perrett, S., & Noble, W. (1997). The effect of head rotations on vertical plane sound localization. *Journal of the Acoustical Society of America, 102,* 2325–2332.

Petrie, H., Schlieder, C., Blenkhorn, P., Evans, G., King, A., O'Neill, A.-M., Ioannidis, G., Gallager, B., Crombie, D., Mager, R., & Alafaci, M. (2002). TeDUB: A system for presenting and exploring technical drawings for blind people. In K. Miesenberger, J. Klaus, & W. Zagler (Eds.), *Computers helping people with special needs* (pp. 537–539). New York: Springer.

Ramloll, R., Brewster, S., Yu, W., & Riedel, B. (2001). Using non-speech sounds to improve access to 2-D tabular numerical information for visually impaired users. *Proceedings of the First Workshop on Human Computer Interaction with Mobile Devices,* 515–530.

Roth, P., Petrucci, L., Pun, T., & Assimacopoulos, A. (1999). Auditory browser for blind and visually impaired users. *CHI '99 Extended Abstracts on Human Factors in Computing Systems* (pp. 218–219). New York: ACM Press.

Savidis, A., Stephanidis, C., Korte, A., Crispien, K., & Fellbaum, K. (1996). A generic direct-manipulation 3-D–auditory environment for hierarchical navigation in non-visual interaction. *Proceedings of the Second Annual ACM Conference on Assistive Technologies* (pp. 117–123). New York: ACM Press.

Shneiderman, B. (2003). *Designing the user interface* (3rd ed.). Reading, MA: Addison-Wesley.

Tobias, J. (2003). Information technology and universal design: An agenda for accessible technology. *Journal of Visual Impairment & Blindness, 97,* 592–601.

van Dam, A. (1997). Post WIMP user interfaces. *Communications of the ACM, 40,* 63–67.

Dan Ratanasit, B.C.E., *research scientist, Computer Information Systems Department, Georgia State University, P.O. Box 4015, Atlanta, GA 30302; e-mail: <dratanas@cis.gsu.edu>.* ***Melody M. Moore, Ph.D.,*** *associate professor, Computer Information Systems Department, Georgia State University, Park Place 10, Suite LLB, Atlanta, GA 30303; e-mail: <melody@gsu.edu>.*

Previous Article | Next Article | Table of Contents

Search JVIB | **JVIB Policies** | **Contact JVIB** | **Subscriptions** | **JVIB Home**

If you would like to give us feedback, please contact us at jvib@afb.net.

www.afb.org | Change Colors and Text Size | Contact Us | Site Map |

Site Search Go

About AFB | Press Room | Bookstore | Donate | Policy Statement

Please direct your comments and suggestions to afbinfo@afb.net