Impact of Spatial Auditory Feedback on the Efficiency of Iconic Human-Computer Interfaces Under Conditions of Visual Impairment

Armando B. Barreto\textsuperscript{1}, Julie A. Jacko\textsuperscript{2}, and Peterjohn Hugh\textsuperscript{1}

\textsuperscript{1}Department of Electrical and Computer Engineering  
Florida International University  
10555 West Flagler Street, Room EC-3956  
Miami, FL., 33174  
tel: (305) 348-3711  
fax: (305) 348-3707  
email: barretoa@fiu.edu  
http://dsplab.eng.fiu.edu  
** CORRESPONDING AUTHOR : Armando Barreto **

\textsuperscript{2}School of Industrial & Systems Engineering  
Georgia Institute of Technology  
765 Ferst Drive  
Atlanta, GA 30332-0205  
tel: 404.894.2342  
fax: 404.894.2301  
email: jacko@isye.gatech.edu  
http://www.isye.gatech.edu/lhci/
Abstract

This paper investigates the addition of spatial auditory feedback as a tool to assist people with visual impairments in the use of computers, specifically in tasks involving iconic visual search. In this augmented interface, unique sounds were mapped to visual icons on the screen. As the screen cursor traversed the screen, the user heard sounds of nearby icons, spatially, according to the relative position of each icon with respect to the screen cursor. A software prototype of the design was developed to evaluate the performance of users in the search of icons within the proposed interface. Experiments were conducted with simulated visual impairments on volunteer participants to evaluate if the addition of spatial auditory feedback makes the interface more accessible to users with impaired vision. Results demonstrated that spatialization of icon sounds provides additional remote navigational information to users, enabling new strategies for task completion. Directions for future research are discussed and prioritized.

Keywords:

Human-Computer Interaction, Visual Impairment, 3D Sound, Spatialization, Multimedia, Icon
Introduction

Vision can be impaired for a variety of reasons. Although most definitions of visual impairment originate in the healthcare community, the concept is also relevant in the realm of human-computer interaction, especially when the capabilities of the user and environmental work contexts are considered. For example, an ocular disease that results in reduced visual acuity, or diminished visual field, will result in computer-oriented disabilities. Similarly, poorly lit work environments or those laden with airborne particles can also result in impaired (i.e., obscured) vision and thus disable a person from accessing a computer as efficiently as they otherwise would have been able. The former is an example of Health-Condition-Induced-Impairments (HCII). The latter is an example of Situationally-Induced-Impairments (SII) (Sears, Lin, Jacko, & Xiao, 2003; Sears & Young, 2003). Both involve degradations in visual processing, albeit due to different causes. It must be explicitly acknowledged that SII is fundamentally different from HCII in that they tend to be temporary and dynamic. In contrast, HCII are generally uncorrectable and sometimes more predictable, in that they tend to become worse over time. The research reported in this paper has implications for both SII and HCII. Experimental results were generated through the use of simulated visual impairments that are representative of both SII and HCII, in order to obtain a baseline of performance in the presence of visual feedback augmented by spatial auditory feedback, within a graphical user interface (GUI). Research that is applicable to not only SII but also to HCII is particularly relevant as The American Foundation for the Blind estimates that 10 million individuals in the United States are visually impaired or blind (American Foundation for the Blind, 2002). Visual impairments and blindness encompass a host of clinical diagnoses, residual visual abilities, and rates of functional decline that in most cases inhibit or prevent access to information technologies (Jacko & Sears, 1998). Various forms of spatial auditory feedback have been investigated as solutions to enabling access to information technologies for people who do not have fully functional vision (e.g., Ramloll & Brewster, 2002). Understanding the role of spatial auditory feedback as a tool for mitigating the influence of HCII during GUI task performance is an important contribution to the scientific knowledge base.
Background

A hierarchy of critical interactions in GUIs was introduced by Emery et al. (2001) in which four main categories of interaction were identified, one of which was object manipulation. Object manipulation requires physical functions/actions such as pointing, moving, and selecting objects within an interface (Jacko, et al., 2003). The direct manipulation paradigm employed in GUIs requires the use of a pointing device, most commonly a mouse, to perform these actions.

Visual icon manipulation and use is a key operation to study because it requires the successful integration of visual and motor functioning by the user. Icons serve as a means through which users can initiate higher-level actions and concepts without the use of complex syntax (Shneiderman, 1998). When using a mouse, two interaction scenarios are critical to successful icon activation: selection and positioning (Foley, Wallace & Chan, 1984). In selection, the user chooses from a set of items displayed on the screen. In positioning, the user specifies a point in a one- two- or three-dimensional space. Successful use of iconic representations within GUIs requires use of visual feedback from the interface, and thus places considerable demands on the human visual system (Jacko, Rosa, Scott, Pappas, & Dixon, 2000).

Auditory feedback, used as a supplement to visual feedback, has been shown in studies of multimodal research to improve the performance of several different direct manipulation tasks for a variety of users. For example, Brewster (1998) found that sonically enhanced feedback on drag-and-drop tasks produced better task performance compared with a visually enhanced version of the task. Jacko et al. (2003) demonstrated that older adults who have visual impairment due to Age-Related Macular Degeneration experienced performance improvements when an auditory component was included in feedback provided during performance of a drag-and-drop task. These results were similar to the results postulated by Fraser and Gutwin (2000), who reported using enhanced feedback in visually stressed environments. As a result of their study, Fraser and Gutwin suggested that auditory feedback adds more value for individuals with reduced visual acuity than for individuals with normal vision under optimal viewing conditions. Interestingly, Jacko et al. (1999) discovered benefits of auditory feedback for
visually healthy, older users as well. This led to the assertion that auditory feedback benefits extend to the visually healthy older adult population, which often experiences a general decline in perceptual abilities. The most common forms of auditory feedback used in multimodal research include auditory icons and earcons (Blattner, Sumikawa, & Greenburg, 1989; Brewster, 1998; Gaver, 1989).

**Auditory Icons and Earcons**

Similar to their visual counterparts, auditory icons capitalize upon metaphors established between sounds encountered in everyday listening and computer events. Computer users must be able to detect auditory icons and decipher their intended meaning (Jacko & Rosenthal, 1997). Gaver (1986) is credited with introducing auditory icons to computer interface design. His work is based upon the principles of ecological perception (Gibson, 1986). The underlying premise of his approach is that humans do not listen to the acoustical dimensions of sound such as pitch and timbre but rather attend to the source that created the sound. An interface based on this premise was developed for the Apple Macintosh computer for sighted users. SonicFinder (Gaver, 1989) extended the visual desktop metaphor into the auditory domain by mapping standard computer events to everyday sounds. For example, the computer event initiated when a user drags a window was mapped to a scraping sound; window scrolling was mapped to a ticking sound. Mappings such as these provide redundant information to what is displayed visually. Providing redundant information to the computer user is especially valuable in the presence of SIIs, such as when a computer user’s attention must be directed elsewhere, or when the computer user is simultaneously being presented with large amounts of information from various sources. For a more detailed review of literature pertinent to auditory icons refer to Jacko (1996).

Blattner, Sumikawa and Greenberg (1989) are credited with pioneering a design approach for the incorporation of earcons into computer interfaces. Earcons are abstract in nature and are composed of synthetic tones that are either single pitch or groups of pitches. They are constructed from fundamental elements called motives. Motives possess specific features; five of the most important features include rhythm, pitch, timbre, register and dynamics. Mappings between computer functions and earcons are
arbitrary. While relatively easy to construct, earcons may pose certain problems for both designers and users because they do not possess a semantic relationship with the computer function that they are designed to represent (Jacko, 1996). However, Brewster, Wright and Edwards (1993) demonstrated that auditory earcons are better for presenting information than unstructured bursts of sound. It has been further demonstrated that high levels of recognition can be achieved by careful use of pitch, rhythm and timbre (Brewster, Wright, & Edwards, 1992). For example, earcons have been shown to enhance computer users’ employment of graphical buttons and icons in interfaces. One common error associated with the use of graphical buttons and icons in computer-based interfaces is slip-off error, which results in serious effects as the mistake may not become immediately evident to the user. By introducing an auditory component to icon selection, researchers (Brewster, Wright, Dix, & Edwards, 1995) demonstrated that computer users were equipped to recover from slip-off errors faster than with the traditional visual design.

Spatial auditory feedback can serve as an aid for localization of features and functions within a computer interface (Blauert, 1996). Although not as prevalent as auditory icons and earcons, auditory spatial cues can be particularly valuable in the presence of SIIs, such as environments in which there are copious distractions, interruptions, or noise (Gaver, 1993). Such cues are often produced via manipulation of loudness, pitch, timbre, and/or localization (Belz, Robinson, & Casali, 1999). Strybel (1995) presented the benefits of auditory spatial information in visual processing for cockpit tasks involving gate identification, blunder avoidance, and traffic identification. Findings showed that both simple detection and identification are quicker and more constant across the frontal hemifield when targets are augmented with auditory spatial cues. For sounds presented in the central visual field, Strybel demonstrated that auditory spatial cues can either supplement or substitute for abrupt visual onsets in directing attention.
Three-Dimensional Sound

The goal of 3-D sound emulation is to manipulate a listener’s spatial auditory perception of a sound source. Figure 1 shows a 3-D sound model adapted from Begault (1994). Using the spherical coordinate system with its origin at the center of the listener’s head, a virtual sound source has a direction given by azimuth ($\theta$) and elevation ($\phi$), at a magnitude distance ($r$). Zero degrees in azimuth and elevation points directly in front of the listener. Clockwise rotations in azimuth from $0^\circ$ to $180^\circ$ are positions to the right of the listener. Counterclockwise rotations from $0^\circ$ to $-180^\circ$ are positions to the left. Upward rotations in elevation from $0^\circ$ to $180^\circ$ are positions above the head. Downward rotations from $0^\circ$ to $180^\circ$ are positions below the head. With 3-D sound, a listener perceives a manipulated virtual sound source to have direction and distance according to these spherical coordinates (Begault, 1994). More advanced 3-D sound systems also account for effects derived from room acoustics, materials and conditions, such as reverberation and humidity (Kendall, 1995).

![3-D Sound Model](image)

Figure 1. 3-D Sound Model (Adapted from Begault, 1994)

Head-Related Transfer Function Technology

The Head-Related Transfer Function (HRTF) is the vehicle that allows for modeling and implementation of artificial 3-D sound. HRTF refers to the spectral filtering of a sound that occurs before it reaches the eardrum of the listener, according to
the position of its source (Begault, 1994). The shape of the pinna, or outer ear, is the primary cause of this spectral filtering. Other significant filtering effects are due to the head and body. This position-dependent filtering is responsible for the natural perception of sound source localization and the basis for its artificial emulation. The brain’s spatial perception of a sound source results from the frequency-dependent interaural time and intensity differences of a sound entering each eardrum. For the artificial emulation of 3-D sound, those differences are imposed on a monaural sound by convolution with a pair of measured HRTFs at a desired azimuth and elevation. Although 3-D sound using HRTF technology can be delivered through loudspeakers, the effect is most dramatic when delivered through headphones.

**Implementation of 3-D Sound**

Microsoft® has developed the DirectX® Application Program Interface (API) to standardize the implementation of multimedia including 3-D sound in the Windows® operating system. The implementation of 3-D sound centralizes its functions around the management of sound sources and a listener in a 3-D space. DirectX® compliant soundcards implementing 3-D sound with HRTFs generally have several 3-D hardware channels. The actual number available depends on how many HRTF filtering operations can be computed within a sampling period. A 3-D audio hardware channel serves as input to a mixer where 3-D sound effects can be applied to the sound before mixing. A sound source is assigned to each 3-D audio hardware channel. The data structure of the sound source is composed of a sound buffer and a location point. The data structure of the listener consists of a listening point and an orientation vector that contains the direction where the head is turned (See Figure 1). The third data structure required for 3-D sound emulation is the matrix of HRTFs, varying in angles of azimuth and elevation. The data from these structures determine the parameters of the 3-D sound effect (Carl, 1998).

The parameters of the 3-D sound effect can be calculated using geometry and physics. The directional vector of the source is computed using its location point and the listening point. When the directional vector is rotated relative to the origin by the orientation vector, the azimuth and elevation angles are used to select the corresponding
pair of HRTFs from a matrix. At this point, the directionality of the sound is determined and HRTF filters can be computed. The next step is to account for the effects of distance through a customizable distance attenuation function. This function is adjustable by a minimum distance, a maximum distance, and a roll-off factor. From zero to the minimum distance, the volume of the 3-D sound is fixed at a maximum value of one. From the minimum distance to the maximum, the volume decreases simulating the effect of sound waves losing energy as they traverse a distance through a medium. The roll-off factor controls the emulation of the dampening level of the medium through which the sound travels. At distances greater than the maximum distance, the sound source can no longer be heard. A distance factor scales the entire 3-D sound model based on a one-meter default scale.

The DirectX® API allows functional control of those parameters, sound source positions, and listener orientation and position. It also provides creation, loading, playing, and looping management of sound buffers. Implementations of spatial auditory icons with the DirectX® API ensures functionality with any DirectX® compliant 3-D soundcard (Carl, 1998).

Summary and Objectives

The performance benefits associated with supplementing visual information with auditory information within a computer interface are well documented in the literature, both for people who are visually healthy and those who are visually impaired. Auditory icons and earcons most often serve as the auditory elements within GUIs, and can be designed to provide spatial localization information as well. The Head-Related Transfer Function is the vehicle that allows for the modeling and implementation of artificial 3-D sound and the implementation of 3-D sound is achievable with currently available software. The objective of this study was to explore the use of spatial auditory feedback to supplement the diminished visual feedback available to people who have visual impairments due to either HCIIIs or SIIs. It was hypothesized that the spatial auditory feedback should assist users with impaired vision while traversing a field of icons with a screen cursor, by enabling them to change their navigational strategy. The expectation is that the user will have a reduced dependency on the impaired visual feedback and utilize
the spatial auditory feedback to help determine the identity and positioning of icons close to the screen cursor.

Methodology
Participants

In order to empirically document the impact of adding sound and spatialized sound to a GUI, 10 volunteers, familiar with the operation of a standard GUI, were recruited to perform 81 trials of an icon identification and selection task, under artificial conditions of visual impairment in each of three interfaces using a) Ordinary Icons, b) Non-Spatial Auditory Icons, and c) Spatial Auditory Icons. The volunteers were recruited from the graduate and undergraduate student population of the Florida International University (FIU) College of Engineering (nine males; one female). The average age of the participants was 25.2 years, with a standard deviation of 5.29 years.

Apparatus

A 17” monitor, with a viewable surface of 32 cm x 24 cm = 768 cm$^2$ was used to display the GUI in this experiment. During the experiments, participants viewed this screen from a distance of 45 cm, and were placed such that the mid-point between their eyes was aligned with the center of the screen. Thus, a two-dimensional plane like the one shown in Figure 2 was presented to each user, where each of the icons and the screen cursor were represented with a specific pair of coordinates. In the interface, an audio cue was associated with each icon and processed through a 3-D audio channel.

As part of the experimental apparatus, the screen was mapped onto the horizontal plane in the 3-D sound model (i.e., elevation = $\phi$ = 0°). The horizontal plane was selected because the positioning of the ears accentuates time and intensity differences in azimuth, enhancing the accuracy of the 3-D sound localization in this plane (Begault, 1994). The audio cue was simulated such that it was emitted from the center of the iconic image. As the screen cursor moved through the field of icons on the screen, the user heard the audio cues in their spatial characteristics relative to the position of the screen cursor. Because too many unsynchronized audio events can be confusing, only icons in near proximity to
the cursor effectively contributed to the 3-D sound mixture heard by the user. For example, given the cursor’s position in the situation depicted in Figure 2, the user would hear the sound of the “My Computer” icon in the front and to the left. The “Netscape Communicator” icon would be heard from behind and to the right. The “Recycle Bin” and “Control Panel” icons would be heard similarly relative to the position of the screen cursor. This method converts auditory icons into spatial auditory icons.

Figure 2. Example of Iconic and Cursor Arrangement for Sound Localization

Data Structure

The heart of any implementation of spatial auditory icons is its data structure. In this study, the structure consisted of the iconic representation, sound, and location. The representation data were composed of a bitmap and its height and width. The interface used this data to render the graphical representation of the icon on the screen. The sound was comprised of a Microsoft WAVE PCM sound (“.wav”) file, its size (the number of samples), and the sampling frequency. The location was the x and y screen coordinates. The interface rendered the sound associated with each icon through a different 3-D sound audio channel. The emulated location of the sound was positioned at the center of the icon. The center of the icon was computed from the dimensions of the bitmap and the screen coordinates. The spatial auditory icon data structure described above is the elementary subset of all implementations in this research.
Spatial Auditory Icon Implementation

The experiment involved a total of nine icons, commonly encountered in the use of Microsoft® Windows® software. As mentioned, each icon sound was assigned to a separate 3-D audio hardware channel. The sound buffers for each channel were created and loaded with the sound samples for each corresponding icon. The 3-D sound API was updated with the parameters of the attenuation function. The center screen coordinates of each icon were recorded for each corresponding 3-D audio channel. The screen coordinates of the cursor were stored in the listener position. Once this initialization was completed, the sound buffers were played continuously in a loop. From this point on, as the screen cursor moved, the 3-D sound API was updated with the listener position changes. In this fashion, as the screen cursor was maneuvered through the field of spatial auditory icons, the spatial auditory feedback provided by the 3-D sound guided the user.

The nine sounds associated with graphic icons for the experiment are described in Table 1. This table shows some key visual and auditory traits of the icons used in this experiment. All icons were of size 32 x 32 pixels. All sounds used to represent the icons were processed at 44.1 KHz. The sounds used in the experiment were selected because they all have broad spectra, and because they semantically map to the computer functions they were designed to represent. This enhanced the participants’ localization process incited by spatial sound (King and Oldfield, 1997). Table 1 indicates, for each sound, the frequency boundary corresponding to 95% of the cumulative spectral power (accumulated from 0 Hz), confirming the wide bandwidth of these sounds. Table 1 also provides the duration of a single loop of each sound used. The disparity in the individual length of the sounds is acceptable, since each icon sound was played in an independent loop by the system.
### Table 1. Icons Used in the Experimental Program

<table>
<thead>
<tr>
<th>VISUAL TRAITS</th>
<th>AUDIBLE TRAITS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function</td>
<td>Picture</td>
</tr>
<tr>
<td>Copy</td>
<td>📄</td>
</tr>
<tr>
<td>Cut</td>
<td>✂️</td>
</tr>
<tr>
<td>Help</td>
<td>🕵️‍♂️</td>
</tr>
<tr>
<td>New</td>
<td>🗝️</td>
</tr>
<tr>
<td>Open</td>
<td>🖍️</td>
</tr>
<tr>
<td>Paste</td>
<td>📝</td>
</tr>
<tr>
<td>Print preview</td>
<td>🔍</td>
</tr>
<tr>
<td>Print</td>
<td>💼</td>
</tr>
<tr>
<td>Save</td>
<td>🎯</td>
</tr>
</tbody>
</table>

### Experimental Design and Dependent Variables

The experimental tasks were generated with a customized software tool developed using Microsoft® Visual C++® for the Microsoft® Windows® operating system. The tool was designed to evaluate the task performance of people who have impaired vision as they use icons that integrate spatial auditory feedback. The testing protocols of the Jacko Low Vision Interaction Assessment (JLVIA) were employed in the software (Jacko et al., 1999). A participant interacted with the software by performing a number of task trials. A task was composed of the following elements. First, the “Stimulus” screen was displayed. In this screen, the participant was visually shown a target icon, rendered at the top and center of the screen. The participant could also be aurally cued with the characteristic sound of the icon. Once the target was assimilated, the participant clicked on the “Target OK” button. The “Response” screen was next presented. In the “Response” screen, a three by three grid of nine icons, including the target icon previously shown, was displayed. The screen cursor was initially positioned at a fixed point at the bottom center of the screen. The participant then attempted to click on the target with the screen cursor. When the spatialized sound was enabled, the participant also had spatial auditory feedback for added guidance as the screen cursor was...
maneuvered to the target. An additional 440Hz sinusoidal tone was played when the
screen cursor was in the “hot-zone” or clicking area of any icon. This tone informed the
participant that clicking at the current position of the screen cursor would select an icon.
Once the user had clicked on an icon, the trial was completed. The number of unique pair
combinations generated from the set of icons and the set of grid locations was 81. This
ensured that the participant searched for each icon at each grid location; thus, a full test
consisted of 81 trials. To protect against order bias, the presentation of trials were
randomized. When the 81 trials were completed, the test concluded.

Data Collection. The program used for the experimental evaluation in this study
collected the following data through the 81 trials of each of the 3 tests in which each
participant was involved:

a) Selection time: Time elapsed from the appearance of the second (“Response”)
screen of each trial and the “Mouse down” event that was recorded when the user
performed a click on any icon (correct or incorrect).

b) Hit/Miss: Binary value indicating the correctness of the icon selection. A “Hit” is
counted if the icon selected in a given “Response” screen matched the target icon
presented to the user in the preceding “Stimulus” screen.

c) X-Y coordinates of the mouse positions, collected every 1 millisecond, while the
“Response” screen of each trial was displayed to the user.

In addition, the identity of the target icon and the specific ordering of the icons in
the grid, for each trial, were recorded.

Auditory Feedback Levels. The experimental program was flexible enough to be
configured to test different types of feedback. When sound was disabled, the condition
was defined as the “no auditory feedback”, or “ordinary icons” condition. When the
minimum distance was set to cover the hot-zone of the icons and the roll-off factor was
set to the maximum value of ten, then the condition was defined as the “non-spatial
auditory feedback” (“non-spatial auditory icons”) condition. When the parameters were
set such that the sounds that radiated from the icon centers overlapped each other, then
the condition was defined as the “spatial auditory feedback” (“spatial auditory icons”)
condition. In summary, the three conditions that served as the experimental treatments in this study were:

i. “Ordinary Icons” (NOSND): No supplementary auditory feedback was provided

ii. “Non-Spatial Auditory Icons” (SND): The characteristic sound associated with each icon was played (non-spatially) only when the cursor was placed on its active icon region.

iii. “Spatial Auditory Icons” (3DSND): The sounds associated with all nine icons were played to the user, spatially. In addition, a pure tone was also played when the cursor was located in the active region of any of the icons.

Experimental Procedure

As mentioned in the Introduction of this paper, experimental results were generated through the use of simulated visual impairments representative of those resulting from either HCIIs or SIIs in order to obtain a baseline of performance in the presence of visual feedback augmented by spatial auditory feedback. This was accomplished through the use of reduced peripheral vision goggles and a blurring glass that effectively reduced the visual field and visual acuity of each participant, respectively.

At the start of the experiment a participant was situated in the testing environment, in a chair whose height was adjusted such that the participant’s default point of gaze (looking directly forward) was aligned with the horizontal and vertical midpoints of the screen. The distance from the participant’s eyes to the screen was approximately 45cm.

To document the visual capabilities of the participants under the simulated visual impairments, visual field assessments were recorded, while the participant wore the reduced visual field goggles. The visual field accessible to each user was recorded as a viewable visual angle (Table 2), which, at the distance allowed between the user and the monitor, translated to a viewable percentage of the total screen surface. After the visual field assessment, visual acuity was measured with the blurring glass placed 5cm in front of the screen. A Snellen-like letter chart was used for visual acuity assessment (Volpe, 1997). Once the visual assessments were recorded, the participant performed three tests under conditions of 1) no auditory feedback; 2) non-spatial auditory feedback; and 3)
spatial auditory feedback. The ordering of the tests was randomized to prevent order effects.

For all three conditions, each participant wore the goggles while the blurring glass was placed 5cm in front of the monitor. Each participant was instructed on how to perform the tasks as described earlier, and a few practice trials were allowed. The custom software automatically stored the user performance data at the end of each test. Under the condition of no auditory feedback, the sound output of the computer was disabled. Under the condition of non-spatial auditory feedback, the participant wore headphones to hear the sound. Similarly, under the condition of spatial auditory feedback, spatialized sound, developed as explained earlier, was delivered to the user through headphones.

**Simulated visual impairment.** The effective levels of simulated visual impairment for each of the 10 experimental participants, achieved by the means described above, are summarized in Table 2.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Left Eye Visual Angle (°)</th>
<th>Right Eye Visual Angle (°)</th>
<th>Screen Visual Acuity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.1</td>
<td>10.6</td>
<td>20/40</td>
</tr>
<tr>
<td>2</td>
<td>10.3</td>
<td>9.2</td>
<td>20/40</td>
</tr>
<tr>
<td>3</td>
<td>6.7</td>
<td>6.7</td>
<td>20/30</td>
</tr>
<tr>
<td>4</td>
<td>7.0</td>
<td>7.3</td>
<td>20/30</td>
</tr>
<tr>
<td>5</td>
<td>7.6</td>
<td>8.7</td>
<td>20/30</td>
</tr>
<tr>
<td>6</td>
<td>8.3</td>
<td>9.3</td>
<td>20/30</td>
</tr>
<tr>
<td>7</td>
<td>7.9</td>
<td>7.9</td>
<td>20/40</td>
</tr>
<tr>
<td>8</td>
<td>8.5</td>
<td>8.5</td>
<td>20/30</td>
</tr>
<tr>
<td>9</td>
<td>6.7</td>
<td>7.2</td>
<td>20/40</td>
</tr>
<tr>
<td>10</td>
<td>6.2</td>
<td>7.0</td>
<td>20/60</td>
</tr>
</tbody>
</table>

Table 2. Visual Impairment Data

The largest visual angle for any of the participants was 10.6°. At a participant position of 45 cm away from the computer screen, the surface of the screen viewable under these circumstances was 222 cm². The effective viewable screen area of the monitor used in the experiment was 768cm². Therefore, a maximum of 28.9% of the screen was viewable by that participant with one eye. At the smallest visual angle of 6.2°, the viewable surface of the screen was 75cm². This area covers 9.8% of the screen.
These quantities summarize the range of field of view reduction achieved by the artificial impairments used.

Table 2 indicates that the simulated impairments used also implemented a significant reduction in the binocular visual acuity of the participants. The best visual acuity score recorded for the participants, using the procedure indicated above, was 20/30. The worst visual acuity score was 20/60. This range of binocular visual acuities is similar to that of some patients suffering from Health-Condition-Induced-Impairments, such as Age-related Macular Degeneration (AMD) (Quillen, 1999). Specifically, the range of binocular acuities is similar to the participants in the study by Jacko et al. (2002), and also similar to the “Group 2” AMD participants in the study by Jacko et al., 2000. This group accounted for 50% of the AMD participants involved in the latter study. Therefore, the level of visual impairment artificially imposed on the participants in this study was not markedly different from the impairments affecting those experiencing HCIIs.

Results

Each of the 10 participants completed an 81-trial test with the artificial visual impairments described and quantified above, for each of the three auditory feedback conditions: “Ordinary Icons” (NOSND), “Non-Spatial Auditory Icons” (SND) and “Spatial Auditory Icons” (3DSND).

The “hits” achieved by a user, in each of the tests were summarized in a “Hit Ratio” for the test, which is the percentage of correct target icon selections, with respect to the total number of trials in a test (81). Similarly, the sets of X-Y cursor coordinates during each trial were stored separately, so that “cursor traces” could be recreated for later examination.

The experimental data were analyzed through Repeated Measures Analysis of Variance (ANOVA), with the level of auditory feedback provided (NOSND, SND, 3DSND), as the
only independent variable, to investigate the impact of this variable on icon selection accuracy and icon selection time.

**Impact of auditory feedback on icon selection accuracy.** For this part of the analysis the average Hit Ratios achieved by each participant, under each one of the auditory feedback levels, were studied through a Repeated Measures ANOVA. Table 3 shows the Hit Ratio averages used in the analysis, as well as the descriptive statistics (mean, standard deviation, N) for the three levels of the independent variable.

<table>
<thead>
<tr>
<th>Participant (N = 10)</th>
<th>Ordinary Icons Hit-Ratio(%)</th>
<th>Non-spatial Auditory Icons Hit-Ratio(%)</th>
<th>Spatial Auditory Icons Hit-Ratio(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>80.2</td>
<td>96.3</td>
<td>98.8</td>
</tr>
<tr>
<td>2</td>
<td>79.0</td>
<td>82.7</td>
<td>97.5</td>
</tr>
<tr>
<td>3</td>
<td>64.2</td>
<td>100.0</td>
<td>97.5</td>
</tr>
<tr>
<td>4</td>
<td>85.2</td>
<td>98.8</td>
<td>100.0</td>
</tr>
<tr>
<td>5</td>
<td>65.4</td>
<td>98.8</td>
<td>95.1</td>
</tr>
<tr>
<td>6</td>
<td>56.8</td>
<td>91.4</td>
<td>95.1</td>
</tr>
<tr>
<td>7</td>
<td>67.9</td>
<td>96.3</td>
<td>97.5</td>
</tr>
<tr>
<td>8</td>
<td>76.5</td>
<td>98.8</td>
<td>100.0</td>
</tr>
<tr>
<td>9</td>
<td>75.3</td>
<td>98.8</td>
<td>97.5</td>
</tr>
<tr>
<td>10</td>
<td>70.4</td>
<td>96.3</td>
<td>95.1</td>
</tr>
<tr>
<td>Mean</td>
<td>72.1</td>
<td>95.8</td>
<td>97.5</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>8.68</td>
<td>5.21</td>
<td>1.88</td>
</tr>
</tbody>
</table>

Table 3. Data and descriptive statistics for accuracy analysis

The analysis was performed using the Statistical Package for the Social Sciences (SPSS) (Field, 2000). The result ($F_{(2,18)} = 64.85; p < 0.001$) confirmed that the level of auditory feedback provided was a significant factor influencing the icon selection accuracy achieved by the participants.

Results of post-hoc pairwise comparisons between the levels of auditory feedback are summarized in Table 4. This table reveals the presence of significant differences between the “ordinary icons” (i.e., without any auditory feedback), and either of the icons
with auditory feedback. No significant differences were found in the accuracy obtained with “Non-spatial Auditory Icons” vs. “Spatial Auditory Icons”.

<table>
<thead>
<tr>
<th>Pair</th>
<th>Mean Difference Hit-Ratio (%)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ordinary Icons – Non-spatial Auditory Icons</td>
<td>-23.7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Ordinary Icons – Spatial Auditory Icons</td>
<td>-25.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Non-spatial Auditory Icons – Spatial Auditory Icons</td>
<td>-1.6</td>
<td>&gt;0.05</td>
</tr>
</tbody>
</table>

Table 4. Results of post-hoc pair wise comparisons for accuracy

**Impact of auditory feedback on icon selection time.** The impact of the three different levels of auditory feedback on average icon selection time was also investigated through Repeated Measures ANOVA. Table 5 shows the average selection times recorded from the 10 participants under the three different auditory feedback levels, as well as the corresponding descriptive statistics (mean, standard deviation, N).

In this case, the result ($F_{(2,18)} = 0.242; p > 0.05$), did not indicate significant differences in icon selection time for the different levels of auditory feedback tested.

<table>
<thead>
<tr>
<th>Participant (N= 10)</th>
<th>Ordinary Icons Selection Time Avg. Time(s)</th>
<th>Non-spatial Auditory Icons Selection Time Avg. Time(s)</th>
<th>Spatial Auditory Icons Selection Time Avg. Time(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16.86</td>
<td>12.91</td>
<td>10.16</td>
</tr>
<tr>
<td>2</td>
<td>7.70</td>
<td>11.64</td>
<td>6.85</td>
</tr>
<tr>
<td>3</td>
<td>14.58</td>
<td>5.30</td>
<td>6.62</td>
</tr>
<tr>
<td>4</td>
<td>8.10</td>
<td>7.70</td>
<td>6.61</td>
</tr>
<tr>
<td>5</td>
<td>4.36</td>
<td>5.22</td>
<td>8.64</td>
</tr>
<tr>
<td>6</td>
<td>6.02</td>
<td>9.12</td>
<td>6.56</td>
</tr>
<tr>
<td>7</td>
<td>9.17</td>
<td>8.65</td>
<td>7.84</td>
</tr>
<tr>
<td>8</td>
<td>4.05</td>
<td>5.70</td>
<td>3.99</td>
</tr>
<tr>
<td>9</td>
<td>6.06</td>
<td>3.81</td>
<td>7.97</td>
</tr>
<tr>
<td>10</td>
<td>5.79</td>
<td>9.66</td>
<td>9.47</td>
</tr>
<tr>
<td>Mean</td>
<td>8.27</td>
<td>7.97</td>
<td>7.47</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>4.27</td>
<td>2.98</td>
<td>1.76</td>
</tr>
</tbody>
</table>

Table 5. Data and descriptive statistics for icon selection time analysis.
Discussion and Conclusions

**Impact on accuracy.** The statistical analysis indicates that icon selection accuracy was significantly enhanced by the addition of auditory feedback. Furthermore, accuracy in the presence of spatial auditory icons was markedly higher than accuracy in the presence of ordinary icons. However, in spite of the slightly higher mean Hit Ratio achieved with spatial auditory icons (97.5%), with respect to the non-spatial auditory icons (95.8%), the difference between these conditions was not statistically significant. For the interpretation of these results one should consider that, at the time of the decision that will result in a hit or a miss (i.e., right before the participant clicks on an icon), both types of auditory feedback (spatialized and non-spatialized) provide clues that are functionally equivalent to the participant (the sound of the icon at the current cursor position). At that specific instant, the spatialization of the sound of the icon currently pointed by the cursor does not add identifying information that could prevent a miss. On the other hand, the lack of any auditory characterization of the icon just before the selection is completed clearly deprives the participant of a second form of icon identification that could prevent a potential miss.

**Impact on selection time.** The results of the statistical analysis indicate that the level of auditory feedback did not introduce significant differences in the average icon selection time, in the experiment. In fact, in Table 5 note that the mean icon selection times for non-spatialized and spatialized icons (7.97 s and 7.47 s, respectively) are similar. On the other hand, from this same table it can be observed that, while most participants recorded a shorter average selection time with spatial auditory icons than with non-spatial auditory icons, participants 3, 5 and 9 exhibited the opposite behavior. This puzzling result led the experimenters to recreate the “cursor traces” of the 81 trials with non-spatial and spatial auditory icons for these participants, as well as for other participants who seemed to benefit from the spatialization of the icon sounds, so that this could be thoroughly investigated.

Figure 3 shows the superimposed 81 cursor traces from the experimental responses of participant 7, using non-spatialized and spatialized auditory icons. These typical traces reveal the strategy followed by most participants under each condition.
With non-spatialized auditory icons, participants engaged in sequential trial-and-error of the identity of the icons, using the residual visual information available to them to hop from one indistinguishable icon location to the next, relying on the non-spatialized icon sound for testing the identity of the current icon they were trying, and determining whether or not to hop to the next icon location. The predominance of cursor movement segments along the rows and columns of the grid of icons is the signature of this strategy, which was verified viewing the “animated” reconstruction of the individual trajectories, reconstructed from the stored X-Y cursor coordinates collected at equal intervals of 1 millisecond.

![Cursor Traces - Subject 7 - Non-Spatial Icon Sounds](image)

**Figure 3.** Cursor traces for participant 7 with: a) Non-spatialized icon sounds, b) Spatialized icon sounds.

The cursor traces collected when spatialized icon sounds were available to participant 7 (Figure 3b) reveal that this participant took advantage of the “remote” navigational information provided by the spatialized icon sounds, executing a much larger number of “diagonal” movements, outside the corridors defined by the horizontal and vertical lines forming the icon grid. Participant 7 benefited from this change of strategy, achieving a smaller average selection time when spatial auditory icons were used. This change of strategy, taking effective advantage of the spatial character of the icon sounds, was observed in seven of the participants, but not in participants 1, 3 and 9 (according to the quantification mechanism explained below). Specifically, the traces
from participant 3 are shown in Figure 4, where it is clear that an unchanged strategy of sequential trial-and-error was applied even when the icon sounds were spatialized.

![Figure 4. Cursor traces for participant 3 with: a) Non-spatialized icons, b) Spatialized icons s](image)

In order to quantify the utilization of “diagonal” cursor movements, away from the horizontal and vertical paths defined by the icon grid, which are made possible by the remote navigational information provided when the icon sounds were spatialized, six “off-grid” areas were defined, as shown in Figure 5.
Figure 5. Definition of six “off-grid” areas for the study, superimposed on a typical set of cursor traces.

For each participant and each auditory feedback condition (non-spatialized vs. spatialized), the total accumulation of cursor positions recorded within these rectangles was computed and compared to the total number of cursor positions recorded throughout the 81 trials of each test. This “Percentage of off-grid utilization” represents the level to which the user departed from the basic strategy of sequential trial-and-error, taking advantage of remote navigational cues. Table 6 shows the off-grid utilization percentages for non-spatial auditory icons, for spatial auditory icons and the ratio of the latter to the former.

<table>
<thead>
<tr>
<th>Participant (N= 10)</th>
<th>Non-spatial Auditory Icons Off-Grid Utilization</th>
<th>Spatial Auditory Icons Off-Grid Utilization</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30.97 %</td>
<td>19.01 %</td>
<td>0.6137</td>
</tr>
<tr>
<td>2</td>
<td>11.13 %</td>
<td>14.22 %</td>
<td>1.2771</td>
</tr>
<tr>
<td>3</td>
<td>2.97 %</td>
<td>2.63 %</td>
<td>0.8840</td>
</tr>
</tbody>
</table>
This table indicates that most participants seem to have changed their navigational strategy when the spatialized icon sounds added remote information that enabled an increased utilization of “off-grid” paths (Ratios larger than 1 in Table 6). Participants 1, 3 and 9 do not seem to have changed their strategy. Coincidentally, participants 3 and 9 are two of the three participants who did not demonstrate shorter average selection times with spatial auditory icons than with non-spatial auditory icons.

From the results, we surmise that the spatialization of the icon sounds does provide additional remote navigational information to the users, enabling diagonal movements towards the icons. However, it is up to the user to take advantage of this additional capability. The effective increase in the utilization of off-grid spaces (i.e., “diagonal movements”) requires an important strategic adjustment on the part of the user, who must exploit the remote navigational information provided by the spatialized icon sounds, abandoning the simpler strategy of sequential trial-and-error search, typically followed with non-spatialized icon sounds. If the user adjusts his/her strategy following the spatialized icon sounds, a shortened average selection time will generally be achieved. Six out of the seven participants who recorded an increased off-grid space utilization (left-most column in Table 6), simultaneously recorded a shorter average selection time with spatialized icon sounds than with non-spatialized icon sounds (two left-most columns in Table 5).

In addition to promoting a reduced mean selection time, the increased off-grid space utilization fostered through the spatialized icon sounds seems to re-enable the more efficient strategy that an unimpaired user would apply for the completion of the selection task. As an example supporting this observation, consider the cursor traces created by

Table 6. Off-grid utilization with non-spatial and spatial auditory icons.

<table>
<thead>
<tr>
<th></th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10.34 %</td>
<td>12.03 %</td>
<td>1.1626</td>
<td>7.48 %</td>
<td>18.16 %</td>
<td>2.4284</td>
<td>12.04 %</td>
</tr>
</tbody>
</table>

Table 6. Off-grid utilization with non-spatial and spatial auditory icons.
participant 6, a left-handed 34-year old male with normal vision, when no artificial impairments were imposed and no auditory feedback was provided (i.e., the normal operating circumstances of an individual with normal vision using an ordinary GUI), as shown in Figure 6. For reference, the six off-grid rectangles are also shown in this figure. The off-grid space utilization of this unimpaired experiment was 24.03%. Thus, it seems that the use of diagonal movement through off-grid spaces is a more natural navigational approach, which is re-enabled, at least in part, when remote directional information is provided by the spatialized icon sounds.

Future research should extend these results to additional degrees of visual impairment, particularly with respect to visual acuity, to determine if the provision of additional navigational information is useful to people with more extreme acuity loss. In addition, while the iconic selection examined in this study is highly representative of direct manipulation, other tasks required of people using GUIs should be explored more extensively. From this study, it can be concluded that spatialized auditory feedback has the potential to influence positively the navigational behaviors and strategies of visually impaired users when engaging in iconic search and selection.
Figure 6. Cursor traces created by participant 6 during the completion of an unimpaired test, while receiving no auditory feedback (Ordinary GUI interaction).

Acknowledgments

This research was made possible through funding awarded to Armando B. Barreto (IIS-0308155 and HRD-0317692) and Julie A. Jacko (BES-9896304 and IIS-0308155), by the National Science Foundation.
References


