Second LACCEI International Latin American and Caribbean Conference for Engineering and Technology (LACCET'2004)
"Challenges and Opportunities for Engineering Education, Research and Development"
2-4 June 2004, Miami, Florida, USA

Platform-independent 3D Sound Iconic Interface to Facilitate Access of Visually Impaired Users to Computers

Joaquin Prendes
Research Associate, Florida International University, Miami, Florida, USA

Armando Barreto, PhD
Associate Professor, Florida International University, Miami, Florida, USA

Abstract

While the introduction of Graphic User Interfaces (GUIs) facilitated the interaction of normally sighted users with computers, these benefits are not equally extended to partially sighted individuals that suffer from limited visual acuity. The Lighthouse Inc. has indicated that approximately 8.7 million Americans who are 45 or older report a severe vision impairment that cannot be corrected by wearing eyeglasses or contact lenses. When a computer user has uncorrected visual impairments, the identification and selection of icons in a GUI may become significantly more difficult than for a normally sighted user.

This paper describes the development of a platform-independent implementation of "3D Sound Icons". In addition to their graphical representation, each icon in this interface has a characteristic spatial sound (3D-sound), which is perceived by the user according to the spatial relationship between the screen cursor (listener), and the graphical icon (sound source), in the plane of the interface screen. This way, the user can supplement the visual information with spatial auditory information to identify the target icon and navigate towards it. The platform-independent implementation uses Digital Signal Processing functions that are capable of transforming an audio signal lacking spatial characteristics, into audio signals that provide the illusion of a point sound source located in a specific spatial location with respect to the listener. Platform-independence is critical in this application, because it will extend the benefits of the enhanced interface to users of a variety of operating systems (Windows, MacOS, etc.,) and types of computers (desktops, notebooks, palmtops, etc.)

Keywords

Assistive Technologies, Human-Computer Interfaces, Universal Access, 3D Sound Auditory Icons.

1. Introduction

While the advent of Graphic User Interfaces (GUIs) has simplified the interaction of many users with computers, it has simultaneously placed additional perceptual demands on the computer users. Approximately 8.7 million Americans who are 45 or older report severe vision impairments. Visual impairments significantly hinder one's access to information technology. Computers that rely on GUIs exclusively for user interaction are, to a large extent, inaccessible to people with reduced visual capabilities. The need for efficient interaction with computer systems is particularly pressing in today's world, where we use computers for many facets of our day-to-day activities. Although a number of "accessibility" features have been added to the most prevalent operating systems, visually impaired individuals continue to encounter barriers in their interaction with interfaces composed of visual icons.

There are different types of visual deficits that hinder the interaction with standard GUIs. Visual acuity and contrast sensitivity allow images to be viewed sharply, which is a prerequisite for proper icon localization and identification. In some individuals, however, these visual capabilities are severely diminished and cannot be restored completely by ordinary optical means. Some conditions, such as Age-related Macular Degeneration (AMD) and Retinitis Pigmentosa (RP) may affect the field of vision of an individual, so that proper visual perception is only available in sub-regions of the field of view (Quillen, 1999).

The focus of this paper relates to one particular aspect of interaction: the movement of the screen cursor on and around a target icon by persons with reduced vision. Low vision computer users experience difficulty identifying the icons they wish to select and positioning a screen cursor in the selectable region of the icons due to diminished visual feedback in the final approach of the screen pointer to the target icon. The system proposed addresses this problem by associating 3-dimensional auditory characteristics (3D-Sound) to the icons and screen cursor in a prototype interface. The intent of the system is to supplement the positional feedback that the subject ordinarily receives through visual perception with a real-time spatial auditory component to facilitate an accurate final approach to a correct target icon.

It has been observed (Jacko et al., 1999) that low vision computer users require more time to identify, position and select a target icon, compared to normally sighted individuals. In some cases these users will click on icons that are not their intended target icon. Additionally, it was observed that a portion of the additional time taken by these users for icon selection can be accounted for by the tendency to overshoot the location of the target icon. Typically, low vision computer users perform a successive approximation to the icon before they are able to successfully select it. The fundamental rationale behind the system proposed is that these inaccurate movements may be due, in part, to the lack of appropriate feedback to the subject, in terms of the instantaneous relative positions of the target icon and the screen cursor, during the final approach of the selection task. Accordingly, a 3D sound system has been developed to supplement the visual feedback provided to the user as the screen cursor approaches an icon with a spatialized auditory cue, characteristic to each particular icon. The relative positions of the cursor and a nearby icon, on the screen, are mapped to a virtual auditory domain of constant elevation. In this domain, 0^o azimuth matches the UP direction in the screen, $+90^o$ azimuth matches the RIGHT direction of the screen, -90^o azimuth matches the LEFT direction of the screen, and $+/- 180^o$ azimuth matches the DOWN direction of the screen. The mapping is such that the screen cursor position is associated with a moving listener, while the icons on the screen are associated to fixed sound sources. This is illustrated in Figure 1. The decay of sound intensity with distance is gauged so that the "listener" cursor will only hear characteristic sounds from icons that are in its close neighborhood. The intent of the system is to provide combined visual / auditory feedback for the user that will enhance the awareness of the instantaneous relative positions of the cursor and the icons on the screen, particularly the target icon, as the final approach of the selection task is executed.

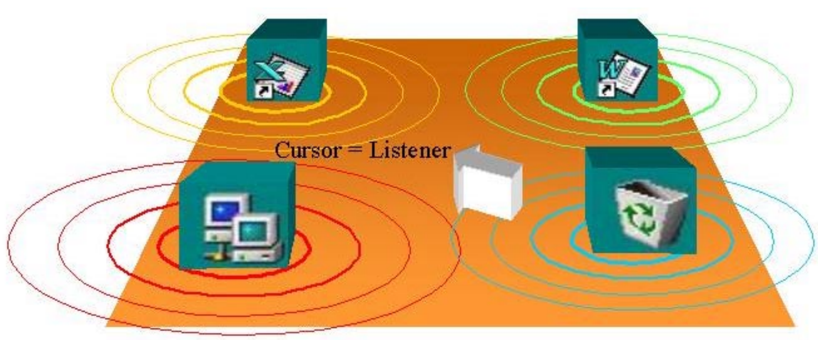

Figure 1. Mapping of icon locations and identities to a spatial sound environment

2. Objectives

The objective of this project is to explore the interaction enhancements that could be provided to partially sighted individuals by augmenting the icons in a standard GUI with characteristic 3D sounds, meant to provide supplementary positional information of the icon, relative to the position of the screen cursor. More specifically, this work aims at extending previous developments in this area to a platform-independent, software-only implementation constructed on the native 3D sound capabilities of Java 3D.

3. Scope of Work

This research pursued the development of a prototype of an augmented interface that provides each icon with a spatial sound, in addition to its regular graphical representation. This work also aimed at identifying the level of enhancement in the interaction with GUIs that partially sighted users may experience if the interface is augmented with 3D sound, as described. Lastly, one of the key goals of this work is the investigation of the use of Java 3D to achieve a software-only, platform-independent implementation of such augmented interface. This last goal is particularly significant because: a) This new type of implementation would take full advantage of the built-in 3D sound capabilities available in emerging programming languages, such as Java 3D; b) This type of implementation would take advantage of the significant computing power and multimedia capabilities that are available in all modern personal computers; and c) Only a platform-independent implementation, like the one sought, will truly make this concept available to a wide range of potential computer users.

4. Methodologies

The following subsections outline the basic method used to achieve sound spatialization and the implementation methods considered.

4.1. Sound Spatialization Through Head-Related Transfer Functions

Digital Signal Processing techniques can be used for transforming a single audio signal, lacking spatial characteristics, into a pair of left and right audio signals that provide the illusion of a point sound source located in a specific spatial location with respect to the listener. This is achieved in real-time by processing the input signal through a pair of Finite-Impulse-Response (FIR) filters (Figure 2) that affect the sound in the same way as it would be affected if it traveled from the emulated source location to the left and right eardrums of the listener, respectively (Kendall, 1995). The special transfer functions that must be implemented in the FIR filters for this purpose are known as Head-Related Transfer Functions (HRTFs). A different pair of HRTFs is needed for the virtual placement of the sound in each different position around the listener (Cheng and Wakefield, 2001).

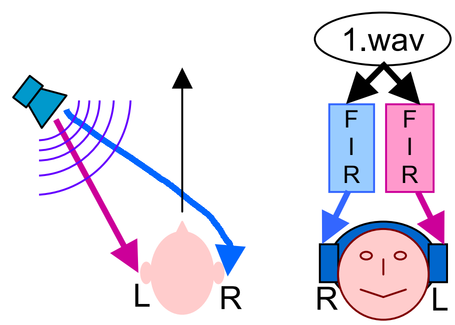

Figure 2. Measurement of the HRTFs (left) and their use to emulate sound spatialization of a monaural sound (right).

The illusion of a moving sound source is achieved by assuming changing azimuth and/or elevation values and using the corresponding HRTF pairs, from available libraries (Algazi et al., 2001), on the input sound as the virtual source travels. Similarly, simulation of a traveling listener in the surroundings of a fixed sound source requires dynamic re-assignment of the HRTFs used, in agreement with the path followed by the moving listener. If the virtual listener moves in an environment with several virtual sources, as many dual-HRTF architectures will be required. In this case, the outputs of all the left HRTF filters will be superimposed to constitute the left signal that will be offered to the user listening to the simulation. The overall right output for the headphones will also be obtained by mixing all the HRTF right outputs.

Our application requires the simultaneous spatialization of sounds from all the icons in the neighborhood of the computer cursor. Furthermore, our application requires real-time monitoring of the relative positions between the icons and the moving screen cursor, so as to dynamically change the HRTFs being used for the spatialization of the different icon sounds. Achieving these goals will provide low-vision computer users with an additional form of feedback to perceive the instantaneous relative placement of all the icons with respect to the screen cursor, supplementing the diminished visual input they obtain from the computer screen. This may help these users in identifying the icons they need to click on and in approaching them with greater accuracy.

4.2. 3D Sound Implementation Approaches.

Due to the amount of computation required to emulate spatialized sounds, real-time sound spatialization of the type required for the augmented interface described here originally required the use of a dedicated sound processor that implemented HRTF pairs in hardware. Previously, our group implemented a prototype of the 3D sound iconic interface based on a Diamond Monster MX300 plug-in board, developed around the Aureal AU3380 DSP processor, dedicated to the implementation of up to 16 3D audio channels. The software developed for the implementation of the augmented interface prototype tracked the positions of all the augmented icons and the cursor on the screen, and dynamically modified the properties of the 3D audio channels, through the Application Programming Interface (API) provided by the manufacturer of the board. While the hardware-based implementation was successful in demonstrating the benefits of the augmented interface concept during an evaluation (described below), this form of implementation is clearly constrained to computers equipped with the required type of add-on board.

In the last couple of years a marked increase in the computational capabilities of most personal computers has been observed. Simultaneously, recent years have witnessed the emergence of versatile object-oriented programming languages, with significant emphasis on multimedia applications, which have buit-in capabilities for the creation of spatialized sounds. For example, Java 3D was developed with platform independence as one of the top priorities. This implies that one should be able to transport a program developed under Java 3D to any machine running the most common operating systems, such as: Microsoft Windows, MAC OSx, Linux, Solaris and SGI IRIX. This kind of broad portability is accomplished by providing a “Java Virtual Machine” which acts as a layer between the Java code and the machine’s operating system. Specifically regarding the generation of spatialized sounds, Java 3D provides API libraries that enable a very complete control over the way in which spatialized sound is rendered, to provide a realistic 3D sound emulation.

The capabilities found in Java 3D allow the development of a “virtual world” in which the user is given visual and auditory clues to indicate where objects are located. Similar to computer games or virtual reality programs Java 3D is able to simulate 3-dimensional graphics and sound. By using Java3D in combination with the core Java API one can provide a GUI in which Java can listen for mouse movements and then translate them over to user movements in the virtual word. Therefore, each icon on the GUI can be associated with a *Point Sound* in the virtual world. The sound source or Point Sound then can indicate the position of the icon. The movement of the cursor on the desktop is paralleled by movement of the listener in the virtual world.

Point Sounds are sounds that radiate equally in a circular manner (in all radial directions). The emulated sounds decay as the listener is placed at locations that are farther and farther away from the virtual Point Sound source. The attenuation of the sound is controlled in Java3D by specifying the attenuation curve through a set of arrays which pre-determine levels of attenuation that must be implemented at different radial distances from the virtual Point Sound source. This attenuation profile is critical in helping the user determine the distance from the icon to the cursor position. Another advantage that Java3D has is its ability to modify the environment around the user. By providing an environment in which sound reverberates in a confined area the listener is better able to locate the virtual position of the sound source.

5. Evaluation and Results

In order to assess the level of interaction improvement that may be brought about by using 3D sound icons, tests have been conducted using the original prototype of the augmented interface. Five volunteers with normal vision participated in these tests, in which artificial visual impairments were simulated. To simulate a controlled, approximately uniform level of diminished visual acuity for the test, a frost glass plate was placed 1" in front of the computer screen during the test, causing the blurring of the images displayed to the experimental subjects. Similarly, a reduced visual field was imposed on the normal subjects by the use of blackened workshop goggles with only a small transparent circular opening (approximately 1/8" in diameter) directly in front of each of the subject's eyes. This reduced the subject's vision (with each eye) to a circular area of about 4" in diameter on the screen, when the subject was placed at a distance of 16" from the computer screen for the tests.

Each test involved 81 icon selection trials. Each trial consisted of two parts: Icon Presentation and Icon Selection. For the Icon Presentation stage the program displayed a screen with only one of the nine MS Windows® icons used in the test (Copy, Cut, Help, New, Open, Paste, Preview, Print and Save). The subject was asked to identify (to the best of his/her ability) this target icon, which should be selected in the second phase of the trial. The subject transitioned to the Icon Selection phase when he/she was ready, by clicking on a button provided in the lower half of the screen. This started a timer in the program and brought up a selection screen with all nine 3/8" x 3/8" icons arranged in a 3 x 3 square grid occupying an area of 3.5" x 3.5" on the upper half of the screen. The user had to displace the screen cursor from the lower half location, where the button of the presentation screen was, to the target icon and click on it, stopping the internal trial timer.

User accuracy was measured as a "Hit Ratio", i.e., the portion of selections made on the appropriate target icon. User speed was assessed through the "Selection Time" recorded by the internal software timer in the program. Each subject performed two evaluation tests: (1) Without any sound guidance for the subject ("No Sound"), and (2) With the spatialized sounds of all the icons provided to the subject through headphones ("3D-Sound"). (In the presentation screen the characteristic sound of the target icon was played for the subject, without spatialization). Table 1 summarizes the performance of each one of the five experimental subjects involved in the evaluation of the system, for both tests.

Table 1. System Evaluation Results

	NO SOUND		3D - SOUND	
SUBJECT	Hit Ratio	Mean Selection Time	Hit Ratio	Mean Selection Time
1	91.36%	5.894 sec.	98.77%	4.895 sec.
2	76.54%	7.012 sec.	100%	3.943 sec.
3	61.73%	4.199 sec.	98.77%	4.090 sec.
4	50.62%	3.790 sec.	96.30%	5.395 sec.
5	34.7%	6.593 sec.	91.36%	6.890 sec.
Average	62.99%	5.498 sec.	97.04%	5.043 sec.

The values in the table indicate that the accuracy of the selection process increases when the 3D-sound supplementary feedback is provided to the subjects. These improvements range from 7.41%, for subject 1, to 56.79%, for subject 5. The average improvement is 34%.

6. Conclusion

This paper has described the need, rationale and design philosophy associated with our development of an augmented user interface in which the icons are not only characterized by graphical representations but, in addition, have specific sound characterizations which are spatialized according to the instantaneous relative position of the icon (which acts as the sound source) and the screen cursor (which acts as the listener).

The results of the experimental evaluation of the augmented interface support our initial expectation that the 3D sound provides additional feedback to the visually limited user for the purposes of target icon identification and selection. Specifically, experimental results indicate that there is a significant improvement (34%) in the accuracy of the icon selection task when the 3D sound support is provided to the user.

Acknowledgements
This research was supported by NSF grants EIA-9906600, HRD-0317692, and IIS-0308155.

7. References

Algazi, V. R., Duda, R. O., Avendano, C., "The Cipic HRTF database", IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2001

Cheng, C.I. and Wakefiled, G.H. (2001). "Introduction to Head Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space", *Audio Eng. Soc*. (AES) vol. 49, No. 4, pp. 231-249.

Jacko, J. A., Dixon M. A., Rosa, R. H., Scott, I. U., & Pappas, C. J. (1999). Visual profiles: A critical component of universal access. Proceedings of the ACM Conference on Human Factors in Computing Systems (Pp. 330-337). New York: ACM Press.

Kendall, G.S. (1995). "A 3-D Sound Primer: Directional Hearing and Stereo Reproduction", *Computer Music Journal*, Vol.19, No. 4, pp. 23-46.

Quillen, D. A., (1999). "Common Causes of Vision Loss in Elderly Patients". *American Family Physician*, 60(1), 99-108.