USE OF 3-D SOUND AUDITORY ICONS IN THE ENHANCEMENT OF HUMAN-COMPUTER INTERFACES FOR PARTIALLY SIGHTED USERS

Barreto, A. B.♣, Hugh, P.O. ♣, Jacko J.A. ♣♣, and Adjouadi, M. ♣

♣Electrical and Computer Engineering Department and Biomedical Engineering Institute
Florida International University, Miami, FL, 33174

♣♣Department of Industrial and Systems Engineering
Georgia Institute of Technology, Atlanta, GA, 30332

KEYWORDS:

3-D Sound, Multimedia, Human-Computer Interaction, Partially sighted users, Low vision users.

ABSTRACT:

Graphical user interfaces (GUIs) have facilitated the interaction between most humans and computers. But they are not equally accessible to users with diminished visual capabilities. Our research proposes the addition of spatial auditory feedback to assist these users in the search for icons in GUIs. In the enhanced GUI, a unique 3-D sound is associated with each icon. As the cursor traverses the screen, the user hears the sounds of nearby icons, spatially, according to the relative position of each icon with respect to the cursor, through the dynamic implementation of Head-Related Transfer Functions that tracks cursor movements. A software prototype of the concept described was developed to evaluate the performance of users under artificially imposed visual impairments, in the search of icons within the proposed interface.

I. INTRODUCTION

According to The Lighthouse Inc. approximately 8.7 million Americans report a *severe* vision impairment not corrected by glasses or contact lenses, (http://www.lighthouse.org). Any visual dysfunction that reduces visual acuity, contrast sensitivity or color perception hinders the capability of an individual to use a standard GUI to interact with computer systems. Our research investigates one aspect of the human-computer interaction: The movement of the mouse cursor on and around a target icon, by persons with reduced vision. These users experience difficulty identifying the icons they wish to select and positioning a mouse cursor in the selectable region of the icons due to diminished visual feedback in the final approach of the mouse pointer to the target icon [Jacko et al., 1999]. This research addresses this problem by associating 3-Dimensional auditory characteristics (3D-Sound) to the icons 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 the correct target icon.

Low-vision computer users may click on icons that are not their intended target, or they may overshoot the location of the target icon, as they displace the cursor on the screen. This leads to the need for successive approximations to the icon before successful iconic selection is achieved. We propose 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 mouse cursor, during the final approach of the selection task. According to our hypothesis, we have developed a system that supplements the visual feedback provided to the user as the mouse cursor approaches an icon with a spatialized auditory cue, characteristic to each particular icon. The relative position of the cursor and a nearby icon, on the screen, is 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 cursor position is associated with a moving listener, while the icons on the screen are associated with fixed sound sources. 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 system aims to provide a combined visual / auditory feedback for the user that will enhance the awareness of the instantaneous relative positions between the mouse cursor and the icons on the screen, particularly the target icon, as the final approach of the selection task is executed.

II. SOUND SPATIALIZATION THROUGH HEAD-RELATED TRANSFER FUNCTIONS

A Digital Signal Processing system is capable of 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 1) 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].

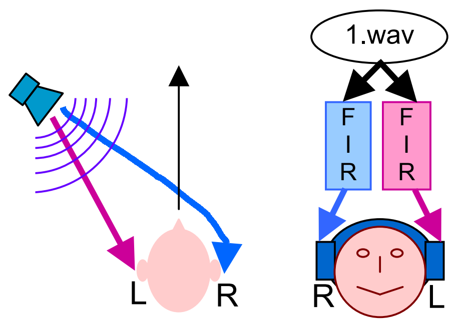

Figure 1. Diagram illustrating the measurement of the HRTFs (left) and their use to emulate sound spatialization of a monaural sound (right).

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 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.

III. SYSTEM HARDWARE

The hardware required for implementation of spatial auditory icons, as described above, includes a pair of stereo headphones and a soundcard implementing HRTF 3D sound. Soundcards that comply with the Microsoft DirectX Application Programming Interface (API) [Carl, 1998] implement HRTF 3D sound generally with an onboard DSP chip and memory for storing buffers and HRTF coefficients at different angles of azimuth and elevation (Begault, 1994). A "3D sound hardware channel" is an input to the mixer where HRTF 3D effects can be applied to the sound before mixing (Figure 2). When the effects are applied, the appropriate pair of HRTFs is selected based on the simulated position of the "listener" and the virtual location of the sound source. The FIR left and right outputs are calculated and passed to a distance attenuation function. The minimum distance, maximum distance, and roll-off factor of this function are adjustable. From zero to the minimum distance, the volume is fixed at unity gain. If the listener is at a distance greater than the minimum from the virtual source, the volume attenuates simulating the effect of sound waves traveling a distance through a medium. The roll-off factor controls the dampening effect of the medium that the sound travels through. The maximum distance defines the point where the sound source can no longer be heard. A distance factor scales the entire 3D sound model, based on a one-meter standard. The DirectX API allows functional control of those parameters, as well as the sound source positions, and listener orientation and position. It also provides the ability to create, load, play, and manage the looping of sound buffers. Implementations of spatial auditory icons with the DirectX API will ensure functionality with any DirectX compliant 3D soundcard. [Carl, 1998].

For the development of our prototype interface, the Diamond Monster Sound MX300 soundcard was selected, which uses the Aureal AU8830 chip. Its full-custom HRTF and atmospheric filters provide complete 16-bit, 48kHz quality through up to 16 hardware channels, whose outputs are mixed together [Shea, 1999] (Figure 2). The drivers for the card are DirectX compliant, but its API is even more flexible than the standard DirectX API.

IV. SYSTEM SOFTWARE

The spatial auditory icon data structure used consists of an image, a pair of screen coordinates, and an icon sound (Figure 3). The image data contains a pixel color matrix and its boundary sizes. The sound data comprises an array of audio samples and its boundary size. The center coordinates used to position the sound source in the 3D sound API can be computed using screen coordinates and the boundary sizes of the icon image. Programs that implement spatial auditory icons are responsible for the maintenance of a collection of these data structures.

An implementation of spatial auditory icons should perform the following tasks. Initially, the icons must be drawn on the screen. Each icon sound should be assigned to a 3D sound hardware channel, i.e., a left-right HRTF FIR pair. Ideally, there should be a 3D sound hardware channel available for every spatial auditory icon on the screen. If the number of icons required is more than the number of available 3D sound hardware channels, then the channels can be used by the 16 icons closest to the mouse cursor. The sound buffers for each channel are created and loaded with the audio samples for each corresponding icon. The 3D sound API is updated with the attenuation function parameters. The center screen coordinates of each icon are assigned to each corresponding 3D sound hardware channel. The screen coordinates of the mouse cursor are stored in the listener position. Once the initialization tasks are completed, the sound buffers are played in a loop fashion. Finally, when mouse or icon movements occur, the 3D sound API must be updated to reflect the appropriate position changes for the listener. Specifically, as the user maneuvers the mouse cursor through the field of spatial auditory icons, he/she will continuously receive updated clues about the identity and relative distance of neighboring icons. These clues will supplement the visual information that the computer display offers and promote an increased accuracy in the icon selection task. It should be noted that the software development of spatial auditory icons, as outlined here, could be adopted as an optional feature of the GUI system components of an operating system, such as the icon shell handler of Microsoft Windows.

V. EXPERIMENTAL EVALUATION AND RESULTS

The effectiveness of the system in improving the performance of users in an icon identification and selection task under conditions of visual impairment were investigated with a Visual C++ test program, which followed the icon presentation, identification and selection paradigm introduced in the Jacko Low Vision Interaction Assessment program: JLVIA [Jacko, et al., 1999]. Optionally, the test program implemented the spatialization of icon sounds, dynamically updating the sound spatialization, as the cursor (listener) navigated among the icons. The composite left and right signals from all the HRTF pairs involved in the simulation were delivered to the computer user through headphones. 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, who had normal vision. 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.

The system was tested with five college-aged volunteers of normal vision, using the artificial impairments described above. 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 mouse 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 Sel. Time	Hit Ratio	Mean Sel. Time
1	91.36%	5.894 s	98.77%	4.895 s
2	76.54%	7.012 s	100%	3.943 s
3	61.73%	4.199 s	98.77%	4.090 s
4	50.62%	3.790 s	96.30%	5.395 s
5	34.7%	6.593 s	91.36%	6.890 s
Average	62.99%	5.498 s	97.04%	5.043 s

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%.

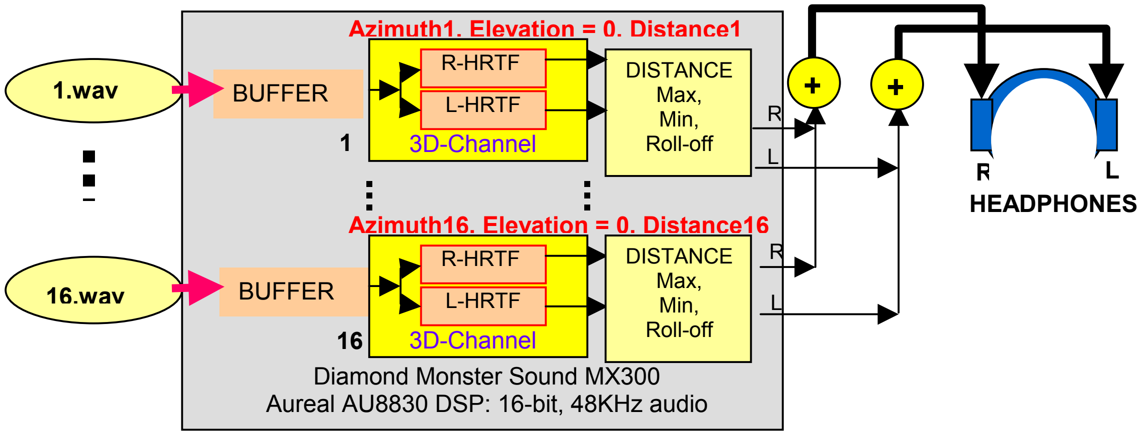

Figure 2. Block Diagram illustrating the hardware components of the system.

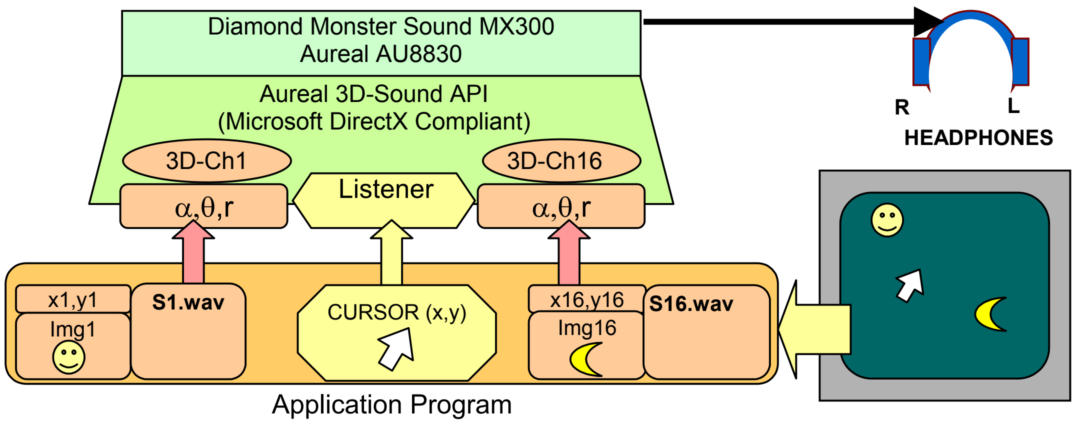

Figure 3. Block Diagram illustrating the software layers of the system.

The average change in mean selection time was a decrease of 0.455 seconds, with three subjects exhibiting a reduction (0.999, 3.069 and 0.109 seconds), and two subjects experiencing an increase (1.605 and 0.297 seconds) in mean selection time. Overall, these results suggest that the addition of 3D-sound to the characteristics of the icons helps increase the accuracy of the subjects in the task.

VI. CONCLUSIONS

The implementation and testing of the 3-D sound icon concept has demonstrated its potential to enhance the interaction of partially sighted users with GUIs. In spite of the uniformity of the artificially imposed visual limitations employed in our tests, inter-subject variability introduced by the use of different search and selection strategies resulted in marked differences in task performance and, most significantly, in the degree of improvement associated with the supplemental information provided by the augmented interface.

VII. ACKNOWLEDGMENTS

This research was supported by NSF grant EIA-9906600 and ONR grant N00014-99-1-0952.

VIII. REFERENCES

[1] Algazi V. R., Duda R. O., Thompson D. M. and Avendano C., "The CIPIC HRTF Database," Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics, pp. 99-102, Mohonk Mountain House, New Paltz, NY, Oct. 21-24, 2001.

[2] Begault, D. R., 3-D Sound for Virtual Reality and Multimedia, Academic Press, Chestnut Hill, 1994.

[3] Carl, J., “DirectSound”, in DirectX 6.0 Programmer's Reference, Microsoft Technical Document, 1998.

[4] Cheng C. I. and Wakefield G. H., “Introduction to Head-Related Transfer Functions (HRTF): Representations of HRTFs in Time, Frequency and Space” Journal of audio engineering society Vol. 49. No. 4, pp. 231 – 249, April 2001.

[5] 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 ACM CHI 99*, Pittsburgh, PA, May 15-20, pp. 330-337.

[6] Kendall, G. S., “A 3-D Sound Primer: Directional Hearing and Stereo Reproduction”, Computer Music Journal, 19:4, pp. 23-46, Winter 1995.

[7] Shea, M., A3D 2.0 API Reference Guide, Aureal Inc., 1999.